Conflict in Time and Space

Article · September 1999

Source: CiteSeer

15 authors, including:

Christopher Gelpi
The Ohio State University
48 PUBLICATIONS **3,296** CITATIONS

Zeev Maoz
University of California, Davis
130 PUBLICATIONS **6,284** CITATIONS

Paul Allison
Anglia Ruskin University
1 PUBLICATION **5** CITATIONS

Gary King
Harvard University
247 PUBLICATIONS **58,309** CITATIONS

Conflict in Time and Space

Richard Tucker and Nathaniel Beck

Weatherhead Center for International Affairs.

Working Paper Series
97-08
November 1997

Thanks to Scott Bennett; Paul Huth, Scott Bennett and Christopher Gelpi; and Frances Oneal, John Oneal, Zeev Maoz and Bruce Russett for supplying their data and to Paul Allison, Barry Bye, Robert Engle, Christopher Ferrall, James Hardin, Gary King, William Sribney, and Glenn Sueyoshi for sharing their expertise. This paper draws partially on research conducted with Simon Jackman. We are grateful to Jonathan Nagler, the director of the Laboratory for High Energy Social Science Research at the University of California, Riverside, and Kolin Hand, its systems programmer, for allowing us to undertake some of our larger computational efforts.

Abstract.

Scholars in international relations (IR) are increasingly using time-series cross-section data to analyze models with a binary dependent variable (BTSCS models). IR scholars generally employ a simple logit/probit to analyze such data. This procedure is inappropriate if the data exhibit temporal or spatial dependence. First, we discuss two estimation methods for modelling temporal dependence in BTSCS data: one promising based on exact modelling of the underlying temporal process which determines the latent, continuous, dependent variable; The second, and easier to implement, depends on the formal equivalence of BTSCS and discrete duration data. Because the logit estimates a discrete hazard in a duration context, this method adds a smoothed time term to the logit estimation. Second, we discuss spatial or cross-sectional issues, including robust standard errors and the modelling of effects. While it is not possible to use fixed effects in binary dependent variable panel models, such a strategy is feasible for IR BTSCS models. While not providing a model of spatial dependence, Huber's robust standard errors may well provide more accurate indications of parameter variability if the unit observations are intra-related. We apply these recommended techniques to reanalyses of the relationship between (1) democracy, interdependence and peace (Oneal, Oneal, Maoz and Russett); and (2) security and the termination of interstate rivalry (Bennett). The techniques appear to perform well statistically. Substantively, while democratic dyads do appear to be more peaceful, trade relations, as measured by Oneal, et al., do not decrease the likelihood of particpation in militarized disputes, Bennett's principal finding regarding security and rivalry termination is confirmed; his finding on common external threats, however, is not; his results on the influence of issue salience are even more robust.

Introduction

A new awareness of research design issues has compelled international relations (IR) researchers to use time-series cross-sectional (TSCS) data to analyze their hypotheses. TSCS data have been used in studies of the democratic peace, economic interdependence, diversionary war, alliance formation, rivalry onset/termination and deterrence. This research often studies binary outcomes -- whether or not nations are at war, allied or in a rivalry. [1] Scholars in these research programs are increasingly using TSCS data to analyze models with a binary dependent variable (BTSCS models).[2] These studies invariably use straightforward logit or probit analysis. This is highly problematic. Straightforward logit or probit analysis assumes that the observations are spatially and temporally independent. Violation of this assumption may lead to an incorrect inference.

Spatial independence refers to units (nations, or more commonly dyads of nations) [3] being statistically independent of other units at the same point in time.[4] This assumption seems untenable since one nation can be a member of many dyads and we would expect a relationship between two observations shared by the same dyadic partner. In contrast, temporal independence refers to units being statistically unrelated over time. Time series data, however, typically exhibit temporal dependence. In this paper we address the problems that spatial and temporal dependence cause for IR BTSCS data, and explore some solutions.

The lack of a solution to the problems associated with BTSCS analysis is not the result of a lack of awareness among IR researchers of the methodological issues involved. Maoz & Russet (1993,631) for example, have noted that "[m]any of the diagnostics appropriate to [TSCS] analyses using multiple regression are unavailable when the dependent variable is dichotomous or ordinal; the necessary computer power is lacking." Huth (1996,266), in assessing his own BTSCS study, remarked that "[t]he drawback of the logit (probit) estimator was that it could not take into account autocorrelation when estimating the standard errors of the coefficients." Ray (1995) and Spiro (1994) have argued that significance tests in pooled cross-sectional time-series studies (of which BTSCS is one form) are substantively meaningless because such designs artificially inflate the information available. The implication is that such designs lead to statistically significant results because of the huge "sample size" in typical IR BTSCS analyses. While these problems have captivated the attention of quantitative IR researchers, there have been virtually no attempts to develop methodological solutions.[5] Our work represents a preliminary effort to address the concerns of Huth, Maoz and Russett, Ray, and Spiro, and to employ the appropriate methods in the analysis of BTSCS data.

TSCS data are characterized by periodic observations on fixed units. In IR the units are dyads or nations, observed (usually) annually. BTSCS data are subject to many of the issues that arise in the analysis of both simple binary data and general TSCS data. In the latter, the critical issue involves trading off rich models which allow for diversity against the practical demands for estimable models that assume considerable homogeneity.[6] Analysts of binary variable models have many choices open to them: choosing methods like probit or logit are often of little consequence, however, such issues are not always trivial. These, and other general issues of model specification, are disregarded here unless they are specifically relevant to BTSCS data.

BTSCS models differ from binary dependent variable panel models. BTSCS models assume fixed units observed over a long time span; panel models assume a large sample of units observed over relatively short time spans.[7] This distinction is important because there has been a vast amount of research on the analysis of panel data with binary dependent variables. Researchers have struggled to find methods appropriate for large samples observed relatively few times. Since this is very different from the typical situation found in IR BTSCS analyses, it remains to be seen whether binary dependent variable panel methods are appropriate here. It is important, however, that IR researchers refrain from blindly adopting methods appropriate for one arena that may or may not be suitable in another.

In the next section we formally delineate the BTSCS model and discuss some difficulties in analyzing such data. In Sections and we discuss two different methods to account for temporal dependence in BTSCS models. One solution is computationally intensive and is just on the verge of becoming computationally feasible; another is relatively easy to implement with currently available hardware and software. In Section we present some methods to deal with cross-sectional dependence; i.e., fixed effects and Huber's robust standard errors. We illustrate these methods in Section through a reanalysis of the relationship between (1) democracy, interdependence and peace (Oneal et al. 1996); and (2) security and the termination of interstate rivalry (Bennet 1996). We conclude with an assessment of the feasibility of our proposed solutions.

BTSCS models

We assume that data are observed on a fixed set of units, $i = 1,...,N$. These units may be countries, or as in many analyses, pair of countries (dyads). By fixed units we mean that analyses are all conditional on the units observed, and it is the units themselves that are of interest. This is contrasted to the typical panel situation, where only a sample is observed, but interest lies in the universe from which that sample was obtained. IR researchers obviously must make choices about the units to be analyzed (such as all dyads versus only "politically relevant" dyads) but once we decide to include, say, the US-Zambia dyad, all thought experiments will contain the same US-Zambia dyad. The N for BTSCS studies may be large or small. But, unlike the panel situation, it does not make sense to think of experiments where $N \rightarrow \infty$.

These units are observed regularly for periods $t = 1,...,T$. In general IR data are observed annually, so we will typically refer to these periods as years. While panel data often has a small T (with single digit T's predominating), TSCS data has a reasonably large T. While we have no fixed lower limit for T, we can think of twenty or thirty years as a lower bound. For our purposes, the larger T is, the better off we are. T's of fifty or more are common in IR, with some studies having several centuries worth of data. In terms of asymptotics we can think of experiments for TSCS data where $T \rightarrow \infty$. Many binary dependent variable panel estimation techniques depend on T not being large; such methods will obviously not work well for typical IR BTSCS data.

We work only with single equation models. The binary dependent variable is denoted $y_{i,t}$, with $\vec{x}_{i,t}$ being a vector of (weakly exogenous) independent variables. For simplicity we assume that we have no missing data, so we have a rectangular data set. This assumption is just made for ease of exposition. It is easy to deal with data sets where units are observed for differing time periods; our reanalyses work with such data. It is always difficult to estimate dynamic models with missing data other than at the beginning or end of the time frame and our methods are similarly unable to handle such data. This problem is endemic to any method which allows for dynamics; the only way that missing data in the interior of a time frame can be easily ignored is if we also ignore dynamics.

TSCS data are subject to many complexities. Many of these complexities would arise with a continuous dependent variable. For TSCS models with a continuous dependent variable, ordinary least squares (OLS) is appropriate if the error terms are independent and identically distributed (iid). While we never observe the error terms, we can use the OLS residuals to test hypotheses about whether the errors are iid. TSCS errors may either be cross-sectionally dependent, temporally dependent or cross-sectionally varying (panel heteroskedastic). If these tests lead us to believe that the errors are not iid then we must turn to some variant of generalized least squares to estimate the model. These estimation techniques use the properties of the OLS residuals to transform the data, so that OLS is appropriate for this transformed data.

With a binary dependent variable, the analogue to OLS is logit or probit analysis. [8] BTSCS analysts estimate a logit equation of the form

$$\Pr(y_{i,t} = 1) = \frac{1}{1 + e^{-\vec{x}_{i,t}\beta}} \qquad (1)$$

which we can write more usefully in terms of a latent, continuous variable, $y^*_{i,t}$ as

$$y^*_{i,t} = \vec{x}_{i,t}\beta + \epsilon_{i,t} \qquad (2a)$$

$$y_{i,t} = 1 \text{ if } y_{i,t} > 0 \qquad (2b)$$

where ϵ has a logistic distribution. We can change this to a probit model by simply allowing ϵ to have a standard normal distribution.

Just as OLS is appropriate for TSCS data with iid errors, the logit is appropriate for BTSCS data if the $\epsilon_{i,t}$ in Equation (2a) are iid. Of course we never observe the ϵ. But the critical problem for BTSCS analysis is that there is no logit counterpart to the OLS residuals which can be used to test for whether the logit is appropriate. The lack of logit residuals also means that we cannot remedy problems via a generalized least squares approach.

