

Real Time, Time Series Forecasting of Political Conflict

Patrick T. Brandt
School of Economic, Political and Policy Science
University of Texas, Dallas
pbrandt@utdallas.edu

John R. Freeman
Department of Political Science
University of Minnesota
freeman@umn.edu

Philip A. Schrodtt
Department of Political Science
University of Kansas
schrodtt@ku.edu

February 10, 2009

Abstract

We propose a framework for forecasting and analyzing regional and international conflicts in real time. The proposed framework would generate forecasts that 1) are accurate, 2) are produced in (near) real time, 3) address the actions of multiple actors in a conflict, 4) incorporate prior beliefs about conflict processes, and 5) allow us to generate policy contingent forecasts. We propose to meet these desiderata by combining the CAMEO event coding framework with Bayesian vector autoregression (BVAR) and Markov-switching Bayesian vector autoregression (MS-BVAR) forecasting models. We outline an example using these methods and present some preliminary conclusions.

1 Introduction

Scholars and policy makers want to anticipate intra- and international conflicts. They also want to evaluate what might have occurred if certain actions had been taken in the past and (or) what might happen if governments take certain actions in a given conflict in the future. To this end, they developed tools for analyzing state failure, political instability, rebellion, repression and civil war. For their part, organizations like the international crisis group, ICG (www.crisisgroup.org) publishes weekly its “Crisis Watch” to help policymakers anticipate and hopefully mediate conflicts world-wide. The Swiss Agency for Development and Cooperation funded a similar effort in recent years (FAST; www.swisspeace.org).

An ideal forecasting tool would have at least five interrelated features:

1. It would produce accurate forecasts. This desideratum sounds straightforward. The tool makes a prediction; the prediction either is realized or not. In fact accuracy is a more complex concept. Forecasts involve uncertainty. Even the simple forecasts from regression models contain forecast error, error derived from the random nature of the dependent variable and estimation uncertainty. A forecast therefore is better characterized by an interval rather than a point, more precisely, it is a probability distribution. As we explain below modern forecast evaluation now focuses on distribution prediction, not point prediction (Gneiting,

2008).¹

2. Useful forecasts need to be produced in real-time. A delay of weeks or months often is too late to anticipate (prevent) a conflict from occurring or escalating, and the postdictive analyses that characterize most academic studies of conflict are useful only for model development. And the forecast must be calibrated in time. Not knowing exactly when the predicted outcome will occur diminishes the usefulness of the forecast.
3. Scholars and policymakers are interested in the behavior of collections of belligerents in conflict systems. For instance, an ideal forecasting tool would allow us to anticipate both the actions of one belligerent towards its adversaries but also the reactions of those adversaries and the subsequent behavior of the original belligerent.
4. To improve its performance, the tool should incorporate our knowledge about intra and international conflict in general and about the history of particular cases. There is a wealth of expert knowledge about this subject (a “wisdom literature”). This knowledge teaches about strategic behavior, the propensity for conflicts to exhibit phase shifts (non-linearities) and other characteristics. We should be able easily to incorporate this knowledge into our forecasting tool.
5. The tool should enable us to make contingent forecasts. We should be able to evaluate counterfactuals about the past history of a conflict as well as to make forecasts contingent on hypothetical behaviors of the belligerents or third parties. For example, we would want to generate contingent forecasts of the impacts of Israeli and U.S. actions on the relations among Israel, Hamas, Fatah, and the U.S.

Existing tools for forecasting intra- and international conflict meet some but not all of these desiderata. Table 1 catalogs some of the most well known; space does not permit a complete review. So we focus on three.² The first is the tool developed by Bueno de Mesquita to predict single events, Expected Utility Approach (EUA). Its accuracy in making *ex ante* point predictions has been heralded for years both by its creator and by government policy makers (Feder 1995; Bueno de Mesquita 1997, 2002, and www.diiusa.com).³ Among EUA’s strengths are its incorporation of rational choice theory and its use of experts to estimate preferences. Contingent predictions can be made with it by altering the specification of the model and deriving an implied (contingent) forecast. Unfortunately this first tool falls short on the first, second, and third desiderata. It has no provision for measuring the forecast errors from estimation, the elicitation of the variables for each subject (capability, salience, intensity) or, the fact that subjects’ utility functions have random elements.⁴ EUA users do not attempt to produce or evaluate the probability distribution over the outcomes of interest. EUA’s forecasts have no time component. Bueno de Mesquita (1997, 264) indicates that predicted outcomes are expected to occur in “reasonable amount of time;” but

¹Forecast errors and intervals are explained in most intermediate regression books. Yet, many political scientists continue to base their model evaluations on point forecasts. See the recent issue of *PS* (61(4), October 2008) on competing forecasts of the 2008 presidential election.

²Some tools could be placed in multiple locations in Table 1. An example is prediction markets Wolfers and Zitzewitz (2004) describe how index contracts in these markets can be used to generate forecasts of mean levels of variables and conceivably of collections of outcomes.

³Recent publications call Bueno de Mesquita the “New Nostradamus”, see www.goodmagazine.com, April 17, 2008.

⁴There are few references in the EUA literature on the processes of elicitation. Elicitation necessarily gauges the assessors degree of certainty in the information she or he is providing the analyst. For a discussion and application of elicitation methods in political science see Freeman and Gill (2008, 2006).

Type	Ex Post	Ex Post Counter-factual	Ex Ante
One Shot Decision or Event	Game theoretic / TOM (Brams and Togman, 2000)		Rational Choice (Organski, 2000); EUA (Bueno de Mesquita, 2002); Expert Systems (Tetlock, 2005); Role Playing (Bennett and McQuade, 1996)
Probability of Binary Outcome	Failed States Project (Esty et al. 1998, King and Zeng 2001); Neural Net Models (Beck, King and Zeng, 2000); Integrated Crisis Warning Systems (ICEWS)	Political Instability Task Force	Prediction Markets (Wolfers and Zitzewitz, 2004)
Conflict System: Directed Dyadic Behavior	Cluster-Density Analysis (Schrodt and Gerner, 2000); BVAR (Brandt and Freeman, 2006); B-SVAR (Brandt, Colaresi and Freeman, 2008)	BVAR (Brandt and Freeman, 2006)	International Crisis Group (ICG); Swiss Peace Foundation (FAST); VAR (Pevehouse and Goldstein, 1999)
Behavioral Phase Shifts	Crisis Early Warning Systems (CEWS); Hidden Markov Models (Schrodt, 2000)		

Table 1: Some Approaches to Forecasting in International Relations

he acknowledges that in practice the EUA forecasts suffer from “off-on-time problems”.⁵ Finally, because EUA produces one-shot, static predictions, its usefulness in forecasting the behavior of whole conflict systems is limited. Also, the possibility that these systems may exhibit phase shifts which alter the parameterization of the EUA model appears not to have been addressed.

The Political Instability Task Force, (PITF; Bates et al. (2003)), the successor to the earlier State Failures Project (Esty et al., 1995; Esty, Goldstone, Gurr, Harff, Levy, Dabelko and Surko, 1998) is a second major contribution to forecasting. It produces probabilistic predictions of state failure and of other discrete outcomes over time horizons of several years. Most of these predictions are based on strongly restricted, single equation models—e.g., neural network models based on a rare event logit specification with independent variables chosen, in part, on the basis of theory and expert consultation. Most of the rigorous accuracy assessments in the open literature have been *ex post* in nature (King and Zeng, 2001), although there have probably been unpublished *ex ante* assessments inside the U.S. government.⁶

The PITF project falls short on three of the desiderata. To begin, forecast error is not incorporated in its accuracy assessments, at least in the published assessments.⁷ Second, the PITF models

⁵The related work of Organski and Lust-Okar (1997) suffers from the same problem. They note that their predicted outcome for the Jerusalem negotiations does not indicate when the final agreement would be reached (See esp., Organski and Lust-Okar, 1997, 350).

