Advances in Bayesian Time Series Modeling and the Study of Politics: Theory Testing, Forecasting, and Policy Analysis

Patrick T. Brandt

School of Social Sciences, University of Texas at Dallas,
Box 830688, Richardson, TX 75083
e-mail: pbrandt@utdallas.edu (corresponding author)

John R. Freeman

Department of Political Science, University of Minnesota,
267 19th Avenue, Minneapolis, MN 55455
e-mail: freeman@polisci.umn.edu

Bayesian approaches to the study of politics are increasingly popular. But Bayesian approaches to modeling multiple time series have not been critically evaluated. This is in spite of the potential value of these models in international relations, political economy, and other fields of our discipline. We review recent developments in Bayesian multi-equation time series modeling in theory testing, forecasting, and policy analysis. Methods for constructing Bayesian measures of uncertainty of impulse responses (Bayesian shape error bands) are explained. A reference prior for these models that has proven useful in short- and medium-term forecasting in macroeconomics is described. Once modified to incorporate our experience analyzing political data and our theories, this prior can enhance our ability to forecast over the short and medium terms complex political dynamics like those exhibited by certain international conflicts. In addition, we explain how contingent Bayesian forecasts can be constructed, contingent Bayesian forecasts that embody policy counterfactuals. The value of these new Bayesian methods is illustrated in a reanalysis of the Israeli-Palestinian conflict of the 1980s.

1 Introduction

Bayesian approaches to the study of politics have become increasingly popular. With a few notable exceptions, few of us employ Bayesian time series methods in the study of politics.

More than a decade ago Williams (1993) wrote a piece on this subject in *Political Analysis.* His article was based on work done at the Minneapolis Federal Reserve Bank in the early 1980s.[1] Recently, Martin and Quinn (2002), drawing on advances in Bayesian time series statistics (West and Harrison 1997), showed how Bayesian multivariate dynamic linear models can be used to study changes in the ideal points of Supreme Court justices. Martin and Quinn only scratched the surface of these advances. In fact, most political scientists are unaware of the improvements and extensions that have been made in Bayesian vector autoregressive (BVAR) methods and Bayesian time series statistics. Studies of international conflict and of other important topics can benefit by incorporating the advances that have been made in Bayesian time series statistics over the last decade.[2]

We review key developments in Bayesian time series modeling for theory testing.[3] Most time series work in political science in the 1980s and 1990s failed to provide any measures of uncertainty for causal inference. Scholars often failed to supply error bands or probability assessments for the impulse responses and dynamic inferences of their models. The error bands that were provided were based on a Monte Carlo procedure that is now viewed as inferior to Bayesian shape bands we discuss. We review the recent work on probability assessment in time series analysis, including the development of means to construct measures of uncertainty for the impulse responses and forecasts of Bayesian multi-equation time series models. We also highlight the special nature of time series analysis vis-à-vis more familiar forms of inference: because of nonstationarity, Bayesian posterior probabilities and classical confidence intervals can be in "substantial conflict" (Sims and Zha 1999).[4]

As Beck et al. (2000, 2004) have recently argued, forecasting is at the root of inference and prediction in time series analysis. Estimation and inference in time series modeling involves the minimization of one-step (or multi-step) forecast errors (Clements and Hendry 1998). Establishing a model's superiority entails showing that it produces smaller forecast errors than its competitors. Such evaluations depend on the structure of the time series model—a structure that at best one believes probabilistically. Assessing how a model specification and beliefs about it are related to inference and forecasting performance (both in and out of sample) are extremely important. Recognizing this, we discuss a popular new reference prior that has performed well in macroeconomics and show how it can be applied in political forecasting. Next, we highlight some potentially useful extensions, such as how to construct from our BVAR models counterfactually contingent forecasts. Closely related

---

[1]See also McGinnis and Williams (1989) for an application of the early 1980s Minneapolis Federal Reserve approach to the study of superpower rivalry.

[2]Martin and Quinn (2002, n. 2) point out that the "machinery" of West and Harrison (1997) can be applied to binary cross-sectional time series models. But to our knowledge no political scientist has attempted such an application. Note that Martin and Quinn's dynamic linear multivariate model does not provide for interdependence between their units of analysis, more specifically, for any interrelationships between judges' decisions (p. 138). The Bayesian multi-equation time series model expressly allows for such interdependence or for endogeneity. It too is a special case of the Kalman filter. The only other applications of Bayesian time series we found are Brandt et al.'s (2000) and Brandt and William's (2001) development of count time series models using an extended Kalman filter, Buckley's (2002) review of Bayesian linear dynamic models, Jackman's (2000) linear regression example in his Workshop piece in the *American Journal of Political Science*, and Western and Kleykamp's (2004) study of change points in the recent special issue of *Political Analysis.* Of course, time series statistics are sometimes used to assess the convergence of computational algorithms used by Bayesians (see Geyer 1992).

[3]We focus here on BVAR models. We consider vector error correction models as special (restricted) cases of VAR models, so much of our analysis applies to error correction and vector error correction models (VECMs).

[4]This is a point that many of the leading Bayesians in our discipline overlook (e.g., Gill 2004, p. 328; see also Jackman 2004, pp. 486, 489).

to this concept are policy contingent or counterfactual forecasts that may be used for policy evaluation.

Our discussion is divided into three parts. Part one reviews the handful of Bayesian time series analyses in political science. It shows how recent advances in time series econometrics and statistics potentially can improve these analyses. In this section, we propose the adoption of an easy-to-specify prior distribution for multi-equation time series models. This reference prior is potentially of enormous value in explaining and analyzing counterfactually political processes like international conflict. Part two provides technical explanations of this reference prior and of how to construct Bayesian error bands and forecasts, including contingent Bayesian forecasts.[5] The usefulness of these advances is illustrated in part three in a reanalysis of the Israeli-Palestinian conflict of the 1980s.

2 Bayesian Multi-equation Time Series Analysis in Political Science: A Review

Theory testing and policy analysis with multiple time series models involves three interrelated enterprises: innovation accounting, forecasting, and counterfactual analysis.[6] Past political science articles explain the tools used in each enterprise (see especially Freeman et al. 1989). New texts also explain these tools (Brandt and Williams forthcoming). Readers unfamiliar with multiple time series models are urged to study these works before proceeding.

2.1 *Innovation accounting*

Innovation accounting is the determination of how a (normalized) shock or surprise in one time series affects other time series. If a variable X_t causes another variable Y_t, a significant part of the response of Y_t will be accounted for by the (normalized) shock in X_t. For the users of multi-equation time series models, these impulse responses or innovation accounting are an essential component of theory testing.

The problem is that political scientists rarely provide measures of the uncertainty of these impulse responses. Usually in political science, no error bands are provided for them. Without such bands, we cannot gauge the soundness of our causal inferences and we have no means to convey how certain we are of the direction and (nonzero) magnitude of the responses.[7] The few political scientists who provide such bands use Monte Carlo methods to construct them. For example, the Monte Carlo method was used by Williams (1993) to construct the error bands for the impulse responses of his unrestricted, frequentist VAR model of Goldstein's long cycle theory. But this same method could not be applied to his Bayesian or time-varying BVAR models because it does not have a tractable posterior that could be easily simulated.[8]

---

[5]New software packages like *Zelig* do not contain code for performing Bayesian time series analyses like that which we describe here. Familiar packages like RATS, as we note below, also are inadequate for this purpose. One author has developed a new software package for *R*, MSBVAR, that will produce Bayesian shape error bands for impulse responses and other advances that we present in this article.

[6]This passage draws from Kilian (1998) and Sims and Zha (1999).

[7]On this point see Runkle (1987). Illustrative of political science research that provides no such error bands are Goldstein and Freeman (1990, 1991) and Freeman and Alt (1994). Examples of works with error bands constructed with Monte Carlo methods (employing classical inference) are Williams and Collins (1997) and Edwards and Wood (1999).

[8]Williams and others used the code provided in RATS to construct their error bands. This code was for many years based on Monte Carlo methods. Monte Carlo and analytic derivative methods are now available in RATS. But these methods and the bootstrap are all based on classical or flat-prior Bayesian inference. Note that these methods were not previously extended to the "Minnesota prior" model used in Williams (1993).

In recent years Monte Carlo and related methods and the classical form of inference associated with them have been criticized by Bayesian time series analysts. They proposed an alternative approach to constructing error bands, one based on the likelihood shape of models. The impulse responses of vector autoregressions are difficult to construct for three reasons:

1. Estimates of the underlying autoregressive form parameters have sampling distributions that depend strongly in shape as well as location on the true value of the parameters, especially in the neighborhood of parameters that imply non-stationarity.
2. Impulse responses are highly nonlinear functions of underlying autoregressive reduced form parameters.
3. The distribution of the estimate of a particular response at a particular horizon depends strongly on the true values of other impulse responses at other time horizons, with no apparent good pivotal quantity to dampen such dependence on nuisance parameters (Sims and Zha 1995, p. 1; see also p. 10, esp. n. 9), Sims and Zha 1999, p. 1127; and Ni and Sun 2003, p. 160).