BTSCS models with temporal dependance

Estimation of BTSCS models using straightforward logit is not appropriate for temporally or cross-sectionally dependent observations or for unit or time variations in the stochastic process. In this section we focus on temporally dependent processes, returning to cross-sectional issues in Section. Ignoring temporal dependence in TSCS models leads to inefficient and inaccurate estimates of coefficients variability (standard errors). The problems increase as the independent variables show temporal features such as autoregression or trend. We would expect that ignoring temporal dependence will have similar consequences for BTSCS data.

TSCS analysts treat temporal issues in one of two ways. The most common approach is to assume that the errors follow some simple process (most commonly a first-order autoregressive process). Such models can be estimated by generalized least squares. A more modern approach is to use a partial adjustment model, which is estimated by entering lagged values of the dependent variable into the model (which usually can then be estimated by OLS). Since BTSCS models do not have simple residuals, estimation of correlated errors is difficult. However, dynamics can be modeled using a lagged dependent variable. This can be easily estimated by any logit program, simply adding $y_{i,t-1}$ to the list of independent variables (and, of course, dropping the first observation for each dyad).

While such a procedure makes sense in the continuous TSCS realm, it might not make sense in the BTSCS realm. Consider the latent variable representation of this model:

$$y^*_{i,t} = \vec{x}_{i,t}\beta + \phi y_{i,t-1} + \epsilon_{i,t}. \qquad (3)$$

Equation (3) may not adequately represent current IR theories. To see why, consider Equation (3) as a model of conflict.

Equation (3) asserts that conflict this year makes the possibility of conflict next year more likely, and that peace this year makes conflict next year less probable. But consider a strain relief model of conflict. Now, imagine a dyad at peace. Over time, strain may build up, making war more likely. This is the opposite of what Equation (3) asserts. Now suppose war breaks out, relieving that strain. Strain relief makes war in subsequent years less likely, again contrary to Equation (3). [9]

To put strain relief in a different context, imagine a model of earthquakes, where $y_{i,t}$ indicates the occurrence of an earthquake in year t in area i. $y^*_{i,t}$ is the underlying propensity for an earthquake, that is, strain. Earthquakes relieve strain. Thus, once the aftershocks end (the analogue of ongoing wars) we would expect that an observed current earthquake makes the occurrence of earthquakes in the following years less, not more likely. But, over time, strain will again build up, so a long sequence of observed zeros makes the likelihood of an earthquake more, not less, likely. This is true only if the covariates in the earthquake model predict high strain, that is, a high value of y^*. Obviously zeros will be followed by zeros when strain, y^*, is low.

Now it is surely possible that peace breeds peace and war breeds war. Nations at peace build trade relations and such that may help ensure future peace; nations frequently at war may take actions that make future wars more likely. This is what Heckman (1981) calls "true state dependence." If being at peace makes it more likely that we will have peace in the future, it makes sense for policy-makers to keep peace, even if the latent propensity for war is high. But Heckman also defined "spurious state dependence" which has the opposite policy implications. It may appear that wars make future wars more likely because high values of the latent propensity for war, y^* are more likely to follow other high values. If this is the case, then attempts to reduce the likelihood of war by preventing current wars, in the face of a high propensity for war, are doomed to failure.[10]

There are two ways that y^* could be "sticky." The simplest explanation is temporally correlated errors in Equation 2a. Suppose that $\epsilon_{i,t}$ is positively correlated with $\epsilon_{i,t+1}$. Then a high value of $y^*_{i,t}$ may be associated with a higher value of $y^*_{i,t+1}$ because the ϵ component of both terms is related. Since the ϵ reflect unmeasured variables, many of which can be presumed to be relatively stable, this story seems quite plausible.

Alternatively, the conflict process might be best modelled by a partial adjustment in the propensity to engage in war. In this case the latent $y^*_{i,t}$ would be partially determined by $y^*_{i,t-1}$. This could be due to y^* being determined by present and lagged values of $\vec{x}$, with the effect of the lagged values declining exponentially. In some situations we could think about rational actors only partially adjusting y^* towards some optimal value, given that adjustment may not be without ramifications.

While these issues also arise in continuous TSCS analyses, the choice of dynamic model is less critical for the continuous dependent variable models. The issue of strain relief does not arise in the continuous models since here we observe the latent propensity, y^*, rather than limiting our observation to its dichotomous realization. While there is a choice between partial adjustment and serially correlated errors, there is a variety of well known tests that discriminate between these specifications; in practice either specification fits the data equally well. 11

We can rewrite Equation 2a to model these various alternatives to Equation 3. Thus spurious state dependence via serially correlated errors is modelled by

$$y^*_{i,t} = \vec{x}_{i,t}\beta + \nu_{i,t} + \theta\nu_{i,t-1}. \qquad (4)$$

where the ν are iid. 12 The partial adjustment model is modelled by

$$y^*_{i,t} = \vec{x}_{i,t}\beta + \phi y^*_{i,t-1} + \epsilon_{i,t}. \qquad (5)$$

(where the ϵ are now assumed to be iid). A simple strain relief model is:

$$y^*_{i,t} = \vec{x}_{i,t}\beta + \phi(y^*_{i,t-1} - y_{i,t-1}) + \epsilon_{i,t}. \qquad (6)$$

where the ϵ are again independent.

These models are not easy to estimate. On examination, it appears that the contribution of the *i*'th unit to the likelihood is given by a complex, high-dimensional integral. The assumption of independence allows us to break that *T*-fold integral into the product of *T* simple integrals. But such an assumption is untenable. However, we can use the laws of conditional probability to simplify the integration, reducing the problem to a series of simpler integrals:

$$\mathrm{Prob}(y_{i,T}, y_{i,T-1}, \ldots, y_{i,1}) = \mathrm{Prob}(y_{i,T} | y_{i,T-1}, \ldots, y_{i,1})\mathrm{Prob}(y_{i,T-1}, \ldots, y_{i,1}) \qquad (7)$$

At this point we could use direct numerical integration, simulation methods (Keane 1994), or the Gibbs sampler (Gilks & Spiegelhalter 1996). Interestingly, unlike in the continuous TSCS case, the simplest dynamic BTSCS model to estimate is the correlated errors model, an inherently conditional probability model, where the current observation depends on its predecessor.

These methods will not come without cost. Beyond the obvious computational issues, mathematical tractability demands that the error processes be normal, rather than logistic because the marginal and conditional distribution of a multivariate normal are normal. Further. these methods do not combine well with other non-standard methods.

We intend to estimate these models using either the Gibbs sampler or direct evaluation of the double integral for the correlated errors model. Computational advances clearly make such models feasible. We now turn, however, to a computationally simpler strategy for estimating some BTSCS models.

The event history approach

A simple computational strategy for estimating BTSCS data is based on the equivalence between event history (duration) models and BTSCS models. 13 While event history data can be analyzed either continuously or discretely (Allison 1982), little has been written on the use of event history ideas to analyze BTSCS data.[14] There is no reason why we should differentiate event history methods from binary dependent variable methods. As Alt, King & Signorino (1996) have recently argued, the use of either method is a choice made by the analyst, and not necessarily due to some inherent feature of the data. [15]

The event history approach seems helpful for IR BTSCS data. The information contained in the dependent variable (i.e., war or no war) is exactly the length of time between wars. We can choose to model this as a sequence of conditional probabilities, that is, the probability of war breaking out in a given year given that nations have been at peace up until then, in which case binary dependent variable methods are appropriate; or, we can choose to model this as the length of time between wars, using event history approaches. [16] Any BTSCS data can be modelled via event history approaches, and event history data modelled by BTSCS approaches.

The advantage of using event history methods for BTSCS data is that they typically do not assume that the data are temporally independent. While temporal dependence is the bane of BTSCS analysis, it is grist for the mill of event history analysis. Event history analysts call temporally independent data "duration independent" and temporally dependent data "duration dependent." Only the simplest event history models assume duration independence. Since we are only interested in discrete time data, we have focused on duration dependence in that context. [17]

To proceed more formally, note that for a given unit, *i*, the BTSCS dependent variables, $y_{i,1},...,y_{i,T}$ are a string of zeros and ones. For many, if not most, IR applications, we are modelling rare events. Thus we will observe long strings of zeros, followed by a one. For some data, e.g. Bennet's (1996) data on the termination of interstate rivalries, the first one ends the string of observations for that rivalry (that is, rivalries do not restart). Here the sequence of zeros followed by a one contains the same information as for the duration of the dyadic rivalry. For other data, a one may be followed by subsequent sequences of zeros terminated by a later one. Thus, in the democratic peace BTSCS data, dyads are observed over a fixed period. After a dyad goes to war, it will eventually return to peace. Here the sequence of zeros, ones, and subsequent zeros and ones contains the same information as a sequence of peace durations. Finally, we may observe units for which the dependent variable is always zero. In the democratic peace data, some dyads may never go to war. In event history jargon, such observations are "right censored." All we know is that they stayed at peace for at least as long as we observed the dyad, that is, we know the minimum duration of peace. (For observations with repeated sequences of zeros and ones, the last duration will typically be censored.) For less rare events, we may observe many sequences of zeros and ones. While each sequence of zeros followed by a one is a duration, it is probably less useful to think of such data as discrete duration data.

Any duration model has as its building the hazard function, which is a function of time. While this is mathematically complicated for continuous time models, in discrete time it is exactly the conditional probability of observing a one in time period *t*, given that zeros were observed at previous times, i.e., the probability of a duration ending in year *t* given that the duration was as long as *t* - 1. Logit analysis of BTSCS data is *equivelent* to the estimation of this discrete hazard function. But the logit analysis assumes duration independence, since the logits do not take time into account. [18]

To see this more formally, let $h(\tau)$ be the discrete hazard function. Let the time counter, τ, count the number of zeros that precede the current observation. Thus τ starts at zero. We use τ rather than *t* so that *t* can still denote annual time (as in $y_{i,t}$). We then have

$$h(\tau) = \text{Prob}(y_{i,t} = 1 | y_{i,t-1}, \ldots, y_{i,t-\tau} = 0, y_{i,t-\tau-1} = 1). \qquad (8)$$

This assumes that our counter starts at *t* = 1. If the counter has already started before *t* = 1, but we only observe data from *t* = 1 on, the data are said to be "left censored." Thus we may observe dyads from 1955 on, but all we know about dyads that were at peace in 1955 is that they have been at peace at least one period. It may be plausible to "restart" time as the end of a major war, such as World War II. But we clearly must worry about this issue in any given application. Even if we ignore left censoring, the event history approach is better than ignoring time information altogether.