⁶King and Zeng’s contribution is a critique of the Failed States project. Their own contribution is a neural net model (Beck, King and Zeng, 2000). This model also uses strongly restricted, single equations of the nested logit type. Their demonstration of the accuracy of the neural net model also is *ex post* in nature.

⁷King and Zeng (2001) produce a more rigorous evaluation of the ability of the earlier, published State Failures

are highly aggregated in space and time. The results for the probability that particular states will fail are based on average effects and therefore subject to ecological inference problems. The use of country-year data means short-term risk assessments (weekly and monthly forecasts of state failure) are not available. Finally, its strong—usually untested—exogeneity assumptions ignore theoretical knowledge about the relationships between political and economic variables, like the endogenous relationships between democratization and certain forms of inequality.

Brandt and Freeman (2006) and Brandt, Colaresi and Freeman (2008) employ a third approach. Using data from the Kansas Events Data System (KEDS/TABARI), they show how Bayesian time series models can be used to analyze conflict dynamics and to make *ex post* forecasts in the Levant over the short-term. Their models are Bayesian vector autoregressions (BVAR) and Bayesian structural vector autoregressions (BSVAR). Their specifications use a modification—incorporating expert judgment of international relations scholars—of the Sims and Zha (1998) prior used in macroeconomic forecasting. The models produce *ex post* forecasts with error bands for weekly and monthly patterns of conflict and cooperation (directed dyadic behavior) of the Israelis, Palestinians and United States. In one of these analyses, the impact of Jewish public opinion is incorporated (*ibid.*). Counterfactual (*ex post*) forecasts are reported in both the pieces.

The work of Brandt et al. suffers from three problems. First, Brandt et al.’s accuracy assessments are based on the mean forecasts from their models; they rely on measures like residual mean square error (RMSE) of the respective point forecasts to evaluate their models. This is in spite of the fact that their approach produces—via computational simulation or Gibbs sampling—the full probability distribution for their forecasts. As we note below, Chatfield (2001) and others show that RMSE and related criteria are sensitive to outliers. Gneiting (2008) and other statisticians recommend using the full predictive density for forecast evaluation. In addition, all of Brandt et al.’s work is *ex post* in nature. They have not produced any *ex ante* forecasts with their Bayesian time series models.

We propose to build on the approach of Brandt et al. because it applies the advances in forecast evaluation from statistics, produces temporally disaggregated forecasts in real-time, allows for the analysis of whole conflict systems, and incorporates expert judgment in the form of Bayesian priors. Specifically, we propose a technology that accurately forecasts the weekly behavior of medium-sized conflict systems in real-time. In our larger-scale project we will employ events data series from KEDS/TABARI, based on real-time text feeds and automatic text parsing, on important intrastate and international conflicts. The choice of the forecasting model will be based on predictive integral transform methods and other new tools developed by statisticians. Provision will be made for nonlinearities (conflict phase shifts) by adding a Markov-switching component to the extant Bayesian time series models. This new technology will produce theoretically grounded, policy relevant forecasts, including contingent forecasts. All the results and tools will be available in the public domain. In this first stage of our project, real-time *ex ante* forecasts for three major conflict systems will be produced. More conflict systems will be included in future stages. Finally, we will actively monitor and compare our forecasting methods to other statistical forecasting models.

2 Research Design

Our design consists of four parts: production of event time series for selected cases, incorporation of statistical advances on forecast evaluation and phase shift analysis, comparison of our forecasting models to others, and website design and management. We explain these parts in the next sections. Figure 1 outlines our overall data generation, model building, and forecasting approach.

model and their neural net model to predict one-shot events. This evaluation uses ROC curves.

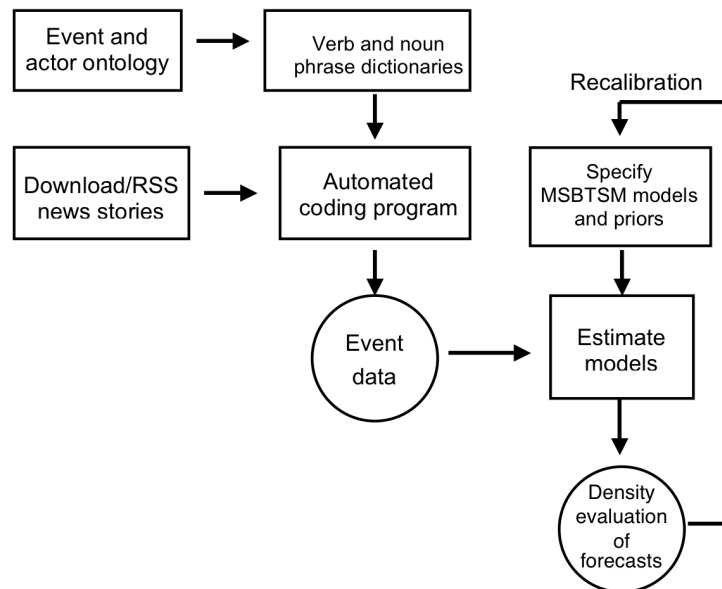


Figure 1: Data generation, model specification and forecasting

2.1 Data and Case Selection

Event data will be used in our project. These are data on the actions between two parties to a conflict, nominal codes indicating who did what to whom and when. These data will be produced by the state-of-the-art TABARI automated coding software⁸ using the CAMEO event and actor coding system⁹. TABARI can analyze text that is downloaded from data services such as Lexis-Nexis or Factiva, or in real-time—at for essentially no cost beyond that of a web connection and web-enabled computer—via real simple syndication feeds (RSS). RSS aggregators such as Google News¹⁰ and the European Media Monitor¹¹ each monitor over 4,000 sources, including both international media sources such as the *New York Times*, *Christian Science Monitor*, *BBC*, *Agence France Press*, and *Reuters*, as well as thousands of local sources. The availability of these aggregators, and RSS feeds more generally, provide an unprecedented ability to monitor global political behavior in real-time.

The automated coding program TABARI works in conjunction with the CAMEO event coding system, which has extended the scope of the Cold War-era WEIS and COPDAB inter-state coding systems to provide detailed coding of intra-state conflict and to incorporate contemporary, post-Cold War theoretical concerns and concepts. TABARI also provides a detailed and systematic framework for coding a wide variety of political actors, whether international, supranational, transnational, or internal. From the coded text we can extract interval measures of inter-state and inter-group cooperation and conflict, count data on various types of behaviors, and nominal data such as event sequences or patterns. Because text data are available and updated frequently, and TABARI’s coding is fully automated, we will have a continuous flow of information on a wide range of variables and actors. Because it is informed by social science theory in how it catalogs actors and actions via dictionaries, the TABARI/CAMEO routines will be much more efficient and accurate than routines used by unsupervised data miners or by manual coding.