While some classical approaches like the nonparametric bootstrap and parametric Monte Carlo integration are asymptotically sound for stationary data, in small samples they can be inaccurate in terms of estimating the location, width, and skewness of the error bands of the responses. These problems also surface when data are nonstationary, since "in a finite sample the accuracy of the asymptotic approximation begins to break down as the boundary of the stationary region of the parameter space is approached" (Sims and Zha 1995, p. 2). Kilian (1998) proposed corrections to the classical approaches to error band construction. He showed that a bias-corrected bootstrap procedure outperforms the nonparametric bootstrap and Monte Carlo integration methods.

However, in a series of papers, Sims and Zha (1995, 1999) raise questions about the adequacy of Kilian's and others' methods for constructing such bands.[9] They argue that classical approaches to constructing error bands for impulses seriously confound information about the model fit and the uncertainty of parameters.[10] Sims and Zha propose an explicitly Bayesian approach to the construction of error bands for impulse responses. They argue that the best way to represent uncertainty about the location and skewness of the impulse responses—particularly the "serial correlation in the uncertainty" over time—is through an analysis of the likelihood shape. Using an eigenvector decomposition of the impulse responses, Sims and Zha produce "probability assessments" for the impulse responses.[11] This method produces bands that are more informative about the corresponding likelihood shape than the bands produced by Kilian's and others' methods. In a series of experiments with artificial and actual (macroeconomic) data, Sims and Zha show that their Bayesian shape error bands are more accurate in terms of location and skewness than the bands produced by other methods.

---

[9]This contribution was made in the late 1990s, hence the absence of any error bands in Williams's piece for his BVAR models (see 1993, Figs. 2–6).

[10]Sims and Zha (1999, p. 1114) argue that the confidence intervals associated with the classical approach to inference "mix likelihood information and information about model fit in a confusing way: narrow confidence bands can be indicators either of precise sampling information about the location of the parameters or of strong sample information that the model is invalid. It would be better to keep the two types of information separate."

[11]The eigenvector decomposition developed by Sims and Zha is similar to a dynamic factor analysis that accounts for the main sources of the variation in the responses over time.

Three additional points should be made with regard to the work on error bands of impulse responses. First, Bayesians like Sims and Zha (1995, n. 15) prefer 68% (approximately one standard deviation) coverage or posterior probability intervals to the more familiar 95% confidence intervals. In their view, the former are much more indicative of the "relevant range of uncertainty" than the latter, which are indicative of "pretesting and data mining."[12] Second, the Sims-Zha method is for identified vector autoregressions, for example, models for which an ordering of the variables and hence an orthogonalization of the variance-covariance matrix of errors has been imposed (Hamilton 1994, section 11.6). Overidentified models require a modified approach to construct posterior probabilities for impulse responses.[13] Finally, the methods developed by Sims and Zha can be extended to any analysis in which one must characterize uncertainty about the values of an estimated function of time and uncertainties about the future values of this function are interdependent (Sims and Zha 1999, p. 1129).

2.2 *Forecasting*

Political scientists recently have been reminded of the importance of forecasting as a means of evaluating statistical models. For example, debates about the relative virtues of neural net models of war focus, to a great extent, on those models' forecasting performances (Beck et al. 2000, 2004; de Marchi et al. 2004). It has been known for some time that unrestricted VAR models tend to overfit the data, attribute unrealistic portions of the variance in time series to their deterministic components, and overestimate the magnitude of the coefficients of distant lags of variables (because of sampling error).[14]

Doan et al. (1984) developed a BVAR model that addresses these problems. Their model is based on a belief that most time series are best predicted by their mean or their values in the previous periods. For nonstationary data this means that the data are first-order integrated, perhaps with drift (deterministic constants), or that the first differences of each series are unpredictable. This and beliefs about the other coefficients in the VAR model—for example, that all coefficients except the coefficient on the first own lag of the dependent variable have mean zero and that certainty about this belief is greater the more distant the lag of the variable to which the coefficient applies—are embodied in the so-called Minnesota prior.[15] One of the key features of this prior is that it treats the variance-covariance matrix of the reduced form residuals as diagonal and fixed. In addition, it does not embody any beliefs an analyst might have about how the prior distribution of the variance-covariance matrix of residuals is related to the prior distribution of the reduced form coefficients. This means the associated likelihood reduces to the product of independent normal densities for the model coefficients Kadiyala and Karlsson (1997). Litterman (1986) concluded that, for the period from 1950 to the early 1980s, a BVAR

---

[12]The concept of highest posterior density region (HPD) is an important related idea here (see Kadiyala and Karlsson 1997; Gill 2004).

[13]In this case there is a possibility of likelihoods with multiple peaks; strong asymmetry in error bands is indicative of this situation. The fitted model must be reparameterized and adjustments made to the flat prior to make the estimation possible. See Sims and Zha (1999, section 8) and Waggoner and Zha (2000).

[14]On the problems with using unrestricted VARs for forecasting, see Zha (1998) and Sims and Zha (1998, pp. 958–60). The poor performance of unrestricted VARs is demonstrated in such works as Fair and Shiller (1990). Interestingly, the new work on neural nets uses in its benchmark models what is, in effect, a deterministic counter of time since the last war (Beck et al. 2000, 2004). This probably makes them very stringent benchmarks vis-à-vis the performance of more theoretically motivated neural net models.

[15]Doan, Litterman, and Sims were all associated with the University of Minnesota or the Minneapolis Federal Reserve Bank at the time.

model based on the Minnesota prior performed as well as or better than the models of major commercial, economic forecasters. Moreover, this model was much cheaper to use and it did not require "arbitrary judgments" to make it perform well.[16]

While the Minnesota prior and BVAR model developed by Litterman is recognized as a valuable tool for forecasting stock prices and other phenomena for which beliefs about reduced form coefficients are unrelated to correlations in the innovations, Sims and Zha (1998, p. 967) found it incompatible with their beliefs about the macroeconomy. Their beliefs are that the macroeconomy is best described by a dynamic simultaneous equation model in which the beliefs (prior) are specified for the structural rather than the reduced form parameters. These beliefs are correlated across equations in a way that depends on the contemporaneous relationship among the variables (the covariance matrix of reduced form disturbances). Operationally, they substituted a normal-inverse Wishart prior for the whole system of VAR coefficients for the Litterman equation-by-equation prior.[17] The Sims-Zha prior introduces a new hyperparameter for the overall tightness of the standard deviation on the observed errors and of their intercorrelations. We argue below that this is more in keeping with ideas like reciprocity in international relations. That is, the Sims-Zha prior more accurately reflects our beliefs that what one belligerent does to its adversary is as likely to reflect the adversary as well as its own past behavior.[18]

Second, Bayesian time series analysts have made fuller provisions for nonstationarity. As noted above, nonstationarity is a key feature of many time series data, one that can create major difficulties for classical inference. Like economists, political scientists have found that many of their series are near-integrated or nonstationary (Ostrom and Smith 1993; Box-Steffensmeier and Smith 1996; DeBoef and Granato 1997; Freeman et al. 1998). For this reason, the Sims-Zha prior is also theoretically relevant. It adds hyperparameters that capture beliefs about the sum of the coefficients of lagged dependent variables (the number of unit roots in the system of variables) and about the possibility of cointegration among these stochastic trends. Sims and Zha (1998, p. 958) argue that in comparison to the standard practice of adding deterministic trends to each equation to represent long-term trends, this

---

[16]In essence, rather than impose exact restrictions on the model's coefficients such as zeroing out lags or deleting variables altogether, the BVAR model imposes a set of inexact restrictions on the coefficients. The key features of the Minnesota prior are a) the tightness of the distribution around the prior mean of unity for the coefficient on the first own lag of the dependent variable, b) the tightness of the distribution around the mean of zero on the coefficients for the lags of the other variables in an equation relative to the tightness of the distribution around value of unity for the first own lag of the respective dependent variable, and c) how rapidly the tightness of the distributions on the lag coefficients goes to zero as the lag length of the variables increases. As regards the constants in each equation, Litterman (1986, p. 29) notes the large degree of ignorance economists had in the 1980s about constants' prior means and, by implication, the nonstationarity of economic processes.

[17]Sims and Zha (1998, p. 955) write, "Thus if our prior on [the matrix of structural coefficients for contemporaneous relationships between the variables] puts high probability on large coefficients on some particular variable j in structural equation i, then the prior probability on large coefficients on the corresponding variable j at the first lag is high as well." An often unappreciated fact about the Litterman prior is that it is not a proper prior for the full VAR model. This is because it is only formed for each of the equations in the model. Hence the resulting posterior distribution is not of a conjugate or standard form. In contrast, Sims and Zha (1998) show how to construct a flexible class of priors for BVAR models. For additional details see Ni and Sun (2003) and Kadiyala and Karlsson (1997).