Equation 8 can be estimated by a logit or probit or any other binary dependent variable method that theory or convention specifies. The simple logit analysis used by most BTSCS analysts assumes that the hazard function is time invariant (duration independent) so that all observations can be pooled into one grand logit analysis. A general form of a logit analysis that allows for duration dependence (that is, $h(\tau)$ varying with τ) is

$$\Pr(y_{i,t} = 1) = \frac{1}{1 + e^{-f_\tau(x_{i,t}\beta)}} \qquad (9)$$

where f_τ is an unspecified function of duration. But such a form is too general to be useful. Event history analysis typically considers the effect of time on the hazard rate as separable from the effect of the covariates. [19] The simplest assumption that still allows for duration dependence is that f_τ (Equation 9) can be written as

$$f_\tau(x_{i,t}\beta) = f(x_{i,t}\beta + s(\tau)) \qquad (10)$$

where s_τ is a function of time.

Since we are working in discrete time, we can specify *s* as a simple arbitrary variation over time and represent *s* as a series of dummy variables, that is, $s(\tau = i) = q_i$ where the *q*'s are a series of dummy variables. Although this approach is workable, it is difficult to get precise estimates of the q_i, that is, it would be hard to estimate the shape of the duration dependence. It would also be hard to test the null hypothesis of duration independence, that is, $q_i = q$. Unfortunately, we cannot simply parameterize *s* as a linear, quadratic, or other simple function of time. Because *s* appears to be fairly smooth, however, we can estimate $s(\tau)$ as a cubic "smoothing spline" (Hastie & Tibshirani 1990, 27-9) with smoothness controlled by the analyst. [20] This is the principal method used in our reanalyses of the various BTSCS studies.

The event history approach also helps us to understand whether BTSCS standard errors are overly influenced by enormous sample sizes. Some critics (e.g., Ray 1995) have argued that the BTSCS analyst is more likely to obtain statistically significant results because of the huge "sample size" in typical IR BTSCS analyses. Thus, with 50 years worth of data, information of 400 dyads becomes 20,000 observations of the yearly binary dependent variable. Since we often assume (incorrectly) that standard errors are determined by sample size alone, it might appear that BTSCS researchers obtain significant results unfairly. [21]

The event history approach shows that this criticism is only partially correct. Insofar as BTSCS analysts assume temporally independent observations, they overstate the amount of information in the sample, in the same way that time series analysts ignore temporally correlated observations and inflate the information in their sample. But the event history interpretation clearly shows that *N* duration observations are essentially equivalent to *NT* binary observations on the yearly duration process. Each of the *NT* binary observations contains considerably less information than the *T* durations, a difference that is contained in the Fisher information matrix. Therefore, BTSCS analysts are not inflating statistical significance by basing their results on a "sample size" of *NT*.[22]

The event history approach to BTSCS data is easy to implement. While the more computationally intensive techniques of Section may be theoretically superior, simply adding a smooth function of time to the logit estimator also corrects for some dependence. This simplicity has several advantages. In particular, it allows for the accounting of duration dependence in conjunction with other methodological corrections, such as for spatial dependence. We turn next to that issue.

Cross-sectional dependence

Unit effects

Cross-sectional dependence results from some unobserved dependency across units. In its simplest form it can be modelled by assuming that there is some unmeasured variable that affects all observations in the same unit in the same way. Thus we can add to the basic latent variable equation a unit specific term, giving

$$y^*_{i,t} = \vec{x}_{i,t}\beta + \mu_i + \epsilon_{i,t}. \qquad (11)$$

We can take the μ as either fixed or random. Fixed effects models are estimated by adding unit dummy variables to the specification; random effects models are estimated by taking account of the new, more complicated, error term, $\mu_i + \epsilon_{i,t}$.

Research on modelling effects has been propelled by panel analysts. With a huge *T* it is impossible to estimate fixed effects.[23] There are some clever solutions to this problem, in particular Chamberlain (1980) conditional logit model, which performs maximum likelihood conditional on the number of ones observed for each unit. Consequently, it has been well accepted by panel analysts, especially those dealing with large *N* panel data. This solution, however, comes with some costs and it may not always be necessary to expend these costs for IR BTSCS data.

The large *T*'s available for IR BTSCS studies might allow us to use simple fixed effects, that is, a series of unit dummy variables. Such a strategy is certainly feasible for continuous TSCS studies with *N*'s and *T*'s comparable to the IR BTSCS data. For continuous dependent variable TSCS models, a *T* of 30 or more allows for accurate estimations of the fixed effects; in such data, fixed and random effects become equivalent. But, as we saw in the previous section, the large *T* for BTSCS models does not indicate that that BTSCS data contain as much information as their continuous counterparts.

We can see the difficulty of estimating fixed effects BTSCS models, regardless of the size of *T*, by thinking about them as duration models. If each unit is observed only until the first "failure" (a dependent variable of one), then using simple fixed effects in BTSCS models is equivalent to inserting a unit dummy as a covariate for each observed duration. Obviously such a covariate will explain its duration rather well! Even in the more common IR situation, where each unit may have more than one failure, few failures per unit will occur. Thus the unit dummy variable is being estimated as the average of a few (perhaps only one) times until failure. We would expect, then, that many IR BTSCS models with fixed effects will yield inadequate results (poor estimates of the effects and no remaining explanatory power for the covariates).

We are more optimistic for the analysis of dyadic IR BTSCS data, where the effects pertain to each nation in the dyad, not the dyad itself. Thus for dyad with members *i* and *j*, modelling $\mu_{ij} = \eta_i + \eta_j$ is more sensible than allowing each dyad to have its own, unspecified, effect. With a complete set of dyadic data (that is, observations on all pairs of N_n nations), we have $N = \frac{N_n(N_n - 1)}{2}$ dyads to estimate N_n dummy variables. This will allow us to accurately estimate the unit dummy variables in several contexts. Of course there are often theoretical reasons for not analyzing complete dyadic data (Maoz & Russett 1993). For example, the Oneal et al. (1996) data on politically relevant dyads have approximately 100 nations but only 900 dyads. We return to this very practical issue in Section. But it is easy to estimate models with fixed effects, and is certainly feasible for dyadic BTSCS (as opposed to panel) models. Fortunately we have good indicators for the inappropriate addition of fixed effects. If, for example, standard errors on the effects dummies are high, and the impact of the substantive variables is neglible, we are almost certainly in a case where the effects explain all the variance, and hence the fixed effects should be dropped.

The alternative to fixed effects is to assume that the effects are random, that is, the μ_i's are drawn from some distribution (independently of the ϵ). Butler & Moffitt (1982) have provided a computationally feasible method for estimating random effects binary panel models. While their interest is in small *T* panels, it appears as though their approach will work with relatively large *T* (perhaps as large as 100). Should analysts use fixed or random effects? The choice here should be made on theoretical grounds rather than computational feasibility.

As with fixed effects, the appropriate way to model random effects in dyadic data is as the sum of the two unit effects. Unlike the Butler and Moffitt single integral computation, this would result in a double integral computation. While such a computation is feasible, we have not yet programmed this method, and so do not know how it would perform in practice.

Panel heteroskedasticity

There are other types of spatial dependence that cause problems for simple logit estimation of BTSCS models. Thus the errors in Equation 2a could be contemporaneously correlated ($\epsilon_{i,t}$ is correlated with $\epsilon_{j,t}$ or show panel heteroskedasticity ($\mathrm{Var}(\epsilon_{i,t}) \neq \mathrm{Var}(\epsilon_{j,t}), i \neq j$). Because we do not observe the latent ϵ, there is no simple BTSCS analogue to "panel correct standard errors" or "panel weighted least squares."

While we could model panel heteroskedasticity by using a unit dummy variable in a heteroskedastic probit (or logit) routine, the heteroskedastic probit simply changes Equation 2a by allowing the error process to have a variance of $(e^{\vec{z}_{i,t}\gamma})^2$, where $\vec{z}$ is a vector of covariates. Once we understand the process which determines error variances, we can model them as a function of covariates, as in standard cross-sectional heteroskedastic probit analysis (Alvarez & Brehm 1995). But BTSCS analysts seldom have such information. While we could employ unit specific dummy variables as the elements of $\vec{z}$ in a heteroskedastic probit approach, it will most likely fail for the same reason that fixed effects for BTSCS data cannot be modeled. Heteroskedasticity could be modelled in dyadic data if we assumed that heteroskedasticity is a function of the two national rather than the dyadic effects. But even if we could do so, atheoretical models of heteroskedasticity may not be the best way to proceed.

Panel heteroskedasticity leads to incorrect standard errors, that is, the estimated standard errors will not be good indicators of sampling variability in the presence of heteroskedasticity. Logit does, however, yield consistent, if not efficient, estimates of the model parameters ($\vec{\beta}$) even in the presence of panel heteroskedasticity. An alternative to modelling heteroskedasticity is to correct the logit (or other maximum likelihood) standard errors, continuing to use the logit estimates of $\vec{\beta}$. This approach, based on the work of Huber (1967), is the basis for Whites (1980) well-known heteroskedasticity-consistent standard errors.

If the logit is the inappropriate stochastic model, but the true model is in the very general "linear exponential family," correct standard errors can be estimated. This is the case if there is panel heteroskedasticity. The Huber standard errors may also be robust to other cross-sectional pathologies. Unfortunately we do not know under what conditions unspecified pathologies will lead to consistent parameter estimates and consistent robust standard errors.

One reasone to believe that the Huber robust standard errors are superior estimates of variability is because of their similarity to the jackknife (Efron & Tibshirani 1993, Ch. 21). Since the jackknife performs well, in general, we would expect the Huber standard errors to perform well. In any event, given that we rarely have perfectly specified models, it seems prudent to use the Huber standard errors. These errors are computed by first noting that the usual maximum likelihood standard errors are simply the square roots of the diagonal of the inverse of the Hessian (*H*), the matrix of second derivatives of the log likelihood. If the model is misspecified, we should instead use the diagonal elements of

$$V = H^{-1} ss' H^{-1} \qquad (12)$$

where *s* is the score vector, the vector of first derivatives of the log likelihood. (If the model is correctly specified, *H* = *ss'*, and so the standard errors can be obtained from the inverse of the Hessian matrix alone.) This would be appropriate for typical cross-sectional heteroskedasticity or other related pathologies. Given the panel nature of our data, we expect cross-sectional pathologies to pertain to dyads, not individual observations. To take this panel structure into account, Huber (1967) indicates that standard errors should be computed using the outer product of the sum of the scores for each dyad. Thus for BTSCS data we can defend against cross-sectional pathologies by computing standard errors as the square root of the diagonal elements of

$$V_{\mathrm{Huber}} = H^{-1} \left(\sum_{t=1}^{T} s_{i,t}\right) \left(\sum_{t=1}^{T} s_{i,t}\right)' H^{-1} \qquad (13)$$

We use this formula to compute robust standard errors for the subsequent reanalyses.