Retrospective automated event data coding from a single source (typically *Reuters* or *Agence France Presse*) is now a well-established procedure, and almost all studies of political conflict in refereed political science literature over the past fifteen years have used this type of data. The

⁸<http://web.ku.edu/keds/software.dir/tabari.html>

⁹<http://web.ku.edu/keds/data.dir/cameo.html>

¹⁰<http://news.google.com/>

¹¹<http://emm.jrc.it/overview.html>

challenge is moving this to multiple sources in real-time. Specifically the three key challenges are

1. Automated detection and classification of new political actors, as well as the re-classification of existing actors (e.g., a political party switching from government to opposition). To a much lesser extent, event coding systems need to be updated as new political vocabulary—such as the emergence of the phrase “ethnic cleansing” in the early 1990s—comes into use.
2. Effective filtering of historical and non-political stories (e.g., sports and business stories that use military metaphors) from a multiple-source stream, as well as the efficient detection of duplicate reports of the same event.
3. Calibration of event reports across multiple sources and regions: it is well known that some areas of the world are covered less extensively than others (e.g., Africa is covered less than East Asia), and as one shifts to multiple news courses, editorial policies such as the mix between political and economic stories also affect the data. Based on some recent experiments, item response theory (IRT) and other scaling methods appear to have great potential in resolving this issue (Schrodt, 2007). But additional work is required to apply them.

In addition to these specific issues, we also will deal with some general issues involved in scaling conflict systems, and some additional refinements to the CAMEO coding systems as we take that out of the Middle East and Asia—the geographical domains where most of the development has been done—into other parts of the world. The open-source TABARI/CAMEO system recently was employed by a defense contractor to code 25 gigabytes of Asian news reports involving more than 6.7 million stories and 253 million lines of text from 70 news sources, so we are fairly confident that the system is sufficiently robust to be able to generate the data that we need.

In this, the first stage of project, we will focus on three cases. The first is the Levant, specifically the Israel-Hamas-Fatah and Israel-Lebanon-Hizbollah cases. The KEDS project has a retrospective dataset on this conflict that provides thirty years of coverage¹², more than ample data for estimating and assessing retrospective models. Furthermore, the news coverage of the Levant is continuous and dense. Schrodt and his collaborators have used this data in several retrospective predictive studies (e.g., Hudson, Schrodt and Whitmer, 2008; Schrodt, 2006; Schrodt and Gerner, 2004; Schrodt, 2000), and Schrodt also has done field work in the region. Brandt and Freeman already have applied some Bayesian time series models to this case *ex post* (Brandt and Freeman, 2006; Brandt, Colaresi and Freeman, 2008). Finally there are a number of competing forecasts to which we can compare the success of our technology. Note however that while we have extensive experience with the collection and analysis of these, our collective past analyses of KEDS data for the Levant have *not* produced real-time forecast densities like what we are proposing here. Our goal is to extend this to real-time forecast density estimation.

In addition, to demonstrate the generality of our new approach, we also will forecast China-Taiwan-US relations and India-Pakistan-US relations. As just noted, TABARI/CAMEO very recently has been used to analyze a large body of Asian news text. So we have experience with the respective data sources.

¹²<http://web.ku.edu/keds/data.dir/levant.html>

2.2 Bayesian Models for Forecasting and Modeling Nonlinearities in International Relations

2.2.1 Bayesian Forecasting in International Relations

To forecast systems of behavior in time we need a multi-equation time series model. Consider the following:

$$\sum_{l=0}^p y_{t-l} A_l = d + \epsilon_t, \quad t = 1, 2, \dots, T \quad (1)$$

where y_t is a $1 \times m$ dimensional vector say for the directed dyadic behavior of a three actor system, A_l is a $m \times m$ matrix of coefficients for a set of lag polynomials representing the memory in the system back to lag p , d is a $1 \times m$ vector of constants or deterministic variables such as electoral calendar counters, and ϵ_t is a $1 \times m$ dimensional error vector assumed to be uncorrelated with the y_{t-s} for all $s > 0$, serially uncorrelated, and with individual variances equal to unity. The reduced form of equation 1 is

$$y_t = c + y_{t-1} B_1 + \dots + y_{t-p} B_p + u_t, \quad t = 1, 2, \dots, T \quad (2)$$

where y_t again is an $1 \times m$ vector of observations at time t , B_l is the $m \times m$ coefficient matrix for the l^{th} lag, p is the maximum number of lags, and u_t are the reduced form residuals.¹³ Such reduced form models have been used by scholars to study conflict dynamics in the Levant and Bosnia (e.g. Goldstein, Pevehouse, Gerner and Telhami, 2001; Pevehouse and Goldstein, 1999; Goldstein and Pevehouse, 1997), as well as to study policy counterfactuals (Goldstein and Freeman, 1990, 1991). But to our knowledge they have not been used to produce *ex ante* statistical forecasts.¹⁴

There are at least two major problems with such models. First they are overparameterized. For a four actor illustration (e.g., the U.S., Israel, Hamas, Fatah) y_t would contain the $m = 12$ directed, dyadic behaviors in the conflict system. If the VAR representation of this conflict uses $p = 6$ lagged values and d is a vector of constants, the VAR model has 876 parameters. Clearly this produces a tremendous amount of estimation uncertainty and therefore a lack of precision in the model's forecasts.¹⁵ Overparameterization also can produce mistaken inferences about persistence, more specifically, about stationarity (Sims and Zha, 1998). But knife-edge tests for non-stationarity (unit roots) also can be erroneous. These tests can produce pretest bias in multivariate time series analysis (Freeman et al., 1998).

Bayesian multivariate time series analysis addresses these and other problems. By using expert knowledge in a Bayesian prior on the coefficients in equations 1 and 2, we are able to reduce the estimation uncertainty and obtain more precise dynamic inferences and forecasts. This prior applies to the weight of behavior at distant lags and the degree to which we believe the series might be non-stationary and possibly cointegrated. In the Bayesian model these restrictions are

¹³The reduced form in equation (2) is related to the simultaneous equation model in equation (1) by

$$c = dA_0^{-1}, B_l = -A_l A_0^{-1}, l = 1, 2, \dots, p, u_t = \epsilon_t A_0^{-1}.$$

¹⁴For example, Pevehouse and Goldstein's (1999) article, "Serbian Defiance or Compliance in Kosovo? Statistical Analysis and Real-Time Predictions" is based solely on causality tests in a reduced form model of the respective conflict.

¹⁵Perhaps this is why Goldstein and Pevehouse and Goldstein et al. do not produce any impulse responses or forecasts for their models (they just report causality test results). Goldstein and Freeman (1990, Chapter 5) also do not produce confidence intervals (error bands) for their policy counterfactuals.

inexact whereas in the frequentist model they are exact and knife-edge. Brandt and Freeman (2006, forthcoming) explain how a prior for the study of intra- and international conflict (cooperation) can be constructed and applied to equations 1 and 2 (see also Brandt, Colaresi and Freeman, 2008). This work also explains how concepts like log marginal data densities Chib (1995) can be used to calculate Bayes factors (Kass and Raftery, 1995) to assess the in-sample fit of these Bayesian time series models. Finally, all these developments are based on posterior Markov chain Monte Carlo (MCMC) samples. So the technology needed to produce predictive distributions for these models is available.¹⁶

2.2.2 Markov-switching Bayesian time series models

Our theories (for example Lebovic, 1994; Lund, 1996; Gurr and Harff, 1996) teach us that human conflicts are subject to phase shifts or fundamental non-linearities. This is another reason why normality assumptions break down in time series forecasting: non-linearities caused by conflict phase shifts are bound to produce fat tailed (leptokuric) forecasting distributions. Unlike financial analysts who model this behavior with “heavy tail distributions,” we propose to model the non-linearities with Markov-switching processes. These models allow us to capture the phase changes and non-normality of the data using Markov and mixture models (Frühwirth-Schnatter, 2006). Specifically, we will include in our evaluation of contending forecasting models, Markov-switching, Bayesian multivariate time series models.¹⁷

A Markov-switching version of the Bayesian VAR model (in equations 1 and 2) can be represented as a set of h state VAR models. The structural Markov-switching Bayesian VAR (MS-BVAR) version is

$$\sum_{l=0}^p y_{t-l} A_l(s_t) = d(s_t) + \epsilon_t(s_t), \quad t = 1, 2, \dots, T, \quad (3)$$

where $s_t = j$ is an h -dimensional vector indexing the state of the process where j is the integer label for the state, with a $h \times h$ Markov transition matrix Q . The rows of Q give the probability of transitioning from state s_{t-1} to s_t , or $\Pr(s_t = k | s_{t-1} = j)$ for states k and j . Note that while the models in equations (1) and (2) generically have $m^2p + m$ regression parameters, the model in equation (3) has $h(m^2p + m)$ parameters. As noted earlier, for a 12 equation, 6 lag model the model in equation (2) has 876 parameters. So if we allow a Markov-switching process with 2 states (say high and low conflict intensity / volatility) there will be 1752 parameters.¹⁸ So a Bayesian approach is even more sensible when allowing for non-linearities and phase shifts, since it allows us to deal with the proliferation of parameters.