[18]Kadiyala and Karlsson (1997) explain the Normal inverse-Wishart prior. They also show how the Diffuse, Normal-Diffuse, and Extended Natural Conjugate priors can be used to relax the specifications in the Minnesota prior. Kadiyala and Karlsson explain and explore in applied work the computational issues for these four priors (the Normal-Diffuse and Extended Natural Conjugate priors, unlike the Normal-Wishart and Diffuse priors, do not have closed form posterior moments). Their illustrations are forecasts of the Swedish unemployment rate and of the U.S. macroeconomy. Kadiyala and Karlsson conclude that when beliefs are like those that underlie the Minnesota prior and computation is a concern, the Normal inverse-Wishart prior is preferred over the abovementioned alternatives. A similar, more recent evaluation of noninformative and informative Minnesota prior is Ni and Sun (2003).

Bayesian approach to capturing nonstationary features of data performs much better in forecasting.[19]

The performance of the Sims-Zha prior in forecasting has been compared to that of the Minnesota prior and to other forecasting models by numerous econometricians. An example is Robertson and Tallman's (1999) article. They find that for the U.S. macroeconomy, the provision for (near) nonstationarity enhances forecast performance more than the provision for cross-equation dependencies.[20] Zha (1998) also compares the performance of his and Sims's prior to the forecasts of commercial services for the U.S. macroeconomy, including the results of the Blue Chip forecasts. Like Litterman before him, Zha contends that his and Sims's prior performs as well as or better than methods of commercial forecasters.[21]

Several points should be highlighted with regard to Bayesian time series forecasting. The first is that there are several ways to assess the accuracy of forecasts. Analysts now routinely produce error bands for their forecasts, including Bayesian shape bands. Again, these are typically 0.68 probability bands that summarize the central tendency of the forecasts (Sims and Zha 1998; Zha 1998; Waggoner and Zha 1999). Analysts also use familiar measures such as residual mean square error (RMSE) and mean absolute error (MAE) to gauge the difference between the posterior means of Bayesian forecasts and the actual data (Robertson and Tallman 1999). Others produce single variable and bivariate posterior probability densities for their forecasts and then compare the location of the (joint) posterior means to the actual data and (or) to the point forecasts from competing models (Zha 1998). In addition, Bayesian time series analysts have developed a set of measures based on cumulative Bayes's factors (CBFs) that can be used to assess the performance of such models over time. Second, a particular set of hyperparameter values for the Sims-Zha prior is often referred to as a "reference prior" (see Gill 2002, section 5.2). These values are based in part on the extensive experience econometricians have had forecasting macroeconomic time series in the post-World War II era and to the widely held beliefs that economists have about macroeconomic dynamics. One of the aims of this article is to develop a similar reference prior for political science, to incorporate in our priors in a systematic way the knowledge we have about international conflict (see Gill 2004, esp. p. 333).[22] Third, the idea of theoretical structure also surfaces in Bayesian time series forecasting. Sims and Zha show how to incorporate a fuller, theoretically informed structural model of the innovations in the variables in Bayesian forecasting. This further extension makes a connection between the correlation of the innovations and beliefs about the correlation of coefficients in the reduced form model: "Once we know that reduced form forecast errors for [two variables] are

---

[19]Such deterministic trends tend to soak up too much of the variance in the time series. Zha (1998) argues that these two new hyperparameters do a better job of accounting for the possibility of near-(co)integration than exact restrictions.

[20]Robertson and Tallman (1999) compare the forecasting performance of an unrestricted VAR model, VAR in differences (exact restrictions) with AIC determined lag length, a BVAR model based on the Minnesota prior, a BVAR model based on the Minnesota prior but with the dummy variables added to capture beliefs about the number of unit roots and cointegration in the system, a BVAR model based on the Sims and Zha prior, and a partial Sims-Zha BVAR model in which the provision for beliefs about unit roots and cointegration are omitted. In brief, it is the provision for unit roots and cointegration that, according to Robertson and Tallman, is most responsible for the improvement in forecasting performance for the U.S. economy in the 1986–1997 period over unrestricted VARs and VARs with exact restrictions.

[21]The Blue Chip forecasts are based on a survey of economic forecasters. Zha (1998, n. 5) uses the "consensus" forecasts from this source.

[22]See Sims and Zha (1998, n. 7). Doan et al. (1984) originally referred to the Minnesota prior as a "standardized prior" (p. 2) and an "empirical prior" (p. 5). When Litterman (1986) and others refer to judgment-based (vs. model-based) forecasting they are referring to the practice of experts literally adjusting the output of models to conform with their hunches about the future.

positively correlated, we are likely to believe that coefficients on [lags of these same variables] differ from the random walk prior in the same way'' (Sims and Zha 1998, p. 967).

2.3 *Counterfactuals*

The third way in which BVAR models are used in theory testing is counterfactual analysis. Counterfactual analysis is a valuable tool in theory evaluation. Counterfactuals are not simply additional tests of theories; they also are tests of theories' logical implications. In international relations, for example, accounts of conflict dynamics often include claims about the hypothetical effects that increases in trade might have on belligerency. By positing a hypothetical increase in trade in a conflict model, a researcher then can analyze the impact of trade levels that, according to liberal peace proponents, ought to greatly reduce international conflict.

Among the most important conditions for a meaningful counterfactual is ''cotenability.'' The hypotheticals should not alter ''other factors that materially affect outcomes'' (Fearon 1991, p. 93). In addition, hypotheticals should be ''in the range of the observed data'' (King and Zeng 2004, p. 1). In terms of the previous example, hypothetical increases in trade should not change the way belligerents react to attacks by their adversaries. The magnitude of these increases should be plausible historically (in sample).

Time series analysts employ conditional forecasting to study counterfactuals. Counterfactuals are translated into constraints on the values that a selected variable may take in the future, either a fixed value (hard condition) or a range of values (soft condition). Forecasts then are drawn from the posterior distribution in a way that satisfies this constraint at all future times and, equally important, takes into account both parameter uncertainty and uncertainty about the random shocks that the system might experience (Waggoner and Zha 1999). This Bayesian approach treats all variables, including that which is manipulated counterfactually, as endogenous.[23] Finally, conditional forecasts of this kind are robust to alternative identifications (triangularizations) of the structural BVAR model. Below we explain Bayesian conditional forecasting in greater detail and illustrate its use in an example from international relations.

Policy analysis with BVAR models is essentially conditional forecasting. The counterfactual is a hypothetical about the fixed or range of value(s) in an endogenous policy variable at all points in the future. Policy outcomes are the corresponding, conditional forecasts for the remaining endogenous variables in the system.[24]

For many years a debate raged about whether such analysis is feasible. If the public could anticipate the decisions of and accurately monitor policy makers, it presumably would nullify the impact of the policy before it was adopted. Analysts would have to address the fact that the parameters in their policy outcome equations are complex, nonlinear functions of agents' expectation formation rules (regarding policy choices). Efforts to use BVAR models to formulate intervention strategies for international conflicts and other applications in political science would have to do this as well.[25]

---

[23]That is, in drawing from the corresponding posterior distribution, one allows for all possible *combinations* of past and present values of the endogenous variables (subject to constraints) and past and present shocks that could have produced the counterfactual value(s) of the selected variable at each future point in time as well as the parameter uncertainty in the model.

[24]Illustrative of this approach to policy analysis in macroeconomics is the practice of fixing values of the Federal Funds Rate at some level or to remain in some range (see Waggoner and Zha 1999). For further discussion of the importance of treating policy as endogenous in such analysis see Freeman et al. (1989).

[25]This is the Lucas critique. One way to think of it is that policy reaction functions cannot simply be substituted into policy output equations because the parameters in the latter are functions of the parameters the former Sims (1987b).

Sims (1987a) and others (Cooley et al. 1984; Granger 1999) refute this critique. If policies were optimal and agents had perfect (exactly the same) information as the policy maker, forecasts conditioned on hypothetical policy choices would be difficult to employ. But, *because of politics*, policy is not optimal and agents are not perfectly informed:

> Actual policy always contains an unpredictable element from this source. The public has no way of distinguishing an error by one of the political groups in choosing its target policy from a random disturbance in policy from the political process. Hence members of such a group can accurately project the effects of various policy settings they might aim for by using historically observed reactions to random shifts in policy induced by the political process. (Sims 1987b, p. 298)

Thus politics produces enough "autonomous variation in policy"—policy variation the source of which agents cannot discern—that we can identify multivariate time series models and then use them to study policy counterfactuals.[26] It is important to note that BVAR models have embedded in them reaction functions and mechanisms by which agents form expectations. These functions and mechanisms usually are not made explicit or separated out from other dynamics. But these functions and mechanisms are assumed to be present in the data-generating process (Sims 1987b, p. 307; see also Zha 1998, p. 19). The bottom line is that thanks to the workings of the institutions on which we as political scientists focus, we should be able to use the recent development in Bayesian time series analysis to produce policy contingent forecasts that will inform policy interventions that are of interest to political scientists.

Table 1 summarizes the key features of frequentist and Bayesian multi-equation time series modeling.