Reanalyses

In this section we reanalyze two recently published IR BTSCS studies. The first examines the effect of democracy and trade on militarized dispute onset (Oneal et al. 1996); the second analyzes the relationship between security and rivalry termination (Bennet 1996). We chose these studies because (1) they represent important research programs in conflict studies; (2) the data and research designs employed are textbook examples for the methods we discuss; and (3) the authors of these papers generously shared their data with us. [24]

Our goal here is to see whether our proposed methods resolve some methodological problems inherent in these studies and to determine their effects on substantive interpretations. The Oneal et al. 1996 findings are the strongest to date that illustrate the pacifying effect of dyadic economic relations. The findings reported in Bennet (1996) convincingly suggest a relationship between security concerns and rivalry termination. However, before we accept the implications of these findings, we need to determine whether they can withstand our methodological adjustments. In what follows, then, we will evaluate the implications of our methodological corrections by reexamining both studies. We begin with corrections for temporal problems, and then address cross-sectional issues.

Temporal dependence

The best analogy to duration dependence is serially correlated errors in time series analysis. In time series we attempt to model the cause of serially correlated errors, rather than treating them as an estimation nuisance (Hendry & Mizon 1978). The same procedure should be followed for models with duration dependence. Thus, the standard for a fully specified model should be the elimination of unexplained duration dependence. Until such a model is found, ignoring duration dependence, like ignoring serially correlated errors, can lead to incorrect inferences. Here we only take duration dependence into account, and do not provide a model of it.

Logits with duration dependence

Oneal et al. We begin the reanalyses with the Oneal et al. (1996) study of international conflict (OOMR). They study the onset of violent international conflict among 899 "political relevant" dyads observed from 1951--85. Their binary dependent variable measures the onset of a militarized interstate dispute (Gochman & Maoz 1984). This study extends Maoz & Russet's (1993) analysis of the "liberal peace" by incorporating a measure of economic interdependence as an explanatory variable. OOMR measures regime type (DEMOCRAT) by a dichotomy which is one if both dyadic partners are scored as democratic. Economic interdependence (INTERDEP) is based on the relative importance of trade between the dyadic partners (using a complicated function of exports and imports between the partners, relative to their GDPs). [25] OOMR also control for a variety of other factors: a dichotomous measure of whether the partners in the dyad were formally allied to each other or the US (ALLIES); a dichotomous measure of geographic proximity (CONTIG); the relative balance of forces between the partners (CAPRATIO), with each partner's capability measured by economic, military and demographic factors; and the average annual growth in real GDP of the dyadic partners over the previous three years (GROWTH).

OOMR used straightforward logit, that is, they assumed temporal independence. Our re-estimation of their major model is shown in Table 1. The two columns are an exact replication of this model; our estimates agree with their published results.

Table 1: Estimates of Logit Models of Post-World War II Disputes[a]

	Logit		Logit with Time	
	$\hat{\beta}$	SE	$\hat{\beta}$	SE
DEMOCRAT	−1.73	0.26	−1.54	0.26
GROWTH	−0.093	0.020	−0.079	0.021
ALLIES	−0.33	0.10	−0.13	0.10
CONTIG	1.17	0.10	0.67	0.11
CAPRATIO(/100)	−0.41	0.06	−0.46	0.07
INTERDEP	−3.59	1.74	−0.88	1.63
PEACEYRS			Figure 1	
CONSTANT	−3.45	0.09	−1.85	0.10
Deviance	5049.4		4244.7	
Degrees of Freedom	22568		22564	

[a]Dependent Variable is whether dyad engaged in a militarized dispute
N = 22575 dyad-years

We then added a smooth function of elapsed time to the logit. The measure of elapsed time, PEACEYRS, is the number of years since the dyad last experienced a militarized dispute (starting at zero with the first observation or the first observation after a dispute). The cubic smoothing spline in PEACEYRS allowed for three knots, using up four degrees of freedom.[26] The logit estimates, accounting for duration dependence, are shown in the third and fourth columns of Table 1. Figure 1 shows the smooth of PEACEYRS. [27]

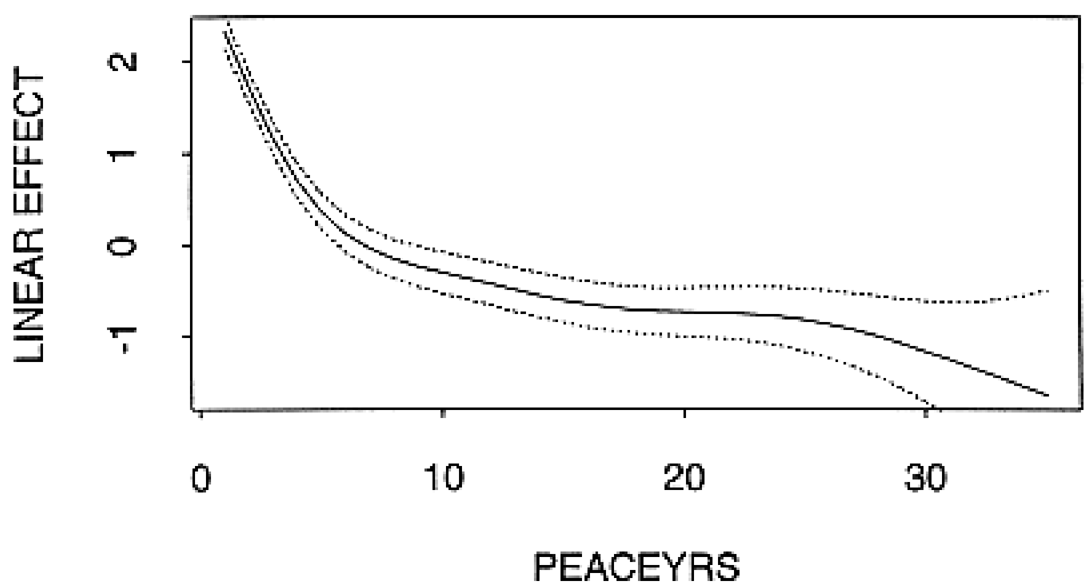

Figure 1: Effect of PEACEYRS on disputes: smoothing spline with approximate 95% confidence interval

The PEACEYRS smooth (Figure 1) and its approximate 95% pointwise confidence bands clearly show that the logit on disputes is duration dependent. A test for whether the smooth of PEACEYRS belongs in the logit clearly shows that it does. Since the two logit estimations are nested, the difference in their deviances (negative twice their log likelihoods) is χ^2_4. The probability of obtaining a decrease in deviance of about 800 by chance is zero (to as many decimal places as one cares to compute). Moreover, a test of whether the cubic spline is superior to a linear fit similarly provides overwhelming evidence against the simple linear form. While the probability of a dispute lessens the longer a dyad has been at peace, this effect is most noticeable in the years immediately following the end of militarized interstate disputes. The 95\% confidence bands indicates that duration dependence *may* disappear after a decade or more of peace.

The consideration of duration dependence changes some, although not all, of the OOMR substantive findings. Although the regime type component of the democratic peace hypothesis (DEMOCRAT) survives, the effect of the economic component (INTERDEP) on militarized disputes all but disappears. Thus, our reanalysis upholds one, and overturns the other, of the two fundamental findings of OOMR. The effects of economic growth and relative capabilities, however, remain virtually unchanged. Accounting for duration dependence, however, does change some of the other OOMR findings. Alliances are statistically insignificant and there is a considerable drop in the effect, and significance, of geographical proximity (CONTIG). Simple logit analysis, which assumes duration independence, overstates the effects of the parameters such as geographical contiguity that remain stable over the life of a unit.

Our new results provide only a starting point for continuing analysis. While we cannot trust estimation methods that ignore duration dependence, we cannot be satisfied with a model that simply says that disputes become less likely over time. Duration dependence must be explained.

Bennett Our second reanalysis examines Bennet's (1996) study of the effect of security concerns on the termination of interstate rivalries. The first major quantitative analysis of this topic, it examined 34 interstate rivalries from 1816 through 1988, using the dyad-half decade as the unit of analysis. [28] The dependent variable is whether a rivalry ended in any five year period (or, in a straightforward duration analysis, the length of each rivalry). Bennett hypothesized that rivalries were more likely to end when the rivals faced strong threats (SECURITY) or shared threats (COMMON THREAT) but that the more important the issues at stake (ISSUE SALIENCE), the less likely the rivalry would end. He also hypothesized that the relative military power of the rivals (BALANCE OF POWER), war between the rivals (WAR) or the polarity of the international system (BIPOLAR) would not influence rivalry termination.

Table 2 presents our reanalysis of Bennett's results on rivalry terminations. The first two columns show our reanalysis of Bennett's original BTSCS analysis, with results identical to his published results. We then added a variable, LENGTH, to the specification. which measures the elapsed time, in years, since the onset of a rivalry. Estimates using a smooth of LENGTH showed that we could not reject the hypothesis that LENGTH enters the specification linearly, so we only use the linear LENGTH term in the logits. The coefficient on LENGTH is positive and quite significant, showing that the Bennett BTSCS model is duration dependent, with the hazard (of rivalry termination) increasing with a continuation of the rivalry.

Table 2: Estimates of Logit Models of Rivalry Terminations[a]

	Logit		Logit with Time	
	$\hat{\beta}$	SE	$\hat{\beta}$	SE
CONSTANT	−2.31	1.51	−3.72	1.60
SECURITY	−0.05	0.02	−0.04	0.02
COMMON THREAT	2.29	1.12	1.55	1.21
BALANCE OF POWER	−1.39	2.31	−0.99	2.32
ISSUE SALIENCE	−0.45	0.27	−0.92	0.32
WAR	0.33	0.62	0.57	0.66
BIPOLAR	0.14	0.50	−0.25	0.57
LENGTH			0.036	0.0096

[a]Dependent Variable is whether a rivalry terminated
N= 414 dyad-half decades

Taking duration dependence into account, some of Bennett's findings are modified, and others strengthened. Accounting for duration dependence does not change the estimated effect of threats (SECURITY) or its standard error. The impact of common threats is cut by about a third and is no longer statistically significant; we can no longer conclude that rivals facing a common threat are more likely to end their rivalry. Issue salience, on the other hand, almost doubles its impact when we duration dependence is taken into account. Other variables, which Bennett hypothesized would have no effect on rivalry termination, and which were insignificant in the original analysis, remained insignificant even after accounting for duration dependence. Thus correcting logits for duration dependence may make estimated effects weaker or stronger, or leave them unchanged.