A prior for this model has to include beliefs about the A_l and d parameters as well as the Markov process Q . We already have been able to adapt the random walk prior in Sims, Waggoner and Zha (2008); Sims and Zha (1998) to serve as a baseline set of beliefs in an analysis of international conflict event data (Brandt and Appleby, 2007). The sensitivity of this prior can be evaluated using methods like those discussed in Brandt and Freeman (forthcoming).

¹⁶Brandt has implemented these models in an open source R package called MSBVAR, <http://yule.utdallas.edu/code.html>

¹⁷On the prevalence of fat tails in forecast distributions in economics and finance see such works as Granger (2005). Note that our models will allow us to capture more permanent structural breaks in conflict as well. These breaks are simply switches with high degrees of permanence.

¹⁸In fact, the Bayesian estimation of these models is even more demanding, since the model in equation 3 has $h(m^2p + m)$ parameters spread over $2^m (h!)$ possible posterior modes. Sampling from these modes is computationally intensive but feasible (Frühwirth-Schnatter, 2001; Scott, 2002). Our plan is to implement the multivariate MCMC permutation samplers necessary for estimation and forecasting using the Teragrid.

Forecasting with these models is a straightforward application of data augmentation via Gibbs sampling. So we can generate dyadic, phase or state specific *ex ante* and *ex post* counterfactual forecasts. The MCMC output of the Bayesian model in equation (3) also can be used to estimate the number of states or phases in the model (Sims, Waggoner and Zha, 2008; Frühwirth-Schnatter, 2006). We will also be able to evaluate these forecasts in the manner discussed below: forecast densities that account for the possible state-dependent mixtures of forecast distributions. This allows us to capture both fat-tailed and other non-linearities in a sensible manner.¹⁹ Finally, the forecasts from a Markov-switching VAR model like equation (3) are a weighted combination of the forecasts for each state or phase. This means that the forecasts account for both the inherent variability of the time series themselves, but also the parameters and the phases. Thus, the uncertainty of changepoints or phase shifts is captured directly in the estimation and forecasting process enabling additional analyses of how regime changes impact the ability to predict conflict (cooperation).

2.3 Forecast Densities and Evaluations: Making the Transition from Point to Distribution Prediction

How then do we build the most *scientifically sound and useful* Bayesian time series model for forecasting intra and international conflict (cooperation) out-of-sample (*ex ante*) in real-time? The answer is that we need to make the transition from point prediction to distribution prediction and to apply the technique of “recalibration” (Gneiting, 2008; Timmerman, 2000; Tay and Wallis, 2000). Statistical forecasting in political science still uses point prediction to evaluate model performance. In fact, we usually use a residual mean square error (RMSE) criterion for this purpose. First, there are many such criteria for point forecasts (Wallis, 1995; Diebold and Lopez, 1996; Chatfield, 2001). And RMSE is one of the weakest criteria because it is sensitive to outliers. But more important, point prediction does not tell us anything about the (un)certainty of a forecast. And reliance on normality assumptions to construct forecast confidence intervals is prone to mistaken inferences since they amount to strong, often inaccurate assumptions about the symmetry and about the size of the tails of a forecast distribution.²⁰

One alternative is to use prediction intervals. Put simply, this is the practice of constructing the upper and lower limits between which the future values of a series are expected to lie with some prescribed probability (Chatfield, 2001; Granger, 2005). Tools like fan charts are used to represent these intervals. Since the late 1990s the Bank of England has presented its forecasts for inflation and growth in this way (Gneiting, 2008). For various reasons these intervals are often difficult to apply. One is that until recently many software packages did not produce them. More fundamentally, they are apt to be too narrow on average (Chatfield, 2001, 479ff.).²¹

The favored approach in statistics is density evaluation, or what is called “probability forecasting” (Rosenblatt, 1952; Dawid, 1984). A density forecast is “a complete description of the uncertainty associated with a prediction, and stands in contrast to a point forecast, which by itself, contains no description of the associated uncertainty” (Tay and Wallis, 2000, 235). *Our plan then is to use our Markov-switching Bayesian multivariate time series model to produce entire forecast densities for all our variables for conflict and cooperation between belligerents and to evaluate these densities for their accuracy through time. On the basis of the results we will recalibrate our model*

¹⁹The MS-BVAR model outlined here can capture changepoint, SETAR, TAR, and other switching and non-linear time series models as special cases (Frühwirth-Schnatter, 2006; Krolzig, 1997).

²⁰For a succinct summary and critique of these normality assumptions see Tay and Wallis (2000, esp. 235-236).

²¹Space does not allow for a full presentation of interval forecasting or for a discussion of the methods that have been developed to evaluate interval forecasts. (See Christoffersen, 1998; Taylor, 1999; Tay and Wallis, 2000).

gradually improving its performance. We will make all of the data forecasts publicly available via a project website. We will also compare our forecasts to those generated by other methods and parties such as the recently announced political forecasting website—The Call—by the Eurasia Group in collaboration with *Foreign Policy* magazine (<http://eurasia.foreignpolicy.com/node>) and any additional such sites that come on-line during the period of the grant.

Several tools are used in this process. The primary tool is the probability integral transform (PIT).²² The PIT is defined for a series of realized values and forecasts. Let $\{y_t\}_{t=1}^m$ be a series of realizations from some conditional density $\{f(y_t|\Omega_{t-1})\}_{t=1}^m$ where Ω_{t-1} is the (past) information set at time t . Then if the one-step ahead density forecasts $\{p_{t-1}(y_t)\}_{t=1}^m$ equal $\{f(y_t|\Omega_{t-1})\}_{t=1}^m$ (and assuming a non-zero Jacobian), the PIT of $\{y_t\}_{t=1}^m$ with respect to $\{p_{t-1}(y_t)\}_{t=1}^m$ is i.i.d. $U(0, 1)$, or

$$\{z_t\}_{t=1}^m = \left\{ \int_{-\infty}^{y_t} p_{t-1}(u) \partial u \right\}_{t=1}^m \sim U(0, 1) \quad (4)$$

If the forecast density is ‘correct,’ then the z_t is an independent uniform $U[0, 1]$ variate and the model is “well-calibrated.” This result extends to the case of time-dependent density forecasts and to multivariate forecasts. Deviations in z_t from uniform iid are indications that the forecasts do not capture some aspect of the data generating process. One uses graphical tools—comparison of the plot of the z_t with the 45 degree line that is the uniform’s distribution function, histograms, Kolmogorov-Smirnov and Cramer-von Mises tests, and correlograms of the levels and powers of the PIT to make this determination. Still newer tools that complement PIT histograms are calibration plots and sharpness diagrams.²³ Using these tools we will be able to rank the performance of alternate forecasting models. By recalibrating the best of them, we will produce better forecasts of the Levant, South Asia and East Asian cases. Finally, if forecasting models can be found that capture the true data generating process, from a decision theoretic perspective, all analysts will prefer it regardless of their loss functions (Diebold, Gunther and Tay, 1998; Diebold, Tay and Wallis, 1998). Hence our best models will aid decisionmaking.