3   Technical Development

This section presents the technical details of the Bayesian VAR models. We first describe the specification of the Sims-Zha BVAR prior. We then present the Bayesian approach to innovation accounting. We explain how to construct Bayesian-shape bands for impulse responses, highlighting how and why the coverage of these response densities is superior to those produced by frequentist methods. Finally we present methods for forecasting and policy analysis.

3.1   *Bayesian Vector Autoregression with Sims-Zha Prior*

We begin by describing the identified simultaneous equation and reduced form representations of a VAR model. We develop both representations of the model, because unlike Litterman (1986), who proposed it for the reduced form of the model, Sims and Zha (1998) specify the prior for the simultaneous equation version of the model. The advantage of the latter approach is that it allows for a more general specification and can produce a tractable multivariate normal posterior distribution. A consequence is that the estimation of the VAR coefficients is no longer done on an equation-by-equation basis as in the reduced form version. Instead, we estimate the parameters for the full system in a multivariate regression.[27]

---

[26]In his paper, Sims (1987b) also shows that a unitary public authority that possesses information not possessed by the public can use conditional forecasts to formulate optimal policy.

[27]We employ the standard usage of "multivariate regression" to mean a regression model for a matrix of dependent variables or where the dependent variable observations are multivariate, as opposed to "multiple" regression where the dependent variable is univariate or scalar, regardless of the number of regressors.

Table 1 Frequentist and Bayesian VAR model comparisons

	Frequentist VAR	Litterman	Sims-Zha
Main Elements of Prior			
1) Tightness of dynamics	1) Flat	1) Set by analyst	1) Set by analyst
2) Tightness of scale	2) Flat	2) Not applicable	2) Separate hyperparameter
3) Own-other lag	3) No difference	3) Can be given different weights	3) Not applicable
4) Lag decay	4) Data determined	4) Separate hyperparameter	4) Separate hyperparameter
5) Correlation of beliefs across equations	5) Not applicable	5) Uncorrelated across equations	5) Correlated across equations similar to residual covariance
6) Unit roots/trends beliefs	6) Produce degenerate sampling distributions	6) Does not allow for cointegration	6) Separate hyperparameters for cointegration and trends
Estimation			
Estimator	Equation-by-equation OLS	Equation-by-equation OLS	Multivariate OLS
Prior structure	Not applicable (diffuse)	Prior is set for each equation	Prior is set for the system
Posterior	Normal-Inverse Wishart	Not tractable Requires importance sampling	Normal-Inverse Wishart or Normal-Flat
Inferences			
Error bands for impulses and forecasts	Assume asymptotic normality of quantities of interest	Not applicable because of non-tractable posterior	Eigendecomposition and quantile methods summarize posterior pdf
Examples	Goldstein and Freeman (1990, 1991) Williams (1990) Edwards and Wood (1999)	Williams (1993) Williams and Collins (1997)	None

10

Consider the following (identified) dynamic simultaneous equation model (matrix dimensions indicated below matrices),

$$\sum_{l=0}^{p} \underset{1 \times m}{y_{t-\ell}} \underset{m \times m}{A_\ell} = \underset{1 \times m}{d} + \underset{1 \times m}{\epsilon_t}, \quad t = 1, 2, \ldots, T. \tag{1}$$

This is an m-dimensional VAR for a sample of size T with y_t a vector of observations at time t, A_ℓ the coefficient matrix for the ℓth lag, p the maximum number of lags (assumed known), d a vector of constants, and ϵ_t a vector of i.i.d. normal *structural shocks* such that

$$E[\epsilon_t | y_{t-s}, s > 0] = \underset{1 \times m}{0}, \quad \text{and} \quad E[\epsilon_t' \epsilon_t | y_{t-s}, s > 0] = \underset{m \times m}{I}.$$

From this point forward, A_0, the contemporaneous coefficient matrix for the structural model, is assumed to be nonsingular and subject only to linear restrictions.[28]

This structural model can be transformed into a multivariate regression by defining A_0 as the contemporaneous correlations of the series and A_+ as a matrix of the coefficients on the lagged variables by

$$YA_0 + XA_+ = E, \tag{2}$$

where Y is $T \times m$, A_0 is $m \times m$, X is $T \times (mp + 1)$, A_+ is $(mp + 1) \times m$, and E is $T \times m$. Here we have placed the constant as the last element in the respective matrices. Note that the columns of the coefficient matrices correspond to the equations.

Before proceeding, define the following compact form for the VAR coefficients in Eqs. (1) and (2):

$$a_0 = vec(A_0), \quad a_+ = vec\begin{pmatrix} -A_1 \\ \vdots \\ -A_p \\ d \end{pmatrix}, \quad A = \begin{pmatrix} A_0 \\ A_+ \end{pmatrix}, \quad a = vec(A), \tag{3}$$

where A is a stacking of the system matrices and *vec* is a vectorization operator that stacks the system parameters in column-major order for each equation. Note that a is a stacking of the parameters in A.

The VAR model in Eq. (2) can then be written as a linear projection of the residuals by letting $Z = [Y \; X]$, and $A = [A_0 | A_+]'$ is a conformable stacking of the parameters in A_0 and A_+:

$$YA_0 + XA_+ = E \tag{4}$$
$$ZA = E. \tag{5}$$

---

[28]We use the term identified or "structural," in a manner consistent with the VAR literature, to denote a model that is a dynamic simultaneous system of equations where the A_0 matrix is identified. The model is structural in that its interpretation and estimation require us to make an assumption about the structure of A_0, the decomposition of the reduced form error covariance matrix. In what follows, we assume that this matrix is "just identified" in the sense that A_0 is a triangular Cholesky decomposition of the covariance matrix of the residuals. See Leeper et al. (1996) and Sims and Zha (1999) for a discussion of alternatives such as over-identified models.

In order to derive the Bayesian estimator for this structural equation model, we first examine the (conditional) likelihood function for normally distributed residuals:

$$L(Y \mid A) \propto |A_0|^T \exp[-0.5 tr(ZA)'(ZA)] \tag{6}$$

$$\propto |A_0|^T \exp[-0.5 a'(I \otimes Z'Z)a], \tag{7}$$

where $tr(\ )$ is the trace operator. This is a standard multivariate normal likelihood equation.

Sims and Zha next propose a conditional prior distribution for this model. Note that since this is a structural equation time series model, the prior will be on the *structural* parameters, rather than on the reduced form as proposed by Litterman (more on this below).

The Sims-Zha prior for this model is formed conditionally. Sims and Zha assume that for a given A_0, or contemporaneous coefficient matrix (stacked in a_0), the prior over all of the structural parameters has the form

$$\pi(a) = \pi(a_+ \mid a_0)\pi(a_0) \tag{8}$$

$$\pi(a) = \pi(a_0)\phi(\widetilde{a_+}, \Psi), \tag{9}$$

where the tilde symbol $^\sim$ denotes the mean parameters in the prior for a_+, Ψ is the prior covariance for $\tilde{a}_+$, and $\phi(\ )$ is a multivariate normal density. For now, we leave the prior on the contemporaneous coefficient matrix, $\pi(a_0)$, unspecified and we assume that conditional on the a_0 elements that the a_+ coefficients have a normally distributed prior.

The posterior for the coefficients is then

$$q(A) \propto L(Y \mid A)\pi(a_0)\phi(\widetilde{a_+}, \Psi) \tag{10}$$

$$\propto \pi(a_0)|A_0|^T |\Psi|^{-0.5} \times \exp[-0.5(a_0'(I \otimes Y'Y)a_0 - 2a_+'(I \otimes X'Y)a_0 + a_+'(I \otimes X'X)a_+ + \widetilde{a_+}'\Psi\widetilde{a_+})]. \tag{11}$$

As Sims and Zha note, this posterior is nonstandard. But it is tractable (unlike the posterior for the Litterman prior) for a special case. When the prior in Eq. (8) has the same symmetric structure as the Kronecker product $I \otimes X'X$ in the likelihood, the posterior is conditionally multivariate normal, since the prior has a conjugate form. In this case, the posterior can be estimated by a multivariate seeming unrelated regression (SUR) model. Thus forecasts and inferences can be generated by exploiting the multivariate normality of the posterior distribution of the coefficients.

The reference prior proposed by Sims and Zha for this model is formed for the conditional distribution $\pi(a_+ \mid a_0)$. This is in contrast to the Litterman approach of formulating the prior on the individual parameters of each equation in the reduced form. This difference is not minor. Forming the prior for the reduced form as in Litterman (1986) requires that the beliefs about the parameters in the covariance matrix for the prior on the coefficients be independent across the equations; this prior is nonconjugate and yields a nontractable posterior. Sims and Zha's prior requires that conditional on the prior for A_0 (contemporaneous correlations in the series), the correlation structure for the regression parameters in the prior is correlated in the same manner as the structural residuals. The result is that the Sims-Zha approach yields a posterior distribution that can be easily sampled, while the Litterman equation-by-equation construction of the prior on the reduced form representation of the model does not (see Sims and Zha 1999 and Kadiyala and Karlsson 1997 for a technical treatment of these points).