Our reanalyses of Oneal et al. (1996) and Bennet (1996) found evidence of duration dependence. However, we cannot simply assume that findings based on independent logit analyses are incorrect. Instead, we must test for duration dependence, discarding the simpler methodology when the tests suggest that we should do so. For example, we also looked for specific instances of duration independence (i.e., independent logit is acceptable). We also reanalyzed Huth, Bennet & Gelpi's (1992) study of risk propensity, system uncertainty, and conflict initiation among the major powers. Our results show no evidence of duration dependence. In a test of whether a smooth of rivalry age belonged in their specification, no evidence was found that indicated that conflicts were any more or less likely to develop at different rivalry ages. Thus, for these data, independent logit analyses are acceptable.

Duration models

Bennett Our confidence in the accuracy of the results in the above section is based on comparisons of our BTSCS analysis with standard event history methods. Although our interest here is in BTSCS analysis and not standard duration analysis, Bennet (1996) gives us an opportunity to also compare our proposed method with standard duration models. Bennett went beyond independent logit analysis and, using an identical specification, undertook an exponential duration analysis. His data look like classical duration data in that each rivalry is observed only until its termination (or censoring). Unfortunately, this cannot shed any light on the independence assumption, since it underlies both the exponential and the ordinary logit. However, we can take advantage of Bennett's work to see how well our remedy for BTSCS duration dependence works.

Table 3 represents a reanalysis of Bennett's exponential duration analysis. The leftmost columns replicate his exponential analysis, with results identical to his published results. There are many ways to examine the duration independence assumption that underlies the exponential. The simplest is to estimate a duration dependent model which nests the exponential and then to test whether the exponential model can be rejected. We use the Weibull for this purpose. The Weibull allows for flexible hazard rates so long as the hazard is monotonic. This assumption was proved to be consistent with the data.

The Weibull estimation results are given in the middle columns of Table 3. The σ parameter provides an indication of duration dependence. If $\sigma = 1$, then there is no duration dependence and the Weibull reduces to the exponential. Here, $\sigma = .34$ and is more than eight standard errors below one. Thus we can reject the hypothesis of duration independence at any level of significance. Since $\sigma < 1$, the longer a rivalry has been in existence, the greater the chance of its ending.

Does duration dependence have substantive consequences? Direct comparisons of the exponential and Weibull coefficients are difficult since the effects of the variables in the Weibull model depend on the σ and the β, all non-linearly. But we can compare *t*-ratios. The exponential analysis shows that both security and common threats have significant effects on rivalry durations. Issue salience, on the other hand, just misses being significant. The Weibull indicates the exact opposite. While the Weibull model may not be correct, the data suggest that it is superior to the exponential model, which fails to take duration dependence into account and leads to incorrect inferences.[29]

Table 3: Estimates of Duration Models of Rivalries[a]

	Exponential		Weibull		Cox[b]	
	$\hat{\beta}$	SE	$\hat{\beta}$	SE	$\hat{\beta}$	SE
CONSTANT	3.98	1.36	4.07	0.53		
SECURITY	0.039	0.022	0.011	0.009	0.044	0.026
COMMON THREAT	−2.02	1.01	−0.46	0.44	−1.37	1.31
BALANCE OF POWER	1.31	1.92	0.32	0.73	2.06	2.37
ISSUE SALIENCE	0.42	0.26	0.28	0.10	0.75	0.30
WAR	−0.31	0.63	−0.18	0.24	−0.82	0.65
BIPOLAR	−0.14	0.55	0.030	0.20	0.18	0.62
σ			0.34	0.074		

[a]Dependent Variable is rivalry duration in years: N= 34 rivalries
[b]Signs of coefficients reversed for comparability

Again, we are more interested in BTSCS analysis than standard duration analysis. Substantively, the results of our BTSCS analysis with LENGTH (Table 2) fall between the independent logit and Weibull results. Both the Weibull and the duration dependent logit show ISSUE SALIENCE to be statistically significant but not COMMON THREAT; the analyses differ on the significance of SECURITY. While we are pleased that both the Weibull and the BTSCS analyses correcting for duration dependence had roughly similar effects on both common threat and issue salience, the different findings on security puzzled us. The Weibull makes strong assumptions about the exact form of duration dependence, while the BTSCS analysis estimates this duration non-parametrically (although it did find a parametric, linear form to be adequate here). We therefore re-estimated Bennett's duration models using the semi-parametric Cox proportional hazard method. The results are in the last two columns of Table 3. [30]

The Cox regression results are consistent with the duration dependent BTSCS results in Table 2. Both analyses confirm the accuracy of the duration independent finding on SECURITY, and show a strong role for ISSUE SALIENCE on rivalry termination. However, both analyses cast doubt on the impact of

COMMON THREAT on rivalry termination. Thus we conclude that Bennett's finding on the role of security is probably correct, his finding on common threat is questionable, and the evidence for the role of issue salience is much stronger than Bennett found.

Bennett analyzed straightforward duration data, that is, observations on dyads that end with the first observation of one on the dependent variable. Standard duration methods, such as Cox regression, are well known, widely accepted methods for such data, in the presence of duration dependence. We were extremely pleased that our BTSCS duration dependence analysis agreed with the Cox regression results. This agreement gives us faith in our method for more traditional BTSCS data, where it is more difficult to apply Cox (and other standard duration methods). The interstate rivalry data clearly show the unity of BTSCS and duration methods.

OOMR. We also re-estimated the OOMR model in a duration context. Since the event history approach models durations of peace, not militarized conflict, we used the OOMR data to count the time between militarized disputes. We dropped all dyad-years that contain an ongoing dispute since such years gave us no information about the duration of peace.[31] Since dropping ongoing wars has dramatic consequences of its own, we first repeated our logit analyses with and without duration dependence in Table 4. The only change between Tables 1 and 4 is the elimination, in Table 4, of the 250 observations of ongoing disputes.

Table 4: Estimates of Logit Models of Post-World War II Disputes: Ongoing Disputes Excluded[a]

	Logit		Logit with Time	
	$\hat{\beta}$	SE	$\hat{\beta}$	SE
DEMOCRAT	−1.89	0.34	−1.72	0.33
GROWTH	−0.10	0.03	−0.09	0.03
ALLIES	−0.23	0.12	−0.14	0.12
CONTIG	1.16	0.13	0.86	0.13
CAPRATIO(/100)	−0.30	0.06	−0.35	0.07
INTERDEP	−0.31	1.75	−0.99	1.64
PEACEYRS			Figure 2	
CONSTANT	−4.08	0.11	−3.19	0.14
Deviance	3384.7		3244.5	
Degrees of Freedom	22318		22314	

[a]Dependent Variable is whether dyad engaged in a militarized dispute
N = 22325 dyad-years
250 dyad-years with ongoing dispute excluded

The elimination of ongoing disputes renders INTERDEP small and insignificant (even in an independent logit analysis). The change in the sample somewhat lessens the effect of ALLIES and CAPRATIO, but both effects remain strongly significant. The impact of DEMOCRAT, CONTIG and GROWTH is, however, almost unaffected by the elimination of years with continuing disputes.

The reduced data set still exhibits duration dependence, with a test of the null hypothesis that PEACEYRS does *not* belong in the specification being strongly rejected. The decrease in deviance from adding PEACEYRS is much less marked in the reduced data set. Figure 2 compares the effect of PEACEYRS on the latent dispute variable for the full and reduced data sets. The exclusion of observations with ongoing wars dramatically reduces the effect of PEACEYRS right after the end of militarized conflict. In the reduced data set, PEACEYRS has about half the impact as it does in the full data set (with all the difference occurring within the first three years of peace). 32

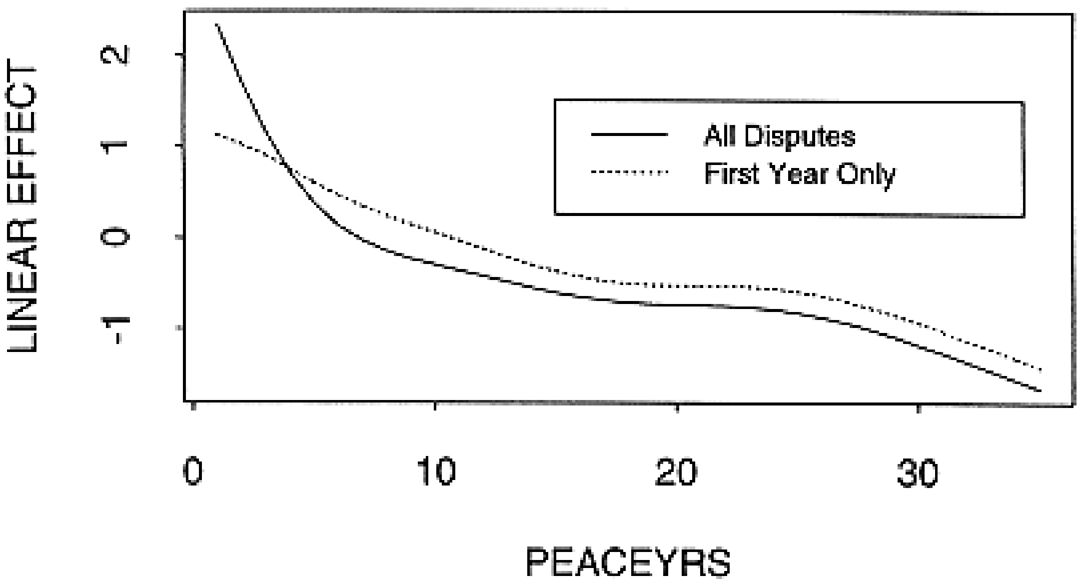

Figure 2: Effect of PEACEYRS on disputes: with and without ongoing disputes

Our interest in the reduced data set is to compare the duration dependent logit results with standard event history analyses of the same data. We present estimates from a Cox proportional hazard analysis, and a Weibull duration analysis in Table 5. 33 The Cox and Weibull results are similar to the duration dependent logit results. All three duration dependent methods find INTERDEP and ALLIES to be statistically insignificant, and the effect of CONTIG is diminished. The Weibull, like the duration dependent logit, finds that the hazard of war decreases over time. Thus these two duration analyses, along with those presented above in the Bennett reanalysis, show that our proposed method for treating duration dependence in BTSCS is consistent with established techniques. 34

Table 5: Estimates of Duration Models of Post-World War II Disputes[a]

	Cox[b]		Weibull	
	$\hat{\beta}$	SE	$\hat{\beta}$	SE
DEMOCRAT	1.82	0.33	2.35	0.44
GROWTH	0.10	0.03	0.11	0.04
ALLIES	0.05	0.12	0.23	0.16
CONTIG	−0.86	0.13	−1.30	0.19
CAPRATIO(/100)	0.40	0.08	0.42	0.10
INTERDEP	−0.69	1.55	−0.41	2.97
CONSTANT			4.36	0.19
σ			1.31	0.09

[a]Dependent Variable is duration of peace
22325 dyad-years, 1233 durations
250 dyad-years with ongoing dispute excluded
[b]Signs reversed for comparability

Cross-sectional dependence

Robus standard errors

We have examined the problems caused by temporal dependence and some solutions. We now turn to cross-sectional issues to determine whether the (independent) logit standard errors failed to take intra-dyad dependence into account. We know from the discussion above that, in the presence of spatial dependence, the estimated standard errors will be incorrect. A remedy for this is to use the Huber (robust) standard errors.[35] For convenience we repeat the OOMR logit results in the first two columns of Table 6. The third column shows the Huber standard errors.