In the first stage of our project, we will evaluate competing forecasting models for intra- and international conflict (cooperation) for our three cases and evaluate them with the PIT and the tools mentioned above. The competing models will include theoretically relevant actor clusters and variables like election forces. To benchmark their performance we will include naïve models like univariate autoregressive moving average specifications. But Bayesian reduced form models like that in equation (2) and its Markov-switching variant in equation (3) will be the main objects of evaluation. Therefore we will need to apply multivariate techniques for density forecasting (Kling and Bessler, 1989; Diebold, Hahn and Tay, 1999; Clements and Smith, 2000). Thus we will produce forecasts that meet our earlier desiderata.²⁴ As with any scientific enterprise, we expect to learn as much from where our models go wrong as from where they are correct. Erroneous predictions are key to determining the gaps and weaknesses in our existing theoretical concepts, variables, and measurement methods. The advantage of systematic forecasting such as the system we are proposing is that we can re-assess the models based on the sets of assumptions based on different theoretical perspectives, particularly as these feed into the model through the priors.

²²The remainder of this paragraph’s definition of the PIT is taken from Diebold, Hahn and Tay (1999, 661–662). See Diebold, Gunther and Tay (1998, 867–868) for a proof of why the PIT has the properties reported here.

²³In a very recent piece Gneiting, Balabdaoui and Raftery (2008) review work (Hamill, 2001) that shows the PIT is only necessary not sufficient for forecasts to be ideal. They propose a “paradigm of maximizing sharpness of the predictive distribution subject to calibration.” The tools of calibration plots and sharpness diagrams are proposed and illustrated in their article. We hope eventually to apply these most advanced tools in our project.

²⁴For example, Kling and Bessler (1989) illustrate how a Bayesian multivariate model with a Litterman prior can be recalibrated in time to produce better forecasts of four variables for the U.S. macroeconomy.

With the high-speed automated coding provided by the TABARI system, we can also experiment with the theoretical constructs embedded in the event coding ontologies. For example by further differentiating sub-state actors, or differentiating types of events that are currently lumped together into a single coding category, but which can be differentiated in the texts from which they were originally derived, we can determine how definitions of actors and events affect our ability to produce high quality forecasts.

3 An Illustration for the Levant

To illustrate our approach, we generate a set of *ex ante* forecasts for material conflict among the Israelis and Palestinians. This is only an example of what we propose here: we plan to expand this example to include more equations (actors), mediation and cooperation event series, less aggregated data (weekly instead of monthly), and more complex dynamics. With these caveats, the goal here is to point out the advantages of MS-BVAR and BVAR models for forecasting conflict processes.

The case study for this analysis is data on the Israeli-Palestinian conflict. We employ monthly CAMEO-coded data from 1996:1-2008:12. We aggregated the data for U.S., Israeli and Palestinian *material conflict events* for each month.²⁵ We use these data because they capture the second intifada and periods of recent high and low conflict. For the six dyadic material conflict series (ISR2PAL, PAL2ISR, USA2ISR, ISR2USA, PAL2USA, USA2PAL) we fit two models. The first is a Bayesian VAR with one lag using the prior specified in Brandt and Freeman (2006). The second is an MS-BVAR model with one lag where the prior for each regime or state is the same symmetric prior used in the BVAR model. The diffuse prior for the regimes is Dirichlet.²⁶

Figure 2 shows the probabilities of the two regime MS-BVAR model for the sample. Note that the high conflict periods (the dashed or red regimes) are those that correspond to recent episodes of very high regional material conflict. The regime switches come at clearly defined points in the conflict: October 2000 (the outbreak of the second intifada after Sharon’s visit to the Temple Mount on September 28, 2000), March 2002 (Hamas suicide bombings in Natanya and Arab Peace Plan proposed), October 2003 (Islamic Jihad bombings in Haifa with Israeli retaliatory bombing in Syria). Thus, we see that there is evidence for changes in parameters throughout the recent years of the conflict. The first regime is one where the conflict stays near a *status quo* level. The second regime is rare and one where the conflict dynamics are nearly non-stationary, have an equilibrium level of twice as many material conflict events than the lower conflict regime, and a variance that is about 50% larger than the lower conflict regime.

Most critically, and in spite of recent events in the Israeli-Palestinian conflict, current state of the conflict is the low conflict regime. Figure 3 presents 12 months of forecasts (covering all of 2009) for the four material conflict series. Note that these are a weighted combination (based on the probabilities of each regime, forecast 12 months into the future). Thus, this is a mixture distribution of the high and low conflict regimes. An important caveat to these current forecasts is that they are probably weakened by the fact that in this model we are not differentiating between Gaza and the West Bank. At present, those two systems seem entirely decoupled, as even very high levels of violence in Gaza have had almost no effect (other than a few demonstrations) on violence

²⁵These counts of material conflict events most correlate strongly with the Goldstein-scaled WEIS data from earlier KEDS coding.

²⁶For the MS-BSVAR model we find the initial posterior mode and the regime probabilities via the EM algorithm. We then draw 10000 values from this posterior mode to construct the subsequent inferences and forecasts. The BVAR model posterior has a known analytic posterior density. We sample 10000 draws directly from this density to construct the BVAR forecasts. See Brandt and Freeman (2006) for details on the latter.

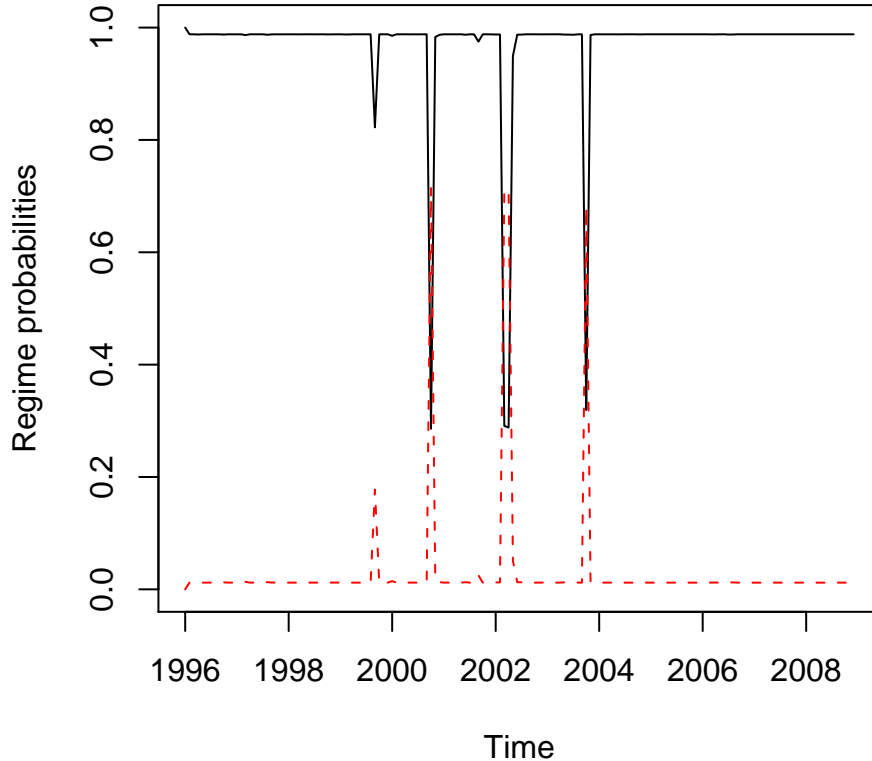


Figure 2: Regime Probabilities: Monthly Levant Material Conflict Data, 1996-2008. Solid (dashed) line is regime one (two)

in the West Bank. We should be able to correct this in the future by looking separately at Fatah and Hamas actions.