Since the residuals of the structural models are standardized to have unit variance, we are working with a prior on "standardized" data. This simplifies the specification, since it removes issues of relative scale and focuses the specification on the dynamics. The Sims-Zha prior is specified by positing a conditional mean for $a_+ | a_0$. The prior mean is assumed to be the same as the Litterman prior: that the best predictor of a series tomorrow is its value today.[29] The unconditional prior has the form $E[a_+] = (I | 0)$ so the conditional prior has the form $a_+ | a_0 \sim N((A_0 | 0), \Psi)$, where these conditional means have the same $mp \times m$ dimension and structure as the A matrix in Eq. (5). Combining these facts, we can write the normal conditional prior for the mean of the structural parameters as

$$E(A_+ | A_0) = \begin{bmatrix} A_0 \\ 0 \end{bmatrix}. \tag{12}$$

The conditional covariance of the parameters, $V(A_+ | A_0) = \Psi$, is more complicated. It is specified to reflect the following general beliefs and facts about the series being modeled:

1. The standard deviations around the first lag coefficients are proportionate to all the other lags.
2. The weight of each variable's own lags is the same as those of other variables' lags.
3. The standard deviation of the coefficients of longer lags are proportionately smaller than those on the earlier lags. (Lag coefficients shrink to zero over time and have smaller variance at higher lags.)
4. The standard deviation of the intercept is proportionate to the standard deviation of the residuals for the equation.
5. The standard deviation of the sums of the autoregressive coefficients should be proportionate to the standard deviation of the residuals for the respective equation (consistent with the possibility of cointegration).
6. The variance of the initial conditions should be proportionate to the mean of the series. These are "dummy initial observations" that capture trends or beliefs about stationarity and are correlated across the equations.

Sims and Zha propose a series of hyperparameters to scale the standard deviations of the dynamic simultaneous equation regression coefficients according to these beliefs. To see how these hyperparameters work to set the prior scale of A_+, remember that $V(A_+ | A_0) = \Psi$ is the prior covariance matrix for $\tilde{a}_+$. Each diagonal element of Ψ therefore corresponds to the variance of the VAR parameters. The variance of each of these coefficients is assumed to have the form

$$\bar{\psi}_{\ell,j,i} = \left(\frac{\lambda_0 \lambda_1}{\sigma_j l^{\lambda_3}} \right)^2, \tag{13}$$

---

[29]This does not mean we are assuming the posterior distribution of the parameters and the data follow a random walk. Instead, it serves as a benchmark for the prior. If it is inconsistent with the data, the data will produce a posterior that does not reflect this belief. We hope to investigate other theoretically derived and consistent specifications for the mean regression coefficients in future work.

Table 2 Hyperparameters of Sims-Zha reference prior

Parameter	Range	Interpretation
λ_0	$[0,1]$	Overall scale of the error covariance matrix
λ_1	> 0	Standard deviation around A_1 (persistence)
λ_2	$= 1$	Weight of own lag versus other lags
λ_3	> 0	Lag decay
λ_4	≥ 0	Scale of standard deviation of intercept
λ_5	≥ 0	Scale of standard deviation of exogenous variable coefficients
μ_5	≥ 0	Sum of coefficients/Cointegration (long-term trends)
μ_6	≥ 0	Initial observations/dummy observation (impacts of initial conditions)
ν	> 0	Prior degrees of freedom

for the element corresponding to the ℓth lag of variable j in equation i. The overall coefficient covariances are scaled by the value of error variances from m univariate AR(p) OLS regressions of each variable on its own lagged values, σ_j^2, for $j = 1, 2, \ldots, m$.[30] The parameter λ_0 sets an overall tightness across the elements of the prior on $\Sigma = A_0^{-1'}A_0^{-1}$. Note that as λ_0 approaches 1, the conditional prior variance of the parameters is the same as in the sample residual covariance matrix. Smaller values imply a tighter overall prior. The hyperparameter λ_1 controls the tightness of the beliefs about the random walk prior or the standard deviation of the first lags (since $l^{\lambda_3} = 1$ in this case). The l^{λ_3} term allows the variance of the coefficients on higher order lags to shrink as the lag length increases. The constant in the model receives a separate prior variance of $(\lambda_0\lambda_4)^2$. Any exogenous variables can be given a separate prior variance proportionate to a parameter λ_5 so that the prior variance on any exogenous variables is $(\lambda_0\lambda_5)^2$.[31] Sims and Zha also propose adding two sets of dummy observations to the data, consistent with Theil's mixed estimation method (Theil 1963). These dummy observations account for unit roots, trends, and cointegration. The parameter $\mu_5 > 0$ is used to set prior weights on dummy observations for a sum of coefficient prior that implies beliefs about the presence of unit roots. The parameter μ_6 is the prior weight for dummy observations for trends and weights for initial observations. Table 2 provides a summary of the hyperparameters in the model.

Several points should be made about this prior. First, it is formulated for the structural parameters—the true parameters of interest in these models. Second, the conditional prior that is a partition of the beliefs about the stacked structural parameters in a (see Eq. [8]) is independent across the equations and thus across the columns of A_+. The interdependence of beliefs is reflected in the structural contemporaneous correlations, A_0. Beliefs about the parameters are correlated in the same patterns as the reduced form or contemporaneous residuals. As such, if we expect large correlations in the reduced form innovations of two equations, their regressors are similarly correlated to reflect this belief and ensure that the series move in a way consistent with their unconditional correlations. This is probably the most important innovation of the prior, since earlier priors proposed for VAR models worked with the reduced form and assumed that the beliefs about the parameters were uncorrelated across the equations in the system (e.g., Kadiyala and Karlsson 1997).

---

[30]This is the only use of the sample data in the specification of the prior. The only reason the data are used is so that the scale of the prior covariance of the parameters is approximately the same as the scale of the sample data.

[31]The prior on any exogenous or deterministic variable coefficients should be set tighter than the prior for the intercept, or $\lambda_5 < \lambda_4$. Otherwise, the exogenous variables will overexplain the variation in the endogenous variables relative to the endogenous variables. We thank the late John T. Williams for clarifying this point for us.

The more common representation is the reduced form VAR model. Writing the model in Eq. (1) in reduced form helps connect the previous discussion to extant VAR texts (Hamilton 1994), multivariate Bayesian regression models (Zellner 1971; Box and Tiao 1973), and the Litterman prior. The reduced form model is

$$y_t = c + y_{t-1}B_1 + \cdots + y_{t-p}B_p + u_t. \tag{14}$$

This is an m-dimensional multivariate time series model for each observation in a sample of size $t = 1, 2, \ldots, T$ with y_t an $1 \times m$ vector of observations at time t, B_ℓ the $m \times m$ coefficient matrix for the ℓth lag, p the maximum number of lags, and u_t are the reduced form residuals. The reduced form in Eq. (14) is related to the simultaneous equation model in Eq. (1) by

$$c = dA_0^{-1}, \quad B_\ell = -A_\ell A_0^{-1}, \quad \ell = 1, 2, \ldots, p, \quad u_t = \epsilon_t A_0^{-1} \quad \text{and} \quad \Sigma = A_0^{-1}{}'A_0^{-1}.$$

The Sims-Zha prior for this model is defined with respect to the normalized simultaneous equation parameters and can be translated to the reduced form.

The matrix representation of the reduced form (analogous to Eq. [2]) is formed by stacking the variables for each equation into columns:

$$\underset{T \times m}{Y} = \underset{T \times (mp+1)}{X} \underset{(mp+1) \times m}{\beta} + \underset{T \times m}{U}, \quad U \sim MVN(0, \Sigma). \tag{15}$$

Here the columns of the matrix β correspond to the coefficients for each equation, stacked from the elements of B_ℓ. Note that the only exogenous variable in this representation is a constant, but extensions with additional exogenous variables pose no difficulties.

We can construct a reduced form Bayesian SUR model with the Sims-Zha prior as follows. The prior means for the reduced form coefficients are that $B_1 = I$ and $B_1, \ldots B_p = 0$. We assume that the prior has a conditional structure that is multivariate normal-inverse Wishart distribution for the parameters in the model. Using this prior for the parameters, denoted $\bar{\beta}$ and $\bar{S}$ for B and Σ, respectively, we estimate the coefficients for the system of Eq. (15) with the following estimators:

$$\hat{\beta} = (\bar{\Psi}^{-1} + X'X)^{-1}(\Psi^{-1}\bar{\beta} + X'Y) \tag{16}$$

$$\hat{\Sigma} = T^{-1}(Y'Y - \hat{\beta}'(X'X + \Psi^{-1})\hat{\beta} + \bar{\beta}'\Psi^{-1}\bar{\beta} + \bar{S}) \tag{17}$$

where the normal-inverse Wishart prior for the coefficients is

$$\beta \,|\, \Sigma \sim N(\bar{\beta}, \bar{\Psi}) \quad \text{and} \quad \Sigma \sim IW(\bar{S}, v). \tag{18}$$

Equation (12) can be used to specify $\bar{\beta}$. The aforementioned univariate AR(p) regression prediction variances are used to determine the diagonal elements of $\bar{S}$. Equation (13) is used to specify the elements of $\bar{\Psi}$. This is the same Bayesian representation of the multivariate regression model found in standard texts (Zellner 1971; Box and Tiao 1973). This representation translates the prior proposed by Sims and Zha from the structural model to the reduced form.