Table 6: Estimates of Logit Models of Post-World War II Disputes: Huber Standard Errors[a]

	$\hat{\beta}$	SE	HSE
DEMOCRAT	−1.73	0.26	0.38
GROWTH	−0.09	0.02	0.02
ALLIES	−0.33	0.10	0.24
CONTIG	1.17	0.10	0.20
CAPRATIO(/100)	−0.41	0.06	0.12
INTERDEP	−3.59	1.74	2.56
CONSTANT	−3.45	0.09	0.16

[a]Dependent Variable is whether dyad engaged in a militarized dispute
N = 22575 dyad-years

Clearly, the Huber standard errors are larger than their "naive" counterparts; three of the Huber standard errors are more than double their naive counterparts. Even without accounting for duration dependence, neither INTERDEP nor ALLIES would have a statistically significant impact on war using Huber instead of naive standard errors. Since the Huber standard errors are more conservative, this casts even more doubt on the original OOMR findings.[36]

Fixed effects

Finally, we examined the use of fixed effects in the dispute and rivalry models, both of which use dyadic data. We did not simply use a different dummy for each dyad; such an attempt would have been doomed to failure. Instead, we modelled the fixed effect for a dyad composed of nations *i* and *j* by $a_i + a_j$ where the *a*'s are fixed effects (dummy variables). We expected this approach to work better for the Oneal et al. data than for the Bennet data.

Our approach to fixed effects is not completely trivial to implement. In both sets of data there are some nations that enter only a single dyad. If this dyad does not engage in a dispute, or is censored in the rivalry data, we cannot estimate the fixed effect for such a nation. We therefore eliminated from each set of data all dyads with nations of this type. This led to the loss of 3100 dyad-years (approximately 14\% of the total observations) or 65 dyads (approximately 7\% of the total dyads) in the OOMR data set.[37]

We begin with the OOMR analysis. Since their results are sensitive to small changes in sample, we re-estimated their original model on the smaller sample. Results are reported in the first two columns of Table 7. We then added fixed effects to the model, remembering that a dyadic effect is modelled as the sum of its two component effects. Results from this estimation appear in the righthand columns of the table.

Table 7: Estimates of Logit Models of Post-World War II Disputes With Fixed Country Effects[a]

	Logit: No Effects		Logit[b]	
	$\hat{\beta}$	SE	$\hat{\beta}$	SE
DEMOCRAT	−1.52	0.25	−1.63	0.30
GROWTH	−0.09	0.02	0.08	0.02
ALLIES	−0.40	0.10	−0.16	0.14
CONTIG	1.30	0.10	1.56	0.14
CAPRATIO(/100)	−0.38	0.07	−0.50	0.08
INTERDEP	−3.39	1.79	−1.69	2.12
CONSTANT	−3.40	0.09		
Deviance	4898.1		4279.6	
Degrees of Freedom	19452		19343	

[a]Dependent Variable is whether dyad engaged in a militarized dispute
N = 19459 dyad years
[b]110 country effects not shown

As we can see, the original OOMR results hold in the smaller sample. However, when we take into account the composition of the dyads, through the addition of national effects, INTERDEP and ALLIES no longer have a significant effect on peace. The other independent variables, including DEMOCRAT, remain highly significant (and relatively unchanged) despite the inclusion of the 110 unit effects. The inclusion of fixed effects provides a strong test of whether a substantive variable belongs in a model. Here the unit effects do not explain so much of the variation in rivalry models, both of wmilitarized disputes so that the substantive variables no longer have an effect. In addition, the estimates of the 110 national effects (not shown) also were reasonable. They are estimated sufficiently precisely so that we can differentiate them from the base effect-- which we took as the United States. Most of the fixed effects had *t*-ratios of two or more, showing that they can be estimated with enough precision to differentiate between more and less peaceful nations. These results increase our confidence that our approach to fixed effects modelling is possible for dyadic IR BTSCS data.

This approach depends on nations' membership in many dyads, so that we have enough information to reliably estimate national effects, and so that the dummy variables do not merely reflect the effect of a single (or a very few) dyads. The relative proportion of dyads to nations is much lower in the rivalry data than in the dispute data. When we added country effects to the Bennett model we found that no estimates remained significant, the country effects had large standard errors (no effects had *t*-ratios over two), and we had to disregard over one third of the observations because they were perfectly explained by the country effects. Thus while fixed country effects may be a good way to estimate dyadic BTSCS models when most countries enter many dyads, this technique is clearly not suitable for all dyadic data. It is even less likely that fixed effects will work for monadic BTSCS data, where separate dummy variables explain each "duration". Fortunately, the estimations clearly signal when fixed effects are not suitable. Researchers can attempt a fixed effects estimation and then decide if there is enough information in the data to warrant using such a model. BTSCS data differs from binary panel data, and IR researchers need not use the complicated methods required for binary panel data.

Thus it appears easier to handle spatial dependence in IR BTSCS data than in common panel data. Fixed effects and Huber standard errors can both be used to mitigate some of the consequences of such dependence. These methods can also be combined with the duration spline method used to treat temporal dependence, giving the analyst a series of remedies for problems that plague BTSCS data. [38]

Conclusion

We have proposed some remedies for common estimation problems in IR binary dependent variable time-series--cross-section (BTSCS) models. Such models present great difficulties for the analyst. For instance, there is no simple notion of a residual in binary dependent variable models. Thus, the various testing and estimation procedures available for continuous dependent variable time-series--cross-section models are not available for BTSCS analysts. This does not mean, however, that BTSCS analysts can ignore the threats to inference from both temporal and cross-sectional dependence. The most common method used by almost all IR scholars to estimate BTSCS models, simple logit, ignores both types of dependence.

One way to correctly understand the dynamics of BTSCS models is to focus on the latent variable representation of such models, i.e., to decide whether to model dynamics via correlated errors, lagged latent variables, or, more rarely, by using lagged realized values. While the latter often works well for continuous models, it usually does not lead to sensible results in the IR BTSCS context. Computational breakthroughs make it possible to estimate BTSCS models with correlated errors or lagged latent dependent variables. Since we have not yet implemented these methods, we cannot determine how well they will work in practice. In particular, it is not clear if BTSCS data contain enough information to allow for precise parameter estimates using these methods.

A simpler, at least computationally, approach to dynamics comes from the recognition that BTSCS models and discrete duration models both represent the same underlying phenomenon. The common logit model estimates a discrete hazard model given the assumption of duration independence. We can allow for duration dependence by simply adding a smooth function of time to the logit equation; using either smoothing or natural splines; while the former is preferable, the latter is available in commonly available software packages. This is the simplest way to account for duration dependence in a BTSCS model and appears to work well; It also allows to test the assumption of duration independence, e.g., if the starightforward logit is appropriate, or if the specification ought include the smmoth function of time.

While it might appear that the simple duration splines are inferior to the more impressive latent dynamic models, the simple splines can be easily incorporated in other, more complicated, models. In a related project, for example, we are estimating the probability of conflict as a non-parametric function of dyadic democracy. This is easy to do with duration splines, but unfeasible with the latent dynamic model. Finally, we also successfully combined duration splines with the various recommended cross-sectional corrections. There is nothing wrong with simple solutions so long as they are solutions.

The standard event history method approach to estimate BTSCS data may be preferred in some cases. At present we cannot determine when the duration-dependent logit approach is better or worse than standard continuous time event history methods. The two related sets of techniques may yield similar findings. It is therefore prudent to employ both types of methods.

Much of the conventional wisdom about modelling cross-sectional effects in BTSCS data comes from the analysis of binary dependent variable panel models. These models, with large cross-sectional, but short-time samples, are subject to problems that differ from the typical IR BTSCS model, where fixed effects are often used. This is especially the case for the most common IR data sets that use the dyad-year as the unit of analysis. Here, estimating the dyadic fixed effect by taking the sum of its two component effects is theoretically sensible, computationally simple, and, at least in some instances, provides reasonable results.

Estimates of robust standard errors for BTSCS models are also computationally simple. BTSCS data may show panel heteroskedasticity or other unaccounted for (by the logit model) unit-to-unit variability, that make the usual logit standard errors overly optimistic. Huber's robust standard errors easily guard against this over-optimism. Since the robust errors are a form of the jackknife, we can expect them to perform reasonably well in a variety of situations. At a minimum, if the Huber standard errors differ (non-trivially) from the usual logit standard errors, we know that the simple logit model is misspecified (in some unknown manner).

We currently lack the experience to estimate BTSCS models other than by logit. Reanalyses indicates that our proposed methods are feasible and appear to produce reasonable and substantively interesting results. Since BTSCS models parallel discrete duration models, our proposed methods should perform well. More experience will enable us to more confidently evaluate the performance of our proposed methods.

As computational advances continue at an accelerated rate, theoretical methods of modeling temporal dynamics will become part of the toolkit of the quantitative IR scholar. Advances can also be made in the analysis of BTSCS data by clarifying the data modelling approach and utilizing current technology.