The median forecasts across the BVAR and MS-BVAR models are nearly the same. The differences come in how we would characterize the forecast density coverage of these predictions. In general, the BVAR forecast probability intervals that assume constant parameters are too small: they are almost always within the forecast probability intervals of the MS-BVAR forecasts (which allow for the non-zero probability for the high conflict regime). Thus, the last row of Figure 3 indicates that the risk of additional material conflict between the Israelis and Palestinians is higher when we consider the MS-BVAR versus the BVAR forecast models. We believe that this serves as initial evidence for our claims that there are parameter shifts in the data that should be modeled.

This is a rather circumscribed example: in later iterations we plan on expanding this in the manner discussed at the start of this section. The inclusion of additional dyadic conflict and cooperation series (both material and verbal, per the CAMEO scheme) should provide us with more evidence for identifying parameter shifts and regime switching like that uncovered here.

4 Conclusions

Our plan is to expand this forecasting exercise in several directions. First, we plan on covering multiple conflicts around the globe. Second, we have to further scale up the number of actors, dyads, and measures included in our forecasts.

The previous paragraph belies the technical complexity of our proposal. The estimation and

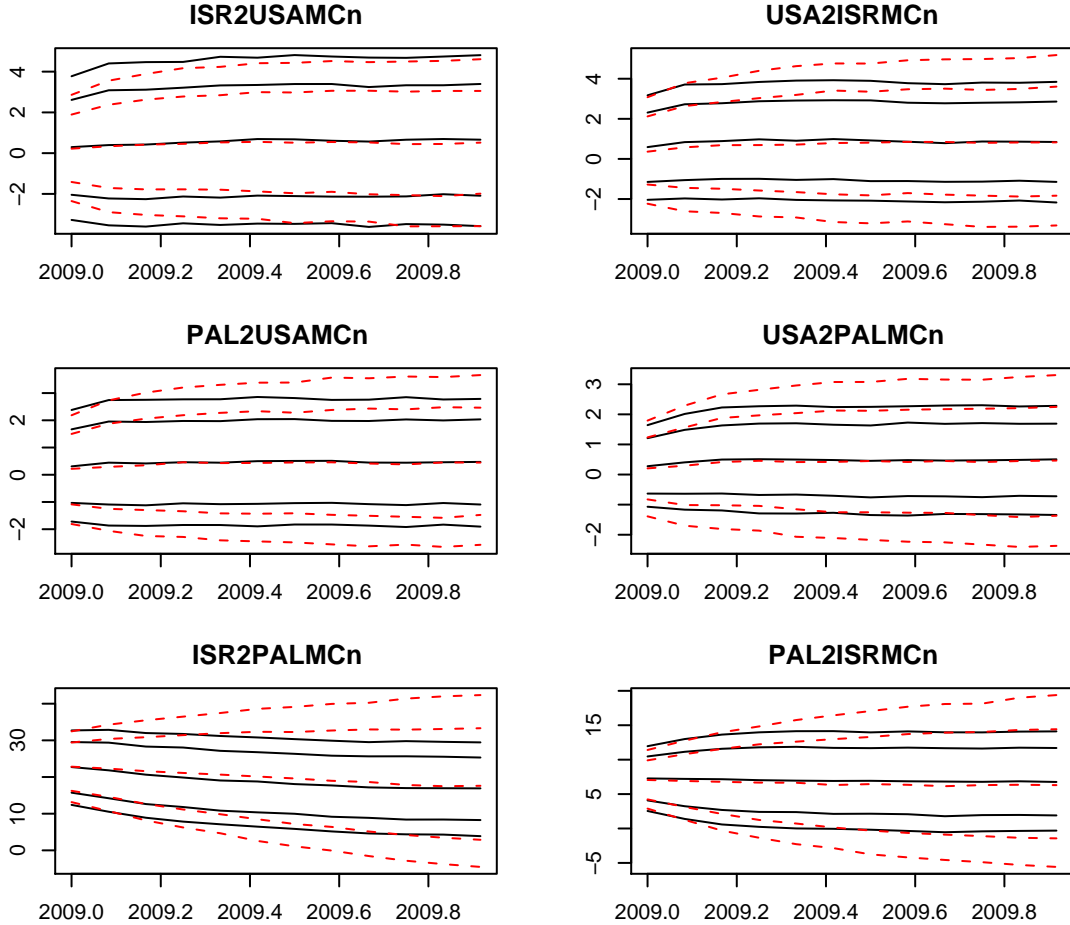


Figure 3: Material conflict forecasts for Israel, Palestine, and U.S. Dyads based on CAMEO material conflict data, 2009:1-2009:12. Black (red) lines are the BVAR (MS-BVAR) forecasts with 16%, 25%, 50%, 75%, and 84% quantiles.

summarization of forecast densities for even a simple BVAR model are computationally intensive: simulating the 100,000 posterior draws of nine month forecasts for six and seven equation (directed-dyad) models in Brandt, Colaresi and Freeman (2008) required over one hour of CPU time²⁷ The computational burdens of estimating Markov-switching versions of these models is several orders of magnitude larger. Further, as we proposed in section 2.1, we will be forecasting multiple conflicts: Israel-Fatah-Hamas, Israel-Lebanon-Hizbollah, India-Pakistan, and China-Taiwan. In each of these cases there are many or more actors than in the *ex post* forecasts of Brandt, Colaresi and Freeman (2008), since the U.S. and other regional actors would also be part of the models. So covering the four conflicts would take at least 4-6 hours for the sample number of periods. Past experience with the BVAR and MS-BVAR models has shown that each additional actor added to a conflict increases the Bayesian posterior simulation time exponentially (since adding one actor increases the scale of the problem from $2^m - m$ to $2^{m+1} - (m + 1)$). Further, each additional Markov-switching state increases the number of posterior modes that must be sampled from $h!$ to $(h + 1)!$ Even for small values of m and h , the initially estimated four hours grows to $4(h!)(2^m - m)$. So covering

²⁷Time measured on an Apple G5 dual core 2.0Ghz PPC 970MP processor.

our basic six actor models ($m = 6$) for a two state model ($h = 2$) would grow to over 464 hours of computational time to simulate the forecast densities over nine time periods for our four conflicts. Our goal is to further work on these issues to reduce the computational burdens and time so that this becomes a feasible exercise.

Finally, once we expand our forecasting capabilities, we will work on further calibration and refinement methods. The plan is that these will be part of an on-going comparison of the *ex ante* forecasts to the actual data. Once we have forecasts generated we will be able to compare the *ex ante* forecasts to the actual data.

A possible reservation to this project may be that it is either already being done in the intelligence community, or that it should be done there rather than by us. We have several responses to this. First, to the best of our knowledge, none of what we are proposing duplicates existing efforts. Hollywood portrayals of an intelligence community endowed with vast technological capabilities for analysis are fiction, not reality, and at present the ability to collect information far outpaces the ability to analyze it.