Litterman's prior is formulated for the reduced form coefficients. Litterman assumed that as a baseline, each of the univariate equations in the system followed a random walk, or that the beliefs are centered on $y_{it} = y_{i,t-1} + u_{it}$ for each series i. As such, his prior is centered on the beliefs about the coefficients $B \,|\, A_0$, rather than on $A_+ \,|\, A_0$. Beliefs are uncorrelated

across the equations and depend explicitly on the reduced form representation of the parameters. This equation-by-equation formulation of the prior then has the form of $\beta \,|\, \Sigma \sim N((I \,|\, 0), M)$, where M is a diagonal matrix of the prior beliefs about the variance of the parameters. In contrast, as indicated in Eqs. (12) and (13), for the Sims-Zha prior the conditional prior is uncorrelated, but the unconditional prior will be correlated across equations (in the same pattern as A_0). Sims and Zha's prior does this by having the correlations of the parameters across the equations match the correlations of the reduced form innovations. It therefore alters the treatment of own-versus-other lags in each equation of the Litterman prior.

What is the benefit of the Sims-Zha prior for political science and international relations research? Our experience in analyzing conflicts and other kinds of data leads us to believe that belligerents reciprocate each other's behavior. We believe that the signs and magnitudes of coefficients in equations describing directed behaviors are related. How one country behaves toward another reflects how that other country behaves toward it. Theories, forecasts, and policy analyses that incorporate this belief in reciprocity offer the best accounts of international conflicts. With the Sims-Zha prior we can explore the possibility that beliefs about the correlations in the innovations in these equations are reflective of this idea. By conditioning our beliefs on these correlations, for the first time we have a reference prior that embodies the belief in reciprocity.

3.2 *Innovation Accounting and Error Band Construction*

Innovation accounting consists of computing the responses of the variables $y_{it} = y_i(t)$ for a specified shock of ϵ_j to variable j. Here we change notation to highlight how the responses are functions of time. These responses are typically found by inverting the VAR model to a moving average representation. This is done to compute the response, for a shock to the system in Eq. (1):

$$c_{ij}(s) = \frac{\partial y_i(t+s)}{\partial \epsilon_j(t)}, \tag{19}$$

where $c_{ij}(s)$ is the response of variable i to a shock in variable j at time s.

The c_{ij} coefficients are the same as the moving average representation coefficients for the dynamic simultaneous equation or VAR(p) model. We define a matrix version of the C coefficients using

$$(A_0 + A_1 L + A_2 + L^2 + \cdots + A_p L^p)y_t = \epsilon_t \tag{20}$$

$$A(L)y_t = \epsilon_t, \tag{21}$$

where L is the lag operator and $A(L)$ defines the matrix lag polynomial in Eq. (20). The impulse response coefficients are then $C(L) = A^{-1}(L)$.

Note several facts about these impulse responses (Sims and Zha 1999, pp. 1122–1124). First, they provide a better, more intuitive representation of the dynamics of the series in the model than the AR representation. Second, the c_{ij} coefficients are a function of time and provide a good method for seeing how the multivariate process behaves over time. Third, constructing measures of uncertainty for the $c_{ij}(t)$ is difficult. Also, the $c_{ij}(t)$ are high dimensional and thus hard to summarize.[32]

---

[32]Technically, the mapping from the matrix A to the matrix C is one to one, but the mapping for the individual a_{ij} to c_{ij} is not, in general, one to one. The subsequent nonlinearity of the responses means that approximations based on linearization and asymptotic normally perform poorly.

As we explained earlier, several methods have been proposed for measuring the uncertainty or error bands for the responses in Eq. (19). Analytic derivatives and related normal asymptotic expansions for these responses are presented in Lutkepohl (1990) and Mittnik and Zadrozny (1993). The approximations from these derivative methods tend to perform poorly as the impulse response horizon is increased. In addition, Kilian (1998) presents a "bootstrap after bootstrap"-based confidence interval for impulse responses. This "bootstrap-corrected" method reduces the bias in the initial estimates of the A coefficients. But it does not adequately account for the non-Gaussian, nonlinear, highly correlated aspects of the responses (again, see the discussion in Sims and Zha [1999], pp. 1125–1127).

The standard approaches to computing the error bands are based on constructing the following interval:

$$\hat{c}_{ij}(t) \pm \delta_{ij}(t), \tag{22}$$

where $\hat{c}_{ij}(t)$ are the mean estimated response at time t and distances $\pm\delta_{ij}(t)$ are the upper and lower bounds of the confidence intervals relative to the mean. These bands are presented by plotting the three functions, $\hat{c}_{ij}(t) - \delta_{ij}(t), \hat{c}_{ij}, \hat{c}_{ij}(t) + \delta_{ij}(t)$, as functions of t. These are effectively known as "connect the dots" error bands and are a standard output from common statistical software (RATS, Eviews, Stata).

There are several ways to compute the error bands and the functions $\delta_{ij}(t)$ (Runkle 1987). One is to take a Monte Carlo sample from the posterior distribution of the VAR coefficients (defined above). From this sample, we then compute a normal approximation to the $\hat{c}$:

$$\hat{c}_{ij}(t) \pm z_\alpha \sigma_{ij}(t), \tag{23}$$

where z_α are the normal pdf quantile and $\sigma_{ij}(t)$ is the standard deviation of the $c_{ij}(t)$ at time t ($z_\alpha = 1$ for 68% bands and $z_\alpha = 1.96$ for 95% bands). This method assumes that the impulse responses are normal in small samples.

While this Gaussian approximation approach originally was available in RATS, an alternative (used currently in RATS) is to calculate the quantiles of the $c_{ij}(t)$ for each response and time point. We then estimate the posterior interval based on the highest posterior density region or pointwise quantiles, namely,

$$[c_{ij.\alpha/2}(t), c_{ij.(1-\alpha)/2}(t)], \tag{24}$$

where the subscript $\alpha/2$ denotes the bounds of the $1-\alpha$ confidence set or interval, computed by taking the empirical quantiles. Yet if the $c_{ij}(t)$ are serially correlated, then the $\delta_{ij}(t)$ and $c_{ij.\alpha/2}(t)$ are likely to be as well. Thus these quantiles will fail to account for the serial correlation in the responses and they will have incorrect posterior coverage probabilities.

To solve these problems, Sims and Zha (1999) estimate the variability of the impulse responses by accounting for the likely serial correlation in the responses. Consider the responses for a single variable i with respect to a shock in variable j over H periods. Denote this vector by the sequence $\{c_{ij}(t)\}_{t=0}^{H}$. A sample of these sequences of responses can be generated using standard methods by sampling from the posterior density of the VAR coefficients and computing the responses. For these H responses, we can compute an $H \times H$ covariance matrix Ω that summarizes the variance of the response of variable i to shock j with respect to time. This is done separately for each of the m^2 impulse responses,

that is, for $i = 1, \ldots, m$ and $j = 1, \ldots, m$. The benefit of using the m^2 covariance matrices Ω for computing the variance of the impulses is that they capture the serial correlation of the responses. The variation of the responses can then be analyzed using the following eigenvector decomposition:

$$\Omega = W'\Lambda W \tag{25}$$

$$\Lambda = diag(\lambda_1, \ldots, \lambda_H) \tag{26}$$

$$WW' = I. \tag{27}$$

The H-dimensional c_{ij} vector can be written as

$$c_{ij} = \hat{c}_{ij} + \sum_{k=1}^{H} \gamma_k W_{\cdot k}, \tag{28}$$

where $\hat{c}_{ij}$ are the mean c_{ij} vector for each of the H periods, γ_k are the coefficients for the stochastic component of each response, and $W_{\cdot k}$ is the $k'th$ eigenvector of W. The variation around each response is generated by the randomness of the γ coefficients. The variances of the γ are the eigenvalues of the decomposition. The decomposition of the responses in Eq. (28) describes the responses in terms of the principal components of their variance over the response horizon as linear combinations of the main components of this variance.

The main variation in the impulse responses can be summarized using this decomposition by constructing the interval

$$\hat{c}_{ij} + z_\alpha W_{\cdot,k} \sqrt{\lambda_k}, \tag{29}$$

where $\hat{c}_{ij}$ is the mean response of variable i to shock in variable j, z_α are the normal pdf quantiles, $W_{\cdot,k}$ is the kth eigenvector of the decomposition of Ω, and λ is the eigenvalue of the kth eigenvector. This Gaussian linear eigenvector decomposition of the error bands characterizes the uncertainty of the response of variable i to a shock in variable j in terms of the principal sources of variation over the response horizon. If the kth eigenvalue explains $100 \cdot (\lambda_k / \sum_{i=1}^{H} \lambda_i)$ percent of the variance, then this band will characterize that component. Note that this method assumes that the responses are joint normal over the H periods. Further, these bands still assume symmetry.