References

Allison, Paul D. 1982. Discrete-Time Methods for the Analysis of Event Histories. In *Sociological Methodology*, ed. Leinhardt, Samuel. San Francisco: Jossey-Bass pp. 61-98

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Data*. Newbury Park, Ca.: Sage

Alt, James E. and Gary King and Curtis Signorino. 1996. "Estimating the Same Quantities from Different Levels of Data: Time Dependence and Aggregation in Event Count Models". Paper presented at the Annual Meeting of the Midwest Political Science Association, Chicago.

R. Michael Alvarez and John Brehm. 1995. "American Ambivalence Towards Abortion Policy: Development of a Heteroskedastic Probit Model of Competing Values." *American Journal of Political Science* 39(4):1055-82.

Anderson, P. K. and R. D. Gill. 1982. "Cox's Regression Model for Counting Processes: A Large Sample Study." *Annals of Statistics* 10:1100-20

Barbieri, Katherine. 1996. "Economic Interdependence: A Path to Peace or a Source of Interstate Conflict?" *Journal of Peach Research* 33(1):29-50.

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

Berry, Frances S. and Willaim D. Berry. 1990. "State Lottery Adoptions as Policy Innovation: And Event History Analysis." *American Political Science Review* 84(2):395-415.

Bremer, Stuart A. 1992. "Dangerous Dyads: Conditions Affecting the Likelihood of Interstate War, 1816-1965." *Journal of Conflict Resolution* 36:309-41.

Bremer, Stuart A. 1993. "Democracy and Militerized Interstate Conflict, 1816-1965." *International Interactions* 18:231-49.

Bremer, Stuart A. 1996. "Power Parity, Political Similarity, and Capability Concentration: Comparing Three Explanations of Major Power Conflict." Paper presented at the Annual Meeting of the Internation Studies Association, San Diego.

Bueno de Mesquita, Bruce & David Lalman 1992. *War and Reason: Domestic and International Imperatives.* First ed. New Haven: Yale University Press.

Butler, J. & R. Moffit. 1982. "A Computationally Efficient Quadraditure Procedure for the One Factor Multinomial Probit Model." *Econometrica* 50:761-4

Chamberlain, Gary. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies* 47:225-38.

Cox, D.R. 1975. "Partial Likelihood." *Biometrika* 62(2):269-76.

Diehl, Paul & Gary Goertz. 1993. "Enduring Rivalries: Theoretical Constructs and Empirical Patterns." *International Studies Quarterly* 37:147-71.

Efron, Bradley & Robert J. Tibshirani. 1993. *An introduction to the Bootstrap.* New York: Chapman & Hall.

Enterline, Andrew. 1996. "Driving while Democratizing." *International Security* 21:183-96.

Enterline, Andrew. 1997. "Fledgling Regimes: Is the Case of Inter-War Germany Generalizable?" *International Interactions* 22(3):245-277.

Farber, Henry & Joanne Gowa. 1995. "Polities and Peace." *International Security* 20:123-46.

Farber, Henry & Joanne Gowa. 1997. "Common Interests or Common Polities? Reinterpreting the Democratic Peacs." *Journal of Politics* 59(2):393-417.

Flemming, T.R. & D.P. Harrington. 1991. *Counting Processes and Survival Analysis.* New York: Wiley.

Gartzke, Erik. 1998. "Kant We All Just Get Along?: Opportunity, Willingness and the Origins of the Democratic Peace." *American Journal of Political Science* 42(1):000-000.

Gilks, W.R. and Richardson, S. & D.J. Spiegelhalter. 1996. *Markov Chain Monte Carlo in Practice.* London: Chapman and Hall.

Gleditsch, Nils Petter & Havard Hegre. 1995. "Peace and Democracy: Three Levels of Analysis." Manuscript, International Peacs Research Institute, Oslo, Norway.

Gochman, Charles S. & Zeev Maoz. 1984. "Militerized Interstate Disputes, 1816-1976." *Journal of Conflict Resolution* 29:585-615.

Hastie, T.J. & R.J. Tibshirani. 1990. *Generalized Additive Models.* London: Chapman and Hall.

Heckman, James. 1981. Statistical Models for Discrete Panel Data. In *Structural Analysis of*

Discrete Data with Econometric Applications, ed. Charles Manski & Daniel McFadden. Cambridge, MA: MIT Press pp. 114-78.

Henderson, Errol A. 1997. *Journal of Conflict Resolution*.

Henry, David & Graham Mizon. 1978. "Serial Correlation as a Convenient Simplification, Not a Nuisance. A Comment on the study for the Demand of Money by the Bank of England." *Economic Journal* 88:549-563.

Hensel, Paul. 1996. "Adding Event Data to the Study of Interstate Rivalry." Paper presented at the Annual Meeting of the Internation Studies Association, San Diego.

Hermann, Margaret G. & Charles W. Kegley. 1996. "Ballots, a Barrior against the Use of Bullets and Bombs: Democratization and Military Intervention." *Journal of Conflict Resolution* 40:436-460.

Hsai, Cheng. 1996. *Anaysis of Panel Data*. New York. Cmabridge University Press.

Huber, Peter J. 1967. The Behavior of Maximum Livelihood Estimates Under Non-Standard Conditions. In *Preceedings of the Fifth Annual Berkeley Symposium on Mathematical Statistics and Probability*, ed. Lucien M. LeCam & Jerzy Neyman. Vol I Berkeley, Ca.: University of California Press pp. 221-33.

Huth, Paul 1996. *Standing Your Ground: Terrritorial Disputes and International Conflicts.* Ann Arbor, Michigam: University of Michigan Press.

Huth, Paul & Bruce Russett. 1993. "General Deterance Between Enduring Rivals: Testing Three Competing Models." *American Political Science Review* 87:61-73.

Huth, Paul, Christopher Gelpi & D. Scott Bennet. 1993. "The Escalation of Great Power Militarized Disputes: Testing Rational Deterrence Theory and Structural Realism." *American Political Science Review* 87:61-73.

Huth, Paul, D. Scott Bennett Christopher Gelpi. 1992. "System Uncertainty, Risk Propensity, and International Conflict." *Journal of Conflict Resolution* 36:478-517.

Keane, Michael P. 1994. "A Computationally Practical Simulation Estimator for Panel Data." *Econometrica* 62(1):95-116.

Krain, Matthew. 1997. "State-Sponsored Mass Murder: The Onset and Severity of Genocides and Politicides." *Journal of Conflict Resolution* 41(3):331-360.

Leeds, Ashley & David Davis 1996. "Domestic Political Vulnerability and International Disputes." Paper presented at the Annual Meeting of Internation Studies Association, San Diego, California.

Lemke, Douglas & William Reed. 1996. "Regimes Types and Status Quo Evaluations: Power Transition Theory and the Democratic Peace." *International Interactions* 22.

Mansfield, Edward & Jack Snyder. 1996. "The Effects of the Democratization on War." *Internation Security* 21:196-207.

Mansfield, Edward & Jack Snyder. 1997. "A Response to Tucker and Thompson." *Journal of*

Conflict Resolution 41:000-000.

Maoz, Zeev. 1996. *Domestic Sources of Global Change*. Ann Arbor: University of Michigan.

Maoz, Zeev. 1997. "Realist and Cultural Critiques of the Democratic Peace: A Theoretical and Empirical Re-Assesment." *International Interactions* 23:000-000.

Maoz, Zeev & Bruce Russett. 1993. "Normative and Structural Causes of Democratic Peace, 1946-1986." *American Political Science Review* 87:624-638.

Morrow, James D. 1991. "Alliances and Asymmetry: An alternative to the Capability Aggregration Model of Alliances." *American Journal of Political Science* 35(4):904-33.

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

Oneal, John & James Lee Ray. 1996. "New Tests of the Democratic Peace: Controlling for Economic Interdependence, 1950-1985." Paper presented at the Annual Metting of the International Studies Association, San Diego, California.

Oneal, John R. & Bruce Russett. 1997. "The Classical Liberals Were Right: Democracy, Interdependance, and Conflict, 1950-85." *Internation Studies Quaterly* 41(2):267-294.

Oneal, John R. Francis H. Oneal, Zeev Maoz & Bruce Russett. 1996. "The Liberal Peace: Interdependance, Democracy and International Conflict." *Journal of Peace Research* 33(1):11-29.

Raknerud, Arvid & Havard Hegre. 1997. "The Hazard of War: The Reassessing Evidence of the Democratic Peace." *Journal of Peace Research* 34(4):000-000.

Ray, Lee James. 1995. *Democracy and Internation Conflict: An Evaluation of the Democratic Peace Proposition.* Columbia, S.C.: University of South Carolina Press.

Russett Bruce. 1993. *Grasping the Democratic Peace.* New Jersey: Princeton University Press.

Spiro, David. 1994. "The Insignificance of Liberal Peace." *Internation Security* 19:50-86.

Stimson, James. 1985. "Regression in Space and Time: A Statistical Essay." *American Journal of Political Science* 29:914-947.

Sueyoshi, Glenn T. 1995. "A Class of Binary Response Models for Grouped Duration Data." *Journal of Applied Econometrics* 10(4):411-31.

Vasquez, John A. 1993. *The War Puzzel*. First ed. Cambridge: Cambridge University Press.

Werner Suzanne. 1995. "Conflict or Cooperation: Positive and Negative Dependane in Dyadic Relations." Paper presented at the Annual Meeting of the Peace Science Society, Columbus, Ohio.

White, Halbert. 1980. "A Heteroskedaticity-consistent Covariance Matrix and a Direct Test for Heteroskedaticity." *Econometrica* 48:817-38.

Notes:

Note 1: Other literatures examine ordered or unordered polychotomous dependent variables (i.e., conflict escalation, war outcomes, forms of engagement, etc.). The methods discussed here might be useful for such models, but we do not discuss these applications. We are more certain that our warnings for binary dependent variables apply to these more complicated cases. Back.

Note 2: During the last few years IR researchers have produced at least 40 BTSCS books, articles, and convention papers. Some prominent examples include: Barbieri (1996), Bennet (1996), Bremer (1996), Enterline (1996,1997), Farber & Gowa (1995,1997), Gartzke (1998), Gleditsch & Hegre (1995), Henderson (1997), Hensel (1996), Hermann & Kegley (1996), Huth (1996), Huth, Bennet & Gelpi (1992), Huth & Russet (1993), Krain (1997), Lemke & Reed (1996), Leeds & Davis (1996), Mansfield & Snyder (1996,1997), Maoz (1996,1997), Morrow (1991), Mousseau (1997), Oneal et al. (1996), Oneal & Ray (1996), Oneal & Russett (1997), Russet (1993), and Werner (1995). Back.