Second, while forty years ago there appeared to be little point in doing work that did not use classified information because this was, in principle (if not always reality, as the experience in Vietnam indicated) vastly superior to the unclassified work, that situation has changed dramatically with the information sources available for strategic analysis on the Internet. Since we are trying to make forecasts at policy-relevant lead times of six months to five years, not trying to figure out where Osama bin-Laden is sleeping tonight, the additional information provided by classified sources—Hollywood notwithstanding—is not only minimal, but often dysfunctional (Lowenthal, 2009; Johnson and Wirtz, 2004).

Third, there is an increasing interest in technical political forecasting in the NGO community. Two recent examples of this would be the SwissPeace FAST project²⁸ which operated from 1998-2008, and the current Armed Conflict Location and Event Data (ACLED) project of the Peace Research Institute Oslo²⁹. Both the ability to generate real-time event data at a very low cost and some of the quantitative forecasting methods will be of interest to NGOs involved in either conflict monitoring or conflict forecasting. If successful, our system would be particularly useful in assisting NGOs who are called upon to provide material services such as food, medical support and refugee housing in areas experiencing violent political conflict. Acquisition and transportation of these resources typically requires three to six months of advance planning, particularly when large-scale relief supplies that must be shipped by sea are required. We are not suggesting that our models will be the sole factor in these plans, but—given the known problems with the accuracy of qualitative forecasting (Tetlock, 2005)—they can be another input. Quantitative models are already used for the advance planning of reactions to natural disasters: for example, the models of the U.S. Agency for International Development Famine Early Warning System Network, FEWS-NET³⁰, the U.S. National Oceanic and Atmospheric Administration (NOAA) Coastal Services hazards assessment toolkits³¹ and the United Nations Food and Agriculture Organization’s Global Information and Early Warning System³². Our project would extend these to human-initiated political disasters.

Fifth, there is little publicly available software for the methods we are proposing to use to forecast conflict (cooperation). We will continue the practice begun with the development of the MSBVAR and TABARI software packages of making all of our code available and open source.

²⁸<http://www.swisspeace.ch/typo3/en/peace-conflict-research/early-warning/index.html>; for several years Schrodt was involved in generating event data for the Israel-Palestine conflict component of FAST.

²⁹<http://www.prio.no/CSCW/Datasets/Armed-Conflict/Armed-Conflict-Location-and-Event-Data/>

³⁰<http://www.fews.net>

³¹<http://www.csc.noaa.gov/bins/resources/hazards.html>

³²<http://www.fao.org/GIEWS/english/index.htm>

The expanded available of software and data that flows from this project will benefit both applied international relations, time series analysis, econometrics, etc.

Finally, the methods we are developing could be extended to many domains beyond political analysis. The event/network methods we develop for the study of political conflict could quite easily, for example, be used for the study of economic competition or small-group dynamics. That synergy will not occur if tool development such as this remains only behind the screen of security classification. Further, the application of BVAR, MS-BVAR, and other time series models will be advanced. As we have already done with developing modified priors for these models, we believe that there will be novel identification methods for the states and phases that can be used by researchers in other fields.

References

- Bates, R. H., D.L. Epstein, J.A. Goldstone, T.R. Gurr, B. Harff, C.H. Kahl, K. Knight, M.A. Levy, M. Lustik, M.G. Marshall, T.M. Parris, J. Ulfelder and M.R. Woodward. 2003. *Political Instability Task Force Report: Phase IV Findings*. McLean, VA: Science Applications International Corporation.
- Beck, Nathaniel, Gary King and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94:21–36.
- Bennett, P.G. and P. McQuade. 1996. "Experimental Dramas: Prototyping a Multiuser Negotiation Simulation." *Group Decision and Negotiation* 5:119–136.
- Brams, Steven J. and J. M. Togman. 2000. Agreement Through Threats: The North Ireland Case. In *Being Useful: Policy Relevance and International Relations Theory*, ed. N. Miroslav and J. Lepingold. Ann Arbor: University of Michigan Press pp. 325–342.
- Brandt, Patrick T. and John R. Freeman. 2006. "Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis." *Political Analysis* 14(1):1–36.
- Brandt, Patrick T. and John R. Freeman. forthcoming. "Modeling Macro Political Dynamics." *Political Analysis* .
- Brandt, Patrick T. and Justin Appleby. 2007. "Conflict Phases and Processes: Bayesian Markov-Switching Models of Endogenous Systems." Presented at Annual Meeting of the Midwest Political Science Association. Chicago, Illinois.
- Brandt, Patrick T., Michael Colaresi and John R. Freeman. 2008. "The Dynamics of Reciprocity, Accountability, and Credibility." *Journal of Conflict Resolution* 52:343–374.
- Bueno de Mesquita, Bruce. 1997. "A Decision Making Model: Its Structure and Form." *International Interactions* 23(2):235–266.
- Bueno de Mesquita, Bruce. 2002. *Predicting Politics*. Columbus, OH: Ohio State University Press.
- Chatfield, Chris. 2001. Prediction Intervals for Time-Series Forecasting. In *Principles of Forecasting*, ed. J. Scott Armstrong. New York: Springer chapter 15, pp. 475–494.
- Chib, Siddhartha. 1995. "Marginal Likelihood from the Gibbs Output." *Journal of the American Statistical Association* 90(432):1313–1321.
- Christoffersen, P. F. 1998. "Evaluating Interval Forecasts." *International Economic Review* 39:841–862.
- Clements, M. P. and J. Smith. 2000. "Evaluating the Forecast Densities of Linear and Non-Linear Models: Applications to Output Growth and Employment." *Journal of Forecasting* 19:255–276.
- Dawid, A. P. 1984. "Statistical theory: A Prequential Approach." *Journal of the Royal Statistical Society Series A* 147:278–292.

- Diebold, Francis X., Anthony S. Tay and Kenneth F. Wallis. 1998. Evaluating Density Forecasts of Inflation: the Survey of Professional Forecasters. In *Cointegration, Causality and Forecasting: A Festschrift in Honour of Clive W. J. Granger*, ed. R.F. Engle and H. White. Oxford: Oxford University Press.
- Diebold, Francis X. and J. A. Lopez. 1996. Forecast Evaluation and Combination. In *Handbook of Statistics 14: Statistical Methods in Finance*, ed. G. S. Maddala and C.R. Rao. Amsterdam: North-Holland.
- Diebold, Francis X., J. Hahn and Anthony S. Tay. 1999. "Multivariate Density Forecast Evaluation and Calibration in Financial Risk Management: High-frequency Returns On Foreign Exchange." *Review of Economics and Statistics* 81:661–673.
- Diebold, Francis X., Tod A. Gunther and Anthony S. Tay. 1998. "Evaluating Density Forecasts with Applications to Financial Risk Management." *International Economic Review* 39(4):863–883.
- Esty, Daniel C., Jack A. Goldstone, Ted R. Gurr, Barbara Harff, Marc Levy, Geoffrey D. Dabelko and Pamela Surko. 1998. *State Failure Task Force Report: Phase II Findings*. McLean, VA: Science Applications International Corporation.
- Esty, Daniel C., Jack A. Goldstone, Ted R. Gurr, Pamela Surko and Alan N. Unger. 1995. *State Failure Task Force Report*. McLean, VA: Science Applications International Corporation.
- Esty, Daniel C., Jack Goldstone, Ted Robert Gurr, Barbara Harff, Pamela T. Surko, Alan N. Unger, and Robert S. Chen. 1998. The State Failure Project: Early Warning Research for U.S. Foreign Policy Planning. In *Preventive Measures: Building Risk Assessment and Crisis Early Warning Systems*, ed. John L. Davies and Ted Robert Gurr. Lanham, Md.: Rowman and Littlefield pp. 27–38.
- Feder, S. 1995. Factions and Policon. New Ways to Analyze Politics. In *Inside the CIA's Private World: Declassified Articles from the Agency's International Journal, 1955-1992*, ed. H. Bradford Westerfeld. New Haven, CT: Yale University Press pp. 274–292.
- Freeman, John R. and Jeff Gill. 2006. "Dynamic Elicitation of Bayesian Priors for Dark Networks." St. Louis, Mo., Washington University.
- Freeman, John R. and Jeff Gill. 2008. "Dynamic Elicited Priors for Updating Covert Networks: Evidence from an Experiment." Paper presented at the 25th Annual Meeting of the Political Methodology Society, Ann Arbor, University of Michigan.
- Freeman, John R., John T. Williams, Daniel Houser and Paul Kellstedt. 1998. "Long Memoried Processes, Unit Roots and Causal Inference in Political Science." *American Journal of Political Science* 42(4):1289–1327.
- Frühwirth-Schnatter, Sylvia. 2001. "Markov Chain Monte Carlo Estimation of Classical and Dynamic Switching and Mixture Models." *Journal of the American Statistical Association* 96(153):194–209.
- Frühwirth-Schnatter, Sylvia. 2006. *Finite Mixture and Markov Swithcing Models*. Springer Series in Statistics New York: Springer.
- Gneiting, Tilmann. 2008. "Editorial: Probabilistic forecasting." *Journal of the Royal Statistical Society Series A* 171 (Part 2):319–321.