To better characterize the uncertainty about the impulse responses, we can look at the quantiles of this decomposition. This may be preferred because the assumption that the error bands are joint normal will likely not hold as the impulse response horizon increases. To compute these likelihood-based (or Bayesian) error bands, we take the Monte Carlo sample of the c_{ij} and compute the quantiles of the γ_k, which summarize the main variation in the c_{ij}. This is done first computing Ω for each impulse, then computing $\gamma_k = W_{k,\cdot} c_{ij}$, where the $W_{k,\cdot}$ are computed from the eigenvector decomposition and the γ_k are then estimated from each of the responses in the Monte Carlo sample of responses. The quantiles of the γ_k across the Monte Carlo sample can be used to construct error bands. Typical quantiles will be one and two standard deviation error bands, or 16–84%, and 2.5–97.5%. We will generally use the rows of $W_{k,\cdot}$ that correspond to the largest eigenvalues of Ω. The bands constructed in this manner will account for the temporal correlation of the impulses:

$$\hat{c}_{ij} + \gamma_{k,0.16}, \quad \hat{c}_{ij} + \gamma_{k,0.84}. \tag{30}$$

Table 3   Impulse response error band computations

Error Band Method	Error Band Interval
Gaussian approximation	$\hat{c}_{ij}(t) \pm z_\alpha \sigma_{ij}(t)$
Pointwise quantiles	$[c_{ij.\alpha/2}(t), c_{ij.(1-\alpha)/2}(t)]$
Gaussian linear eigenvector	$\hat{c}_{ij} \pm z_\alpha W_{.k}\sqrt{\lambda_k}$
Likelihood-based eigenvector	$\hat{c}_{ij} + \gamma_{k,0.16}, \quad \hat{c}_{ij} + \gamma_{k,0.84}$
Likelihood-based stacked eigenvector	$\hat{c}_{ij} + \gamma_{k,0.16}, \quad \hat{c}_{ij} + \gamma_{k,0.84}$
	(with γ_k computed from the stacked covariance)

As such, these bands assume neither symmetry nor normality in the impulse response density. Their location, shape, and skewness are more accurate than bands produced by other methods because they can account for the asymmetry of the bands over the time horizon of the responses.

Finally, we could construct error bands for all of the responses over all of the time periods. For this method, instead of stacking the temporally correlated impulses for each response (as in the computation for Eq. [29]), we stack all m^2 responses for all H periods and compute Ω based on the stacked m^2H responses. This stacked eigenvector decomposition then accounts for the correlation across time in the responses and across the responses themselves. This is appropriate if our series are highly contemporaneously correlated.

In what follows we use the notation and terminology in Table 3 to describe the error bands computed for our impulse responses. We show below that error bands computed using the eigenvector decomposition methods suggested by Sims and Zha provide a better summary of the shape and likelihood of the responses than the alternatives.[33]

3.3   *Forecasting and Policy Analysis*

Sims (1980) notes that one of the major advantages of reduced form multiple equation time series modeling such as VARs is their applications to forecasting and policy analysis. We believe that for the analysis of international conflict and other subjects in political science both of these advantages are present. First, we want to know the trend or overall direction of conflict in the future based on the recent past. Second, we want to know the impact of feasible policy intervention. This type of counterfactual analysis is not easy, however— the presence of dynamic policy rules, and dynamic systems of equations such as those proposed in Eq. (1), lead to complicated forecasting and conditional forecasting problems.

Doan et al. (1984) note that we may know the path of one endogenous variable in a dynamic system of simultaneous equations before we see another (such as unemployment, which is measured monthly, while GNP remains unobserved until the end of the quarter). We also could hypothesize alternative paths for a policy variable such as the level of U.S. mediation or trade sanctions in an international conflict and then look at the resulting forecasts of the conflict. In both cases, we are placing a set of constraints on the forecasts we can make because the estimated error covariance determines the correlation between the forecasts for the variables in the VAR. This idea led Doan et al. to derive the set of linear conditions on forecast innovations implied by the simulated path of policy variables.

---

[33]In what follows, we do not employ the stacked eigenvector decomposition method for all the responses in the system. We present it because in some applications where there is a high contemporaneous correlation in some of the responses it may be a better method. Note, however, that this method is highly computationally intensive, since it requires an eigendecomposition of a m^2H square matrix.

Waggoner and Zha (1999) extend this idea and show how to derive the Bayesian posterior sample based on the mean and variance of these constrained forecasts. They demonstrate how to use information about the forecasts' innovations subject to constraints on the forecast of one or more endogenous variables in a VAR to generate conditional forecast distributions that correctly account for both parameter uncertainty and forecast uncertainty. Waggoner and Zha do this by using Gibbs sampling with data augmentation to generate a sequence of model estimates and forecasts that summarize the conditional forecasts and their associated uncertainty.

In this subsection, we stress policy counterfactuals. But this analysis also applies to historical counterfactuals (inquiries into the hypothetical effects, *ex post*, of a counterfactual path of a variable in time past). There are two ways we can proceed to construct such policy counterfactuals. The first uses a *hard condition* to specify the path of a given endogenous variable. A hard condition sets the value of an endogenous variable to a fixed value or path of values. Alternatively, we could use a *soft condition* and posit a range of values for this policy variable. For instance, a hard condition for an international conflict model assumes that a policy innovation—a surge in cooperation of a third party toward one of the two belligerents, for instance—remains at a fixed level for some time into the future. A soft condition assumes that this policy shock takes on one of a range of values over some future horizon. This is a Bayesian implementation of the analysis of sequences of policy innovations.[34]

Formally, consider an h-step forecast equation for the reduced form VAR model:

$$y_{T+h} = cK_{h-1} + \sum_{l=1}^{p} y_{T+1-l}N_l(h) + \sum_{j=1}^{h} \epsilon_{T+j}C_{h-j}, \quad h = 1, 2, \ldots \quad (31)$$

where

$$K_0 = I, \quad K_i = I + \sum_{j=1}^{i} K_{i-j}B_j, \quad i = 1, 2, \ldots;$$

$$N_l(1) = B_l, \quad l = 1, 2, \ldots, p;$$

$$N_l(h) = \sum_{j=1}^{h-1} N_l(h-j)B_j + B_{h+l-1}, \quad l = 1, 2, \ldots, p, \quad h = 2, 3, \ldots;$$

$$C_0 = A_0^{-1}, \quad C_l = \sum_{j=1}^{i} C_{i-j}B_j, \quad i = 1, 2, \ldots,$$

where we use the convention that $B_j = 0$ for $j > p$, C_l are the impulse response matrices defined in the last section for lag l, K_i describe the evolution of the constants in the forecasts, and $N_l(h)$ define the evolution of the autoregressive coefficients over the forecast horizon. This h-step forecast Eq. (31) gives the dynamic forecasts produced by a model with structural innovations. It shows how these forecasts can be decomposed into the components with and without shocks. The first two terms in Eq. (31) are the sum of the effects of the past lagged values of the series and the constant or trends. The final term is the impulse responses that determine the relationships among the (policy) innovations that affect the series. The C_i

---

[34]The canonical example here is monetary policy where the Federal Reserve Funds rate (FFR) is fixed at a given value as part of a policy rule (hard condition) or a range of values greater than some level is examined (a soft condition). In both cases, the forecast paths are traced out to see the effects on GNP and the economy at large. See Waggoner and Zha (1999). For a political science application, see Goldstein and Freeman (1990, chap. 5).

matrices are the impulse responses for the forecasts at periods $i = 0, \ldots, h$ where the impulse at time 0 is the contemporaneous decomposition of the forecast innovations.[35]

The key point in conditional forecasting is that setting the path of one variable, say y_{1t}, constrains the possible innovations in the forecasts of $y_{2t} \ldots y_{mt}$. To see this, consider the following formulation for a hard condition on a VAR forecast. Suppose that the value of the $j'th$ variable forecast is constrained to be $y(j)^*_{T+h}$. Then from Eq. (31) it follows that

$$y(j)^*_{T+h} - cK(j)_{h-1} - \sum_{l=1}^{p} y_{T+1-l} N_l(h)(j) = \sum_{j=1}^{h} \epsilon_{T+j} C_{h-j}, \qquad (32)$$

where the notation (j) refers to the $j'th$ column matrix.