Note 3: The recent IR literature (Bremer 1992, Bueno de Mesquita & Lalman 1992, Diehl & Goertz 1993, Huth, Gelpi & Bennet 1993, Maoz & Russet 1993, Vasquez 1993) has favored the dyad-year as the unit-of-analysis. We use "dyad" as a generic for unit, with almost all of our arguments and methods applying equally well to monadic analyses which use the nation-year as the unit of analysis. Back.

Note 4: All uses of the word "independent" should be read as "conditionally independent," with conditioning being on other variables in the model. Back.

Note 5: Bennet (1996), Bremer (1993), and Raknerud & Hegre (1997) are excepetions. Back.

Note 6: See Hsaio (1986) for an overview of the TSCS issues. Stimon (1985) provides an excellent overview of these issues from a political science perspective. Back.

Note 7: The distinction between fixed and sampled units can be elusive. IR TSCS data sets typically cover a long time-frame, with 50 years being common, and hundreds of years not being unknown. We formalize some distinctions in the next section. Back.

Note 8: There is generally no theoretical reason to prefer one to the other. Most analysts use logit because it is a bit simpler, and we follow convention here. There are cases where the probit is easier to work with because the multivariate normal has some very nice and well worked out properties. Where convenient we work with the probit. Since most analyses are logit, we will use logit in the text as the generic binary variable estimation method. Back.

Note 9: If annual data and wars persist for more than one year, then Equation 3 would be a good model for years with ongoing wars. Back.

Note 10: This can be seen in the context of Heckman's work on employment. True state dependence asserts that having a job makes one more likely to have a job in the future, regardless of one's latent employability. Spurious state dependence says that this observation is an artifact of the latent employability variable remaining relatively stable. If there is true state dependence, then we can reduce unemployment by simply getting people an initial job; this effort will fail if state dependence is spurious. Zeros following zeros in areas with low propensity for earthquake (strain) is another example of spurious state dependence. Thus residents in areas of high strain should hardly take comfort in not having already had an earthquake! Back.

Note 11: Any continuous dependent variable time series specification can be incorporated in the logit framework by simply using the chosen specification for the latent y^*. Thus, we could easily allow the latent y^* to be modelled as an error correction process; however, it would not solve any of the problems discussed here. Back.

Note 12: This assumes that the error process is a first-order moving average. We could easily allow for higher-order processes, although we expect the simple first-order processes to suffice for yearly data. While autoregressive errors are usually used in continuous TSCS models because they are much simpler to estimate, theoreticians generally using moving average errors. It is shocks, the ν, that we expect to persist. Where theory suggests otherwise, however, analysts should choose the theoretically appropriate error model. Back.

Note 13: A good introduction to the basic mechanics of event history analysis can be found in Allison (1984). Back.

Note 14: An exception is a recent unpublished paper by Raknerud & Hegre (1997) where they analyze BTSCS data on the democratic peace using Cox's (1975) partial likelihood method. We note that they use the event history approach in order to retain information about the exact duration of any peace; the BTSCS approach only allows for peace durations to be known to within the accuracy of a year (or whatever time unit characterizes the BTSCS data). For the conflict data that Raknerud & Hegre use, it is possible to get finer than annual dating. Our interest is in using event history ideas to deal with the temporal dependence of BTSCS data and do we do not contrast our work with that of Raknerud & Hegre. But their approach is the closest to ours of any of the IR conflict studies. Back.

Note 15: Survival, BTSCS and event count models are what Flemming & Harrington (1991) call "counting processes." Survival models record the time until the count reaches one and BTSCS (and discrete time survival) models note whether a count has changed in any year. Back.

Note 16: Event history analysts call the conditional probability of war breaking out in a year given peace until that year the "discrete hazard" of war. Thus BTSCS analysts are estimating discrete hazards. Back.

Note 17: See Sueyoshi (1995) for a full discussion of discrete time duration models. Back.

Note 18: There have been several discrete event history analyses of the passage of legislation in the American states (e.g., Berry & Berry 1990). These interesting studies end up doing logit or probit analysis, where duration independence is assumed. Back.

Note 19: Thus, in continuous time the most common assumption is that of proportional hazards, where $h(\tau, \bar{x}) = h_1(\tau)h_2(\bar{x})$. While the models we propose are not proportional in this regard, they do assume that the effects of time and the covariates are separable. Back.

Note 20: A cubic spline first breaks up the time interval into a series of subintervals; the breakpoints are the interior knots. A cubic equation is fit for each subinterval, subject to the constraint that the spline joins up (is continuous) at each knot, and is also smooth at the knots (the right and left first and second derivatives are the same at each knot). Thus splines are just complicated parameterizations of the relationship between time (or any other variable) and the latent y^*. Fully parametric, "natural splines" require the analyst to specify the placement of the knots. Semi-parametric smoothing splines trade off the fit of the spline against smoothness, measured by the integral of the second derivative of the spline. This tradeoff is controlled by the analyst. While the degree of smoothing is somewhat arbitrary, in practice it matters little (and we are less interested in precise estimates of *s* than in its overall shape, particularly, whether it might be constant over time). We use the "generalized additive models" provided by **SPLUS** to estimate the smoothing splines because we use other features of the generalized additive model for related analyses on the causes of war. If we were limiting ourselves to BTSCS analysis we might choose a simpler fully parametric natural, or even linear, spline, or low-order polynomial. There is probably little loss in most cases of using a computationally simpler fully parametric method, so long as it provides for a flexible model of *s*. The parametric splines can be estimated with commonly available software, such as **STATA**. Back.

Note 21: A critique of conflict results using BTSCS data, although not explicit, can be found in Ray (1995) and Spiro (1994). Back.

Note 22: Compare the results of a regression with a logit on the sign of the dependent variable. Both analyses have the same sample size, but the logit results will have much higher standard errors. This increase comes not through sample size, but through the Fisher information matrix. Anderson & Gill's (1982) setup of the Cox survival model as a "counting process" also makes the informational equivalence of the survival and BTSCS approaches obvious. Back.

Note 23: Panel analysts do asymptotics as $N \rightarrow \infty$. Thus fixed effects lead asymptotically to an "infinite" parameters problem. Back.

Note 24: It is important to note that, as mentioned in the introduction, the methods employed by these scholars are endemic to quantitative IR. Back.

Note 25: Exact details may be found in OOMR, where they analyze a variety of specifications, of which we reanalyze only their major one. We have examined their other specifications, always finding them temporally dependent. We note that very recently Oneal & Russett (1997) have revised their measures of democracy and trade interdependence. Since we have not yet had a chance to either obtain or recreate the new measures, for this paper we work with the data that corresponds to the original published article. We do not use the flawed JOINREG measure of regime type that was also analyzed by OOMR. Back.

Note 26: Results were not sensitive to the degree of smoothing. Back.

Note 27: The vertical axis in this figure is in the units of $\bar{x}_{i,t}\beta$ in Equation 2a. Thus, for example, the effect on the probability of a dyad entering a militarized dispute after having been at peace for eight years is slightly greater than the effect of the dyad being democratic. Eight years of peace decrease the latent propensity for dispute by about two; dyadic democracy decreases this propensity by 1.73. To compute probabilities the latent must be transformed by a logit transformation, so the effect of PEACEYRS depends non-linearly on the other dyadic characteristics. Back.

Note 28: See Bennett (1996) for all details. Back.

Note 29: We are puzzled by Bennett's (1996) claim that an "analysis using a Weibull distribution as a check on model specification yields conclusions regarding the substantive effects of the variables similar to those [he] presented." (p. 169) The results in Table 3 for the Weibull model are rather different from those for the exponential. Back.

Note 30: Note that the signs of the coefficients in Tables 2 and 3 are opposite, since one table presents models on the hazard of exit while the other presents models on the length of time until exit. The Cox regression coefficients were reversed in sign to make them consistent with the other models of Table 3. Back.

Note 31: We do not interpret the democratic peace hypothesis as asserting that democracies fight shorter wars. There were 612 dyad-years of militarized disputes in the full Oneal et al. data set. We drop 250 of these observations as representing ongoing disputes Back.

Note 32: This result should not be surprising. Given the rarity of conflict, and the relatively high proportion of conflicts that last over one year, the estimated probability of a dispute in years with a zero score for PEACEYRS clearly overstates the probability of a *new* war breaking out immediately after the settlement of an old one. Back.

Note 33: As before, we reverse the signs of the Cox coefficients for ease of comparison. The Weibull estimates are not strictly comparable to the other estimates since the effect of each independent variable on the end of peace is a function of its own estimated coefficient and the estimation of duration dependence, σ. Back.

Note 34: Why not simply use duration methods for BTSCS data? Duration methods work well if there are long spells of zeros between ones. However, our proposed method does not require this. Our proposed method is also more similar to the method of analysis now used, logit analysis. In addition, the Cox technique is sensitive to the number of tied durations and BTSCS data exhibit many ties. The Weibull technique does not suffer from this problem, but it does make strong assumptions about the nature of duration dependence, stronger than those made for our proposed method. But for some BTSCS data it may well be the case that traditional event history methods are best. We should see the two approaches as complementary, not in competition. Back.

Note 35: These were computed with **STATA**. Back.

Note 36: Our reanalysis of Bennett's data findings using robust standard errors is consistent with the findings obtained from his original analysis. (These results, and others not reported, are available from the authors). Briefly, the robust standard errors reduce *P*-values for SECURITY and COMMON THREAT, making the former marginally insignificant and the latter only marginally significant. Interestingly, the Huber standard error for ISSUE SALIENCE is lower than its maximum likelihood counterpart. This is yet another signal that the independent analysis understates the role of issue salience. We also estimated Huber standard errors for Huth's rivalry data. The robust errors for that analysis were within 10% of the naive standard errors. Thus, the Huth, et al. findings do not suffer from whatever pathologies the Huber technique guards against. Back.

Note 37: A better strategy would be to assume that all the dyads are similar and model them with the same "miscellaneous" fixed effect (i.e., model them with the same dummy variable). An alternative would be to find some commonality in subsets of these dyads and use several dummy variables. Back.

Note 38: We re-estimated the OOMR model using fixed effects and robust standard errors with duration splines. This reanalysis reinforced the conclusions of our other reanalyses. Back.