- Gneiting, Tilmann, Fadoua Balabdaoui and Adrian E. Raftery. 208. "Probabilistic Forecasts, Calibration and Sharpness." *Journal of the Royal Statistical Society Series B* 69 (Part 2):243–268.
- Goldstein, Joshua and John R. Freeman. 1991. "U. S.-Soviet-Chinese Relations: Routine, Reciprocity or Rational Expectations?" *American Political Science Review* 85(1):17–36.
- Goldstein, Joshua S. and John R. Freeman. 1990. *Three-way street: strategic reciprocity in world politics*. Chicago: University of Chicago Press.
- Goldstein, Joshua S. and Jon C. Pevehouse. 1997. "Reciprocity, Bullying, and International Cooperation: A Time Series Analysis of the Bosnia Conflict." *American Political Science Review* 91(3):515–529.
- Goldstein, Joshua S., Jon C. Pevehouse, Deborah J. Gerner and Shibley Telhami. 2001. "Reciprocity, Triangularity and Cooperation in the Middle East, 1979-1997." *Journal of Conflict Resolution* 45(5):594–620.
- Granger, Clive W. J. 2005. "Preface: Some Thoughts on the Future of Forecasting." *Oxford Bulletin of Economics and Statistics* 67(supplement):707–711.
- Gurr, Ted R. and Barbara Harff. 1996. *Early Warning of Communal Conflict and Humanitarian Crisis*. Monograph Series on Governance and Conflict Resolution Tokyo: United Nations Press.
- Hamill, T. M. 2001. "Interpretation of Rank Histograms for Verifying Ensemble Forecasts." *Monthly Weather Review* 129:550–560.
- Hudson, Valerie M., Philip A. Schrodtt and Ray D. Whitmer. 2008. "Discrete Sequence Rule Models as a Social Science Methodology: An Exploratory Analysis of Foreign Policy Rule Enactment Within Palestinian-Israeli Event Data." *Foreign Policy Analysis* 4(2):105–126.
- Johnson, Loch K. and James J. Wirtz. 2004. *Strategic Intelligence*. Los Angeles: Roxbury.
- Kass, Robert E. and Adrian E. Raftery. 1995. "Bayes Factors." *Journal of the American Statistical Association* 90(430):773–795.
- King, Gary and Langche Zeng. 2001. "Improving Forecasts of State Failure." *World Politics* 53:623–658.
- Kling, John L. and David A. Bessler. 1989. "Calibration-based Predictive Distributions: An Application of Prequential Analysis to Interest Rates, Money, Prices, and Output." *Journal of Business* 62(4):477–499.
- Krolzig, Hans-Martin. 1997. *Markov-Switching Vector Autoregressions: Modeling, Statistical Inference, and Application to Business Cycle Analysis*. Berlin: Springer.
- Lebovic, James H. 1994. "Before the Storm: momentum and the onset of the Gulf War." *International Studies Quarterly* 38:447–474.
- Lowenthal, Mark M. 2009. *Intelligence: From secrets to policy*. Washington: CQ Press.
- Lund, Michael S. 1996. *Preventing Violent Conflicts: A Strategy for Preventive Diplomacy*. Washington, D.C.: United States Institute for Peace.

- Organski, A. F. K. and E. Lust-Okar. 1997. "The Tug of War Over Jerusalem: Leaders, Strategies and Outcomes." *International Interactions* 23(4):333–350.
- Pevehouse, Jon C. and Joshua S. Goldstein. 1999. "Serbian Compliance or Defiance in Kosovo? Statistical Analysis and Real-Time Predictions." *Journal of Conflict Resolution* 43(4):538–546.
- Rosenblatt, M. 1952. "Remarks on a Multivariate Transformation." *Annals of Mathematical Statistics* 23:470–472.
- Schrodt, Philip A. 2000. "Forecasting Conflict in the Balkans Using Hidden Markov Models." Paper presented at the Annual Meeting of the American Political Science Association, Washington, D.C.
- Schrodt, Philip A. 2006. Forecasting Conflict in the Balkans Using Hidden Markov Models. In *Programming for Peace: Computer-Aided Methods for International Conflict Resolution and Prevention*, ed. Robert Trapp. Dordrecht, Netherlands: Kluwer Academic chapter 10, pp. 161–184.
- Schrodt, Philip A. 2007. "Event Data Scaling Using Item Response Theory." Presented at the 24th of the Annual Meeting of the Society for Political Methodology.
- Schrodt, Philip A. and Deborah Gerner. 2000. "Cluster Based Early Warning Indicators For Political Change in the Contemporary Levant." *American Political Science Review* 94(4):803–818.
- Schrodt, Philip A. and Deborah J. Gerner. 2004. "An Event Data Analysis of Third-Party Mediation." *Journal of Conflict Resolution* 48(3):310–330.
- Scott, Steven L. 2002. "Bayesian Methods for Hidden Markov Models: Recursive Computing in the 21st Century." *Journal of the American Statistical Association* 97(457):337–351.
- Sims, Christopher A., Daniel F. Waggoner and Tao Zha. 2008. "Methods for inference in large multiple-equation Markov-switching models." *Journal of Econometrics* 146(2):255–274.
- Sims, Christopher A. and Tao A. Zha. 1998. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review* 39(4):949–968.
- Tay, Anthony S. and Kenneth F. Wallis. 2000. "Density Forecasting: A Survey." *Journal of Forecasting* 19:235–254.
- Taylor, J. W. 1999. "Evaluating Volatility and Interval Forecasts." *Journal of Forecasting* 18:111–128.
- Tetlock, Philip E. 2005. *Expert Political Judgement*. Princeton: Princeton University Press.
- Timmerman, Allan. 2000. "Editorial: Density Forecasting in Economics and Finance." *Journal of Forecasting* 19:231–234.
- Wallis, Kenneth F. 1995. Large-scale Macroeconomic Modeling. In *Handbook of Applied Econometrics*, ed. M. H. Pesaran and M. R. Wickens. Oxford: Blackwell.
- Wolfers, Justin and Eric W. Zitzewitz. 2004. "Prediction Markets." *Journal of Economic Perspectives* 18(2):107–126.