The left-hand side of Eq. (32) implies that the innovations on the right-hand side are constrained. That is, there is a restricted parameter space of innovations that are consistent with the hypothesized conditional forecast. These constraints can be expressed as a set of encompassing conditions. These hard conditions take the form of linear constraints:

$$\underset{q \times 1}{r(a)} = \underset{q \times k}{R(a)'} \underset{k \times 1}{\epsilon}, \quad q \leq k = mh, \qquad (33)$$

where $R(a)$ are the stacked impulse responses—the C matrices in Eq. (32)—for the constrained innovations and $r(a)$ are the actual constrained innovations (the left-hand side of Eq. [32]). The elements of these matrices correspond to the forecast constraints. The notation assumes that there are q constraints, and there can be no more constraints than the number of future forecasts for all the variables, $k = mh$. In any case, the elements of R and r may depend on an estimated parameter of the reduced form, denoted by a (the vectorized coefficients in Eq. [3]).[36]

This last fact leads us to use a Gibbs sampling technique to generate the distribution of the conditional forecasts. Gibbs sampling allows us to account for the path of the conditional shocks and the possible uncertainty surrounding the parameters used to generate the respective conditional forecasts. We start by estimating a BVAR model based on the Sims-Zha prior and generating a conditional forecast from this model. We then use this conditional forecast to augment the data and resample the parameters. This procedure accounts for both sources of uncertainty in the forecasts.[37] We explain this Gibbs sampling algorithm and its notation in the appendix.

---

[35]This raises the issue about the properties of the model for different decompositions—the same issue present in the ordering of the responses in impulse response analysis. For just identified VAR models—such as those we are discussing in the section—the choice of this decomposition for the computations is invariant to the ordering of the variables. See the discussion in Waggoner and Zha (1999).

[36]For a soft condition, $r(a)$ is not a vector but a set that contains the admissible forecast values for the forecast condition on the $j'th$ variable. See Waggoner and Zha (1999) for a discussion.

[37]As Waggoner and Zha (1999, pp. 642–643) note, the sampling of the model parameters is

> a crucial step for obtaining the correct finite-sample variation in parameters subject to a set of hard conditions in constraints. . . . Because the distribution of parameters is simulated from the posterior density function, the prior plays an important role in determining the location of the parameters in finite samples. Under the flat prior, the posterior density is simply proportional to the likelihood function, which, in a typical VAR system, is often flat around the peak in small samples. Moreover, maximum-likelihood estimates tend to attribute a large amount of variation to deterministic components (Sims and Zha 1998). Such a bias, prevalent in dynamic multivariate models like VARs, is the other side of the well known bias toward stationarity of least-squares estimates. These problems can have substantial effects on the distribution of conditional forecasts.

4 Illustration

The conflict between the Israelis and Palestinians is one of the most enduring of our time. For decades these two peoples have battled one another. Since the end of World War II, the United States has been involved in this conflict. For the United States, however, solving the Israeli-Palestinian conflict seems tantamount to "moving mountains."

Political scientists have studied this conflict for many years. Among the recent quantitative investigations of it are Schrodt et al. (2001) and Goldstein et al. (2001). Both these studies employ the Kansas Events Data System; each uses WEIS codes. Schrodt et al. is a collection of exploratory analyses of the impacts of third-party intervention on the behavior of the belligerents in the time period April 1979–September 1999. They use frequentist regression and cross-correlation methods to analyze the conflict. Schrodt et al. find evidence that U.S. intervention is motivated by and has a salutary impact on Israeli-Palestinian relations. Multi-equation time series models are used by Goldstein et al. These researchers find evidence of "triangularity" between U.S. behavior toward Israel and the Palestinians, Israeli behavior toward the Palestinians, and Palestinian behavior toward Israel: "Israeli and Palestinian behaviors were reciprocal, indicating that cooperation or conflict received from the United States was 'passed along' in kind to the neighbor" (2001, p. 612). This triangularity provides the basis for the evolution of cooperation between the Israelis and the Palestinians. In other words, it demonstrates, according to Goldstein et al., the potential for effective U.S. intervention in this conflict.

The Bayesian multi-equation time series methods introduced here can improve these and other studies of the Israeli-Palestinian conflict. BVAR models offer three advantages over the approach of Schrodt et al.: the ability to analyze more complex, simultaneous causal relationships between the actors behaviors, systematically incorporate beliefs about conflict dynamics, and gauge the degree of uncertainty about causal inferences. Because their model is essentially an unrestricted VAR with a flat prior, our BVAR model improves the analysis in Goldstein et al. in many of the same ways. Above all, it provides, for the first time, measures of uncertainty for those investigators' causal inferences. Finally, we use our BVAR model to generate forecasts, including forecasts of the policy contingent type. Neither of these studies attempts to produce forecasts of any kind, let alone provide measures of uncertainty for forecasts.

To illustrate these strengths of the BVAR model, we reanalyze the Israeli-Palestinian conflict in the period between April 15, 1979, and December 14, 1988. The latter date is when Yasser Arafat met U.S. demands to renounce all forms of terrorism and accept United Nations Resolutions 242 and 338 (Gerner et al. 1994, pp. 142–144; Morris 2001, pp. 608–610).

The data we use here are from the Kansas Event Data System (KEDS). We employ weekly measures of Israeli, Palestinian, and U.S. directed behaviors, measures derived from the KEDS Levant dataset. We extracted the events involving the United States, the Palestinians, and the Israelis. We then scaled these events and aggregated them into weekly totals. The KEDS data were scaled into interval data using the scale created by Goldstein (1992). This produces a set of six variables: A2I, A2P , I2A , P2A , I2P , P2I, where A = American, P = Palestinian, and I = Israeli. So, for instance, I2P denotes the scaled value of Israeli actions directed toward the Palestinians.

Our analysis is divided into two parts. First we analyze the dynamics of this conflict in a way that takes into account the serial correlation over time in uncertainty about causal inference. We illustrate the value of the eigenvector decomposition method for constructing error bands for impulse responses. To simplify the exposition, we use a flat prior

BVAR model in this analysis.[38] We then use a BVAR model with a modified Sims-Zha prior (which allows for beliefs to be correlated across equations of the reduced form model in a way that reflects the contemporaneous relationships between the actors' behaviors) to produce *ex post* forecasts for the 12 weeks following Arafat's capitulation. We also produce a counterfactual, (hard) policy contingent forecast for the same 12 weeks under the (counterfactual) assumption of sustained U.S. cooperation toward the Israelis.[39]

4.1 *Bayesian Error Bands*

Users of VAR models usually base their causal inferences on impulse responses. For our six-variable system, there are $6 \times 6 = 36$ such responses. Since many of these responses are not of direct interest, we focus on the subset of responses of Israel and Palestine to each other. That is, we focus on the four dyadic responses: responses of Israeli (Palestinian) actions toward the Palestinians (Israelis) to a positive or cooperative shock in Israeli behavior toward the Palestinians, and responses of the Palestinians (Israelis) toward the Israelis (Palestinians) from a positive shock in Palestinian actions toward the Israelis.[40] Our impulse response analyses are based on a flat prior BVAR model because we want to illustrate methods for constructing the error bands separate from the implications of the choice of the prior.

The impulse responses and their error bands are all based on a Monte Carlo sample of 5000 (not antithetically accelerated) draws. For all the moving average responses, the same procedure is used to draw the sample of impulse responses. A sample is taken from the posterior of the (B)VAR models coefficients. The draw is then used to compute the error bands for that draw. These impulses are then saved and summarized using the methods described earlier. The main difference in the results is in the methods used to construct the error bands. All figures have 95% or approximately two standard deviation error bands.

Figure 1 shows three different sets of error bands. The rows in this figure are the responses of the variable on the left axis. The columns correspond to the variable that has been shocked with a positive one standard deviation innovation. Each 2×2 cluster is therefore the same set of responses but with error bands computed by the different methods in Table 3.

The "Normal Approximation" columns use the standard approach of treating the responses as though they are joint normally distributed. The error bands computed using this method tend to be quite large and are symmetric by design. The high degree of (incorrect) uncertainty in the later periods of the response horizon tend to dominate any inferences, making it appear as though there are no significant reactions to the shocks.

The "Pointwise Quantile"-based error bands do not assume that the responses are normally distributed. These error bands are computed using the quantiles of the responses at each point in time. These error bands show a large degree of uncertainty as well. For

---

[38]This is similar to estimating a frequentist model, since where the prior is assumed to have a large variance the posterior estimates are nearly identical to the maximum likelihood estimates.

[39]Our origination date thus is the same as that used by Schrodt et al. (2001) and Goldstein et al. (2001). But our series terminated at December 15, 1988, the date on which Arafat met U.S. demands. The period of the forecasts is December 16, 1988–March 15, 1988. Note that this estimation period is after the Camp David Accords and before the Madrid conference, the Oslo Accords, and the Gulf War. This period also is one in which there were unity governments in Israel and the PLO was arguably more unified than it is today. The U.S. government was, at least as compared to the Nixon administration, more unified as well. We thank Phil Schrodt for his advice on the selection of this time period and choice of the policy counterfactual.

[40]The ordering of the decomposition of the innovations we use to generate the impulse responses is as follows: A2I, A2P, I2A, P2A, I2P, P2I. We put the American-related dyads at the top of the ordering because we are interested in the impacts of U.S. policy on the Palestinian-Israeli conflict.
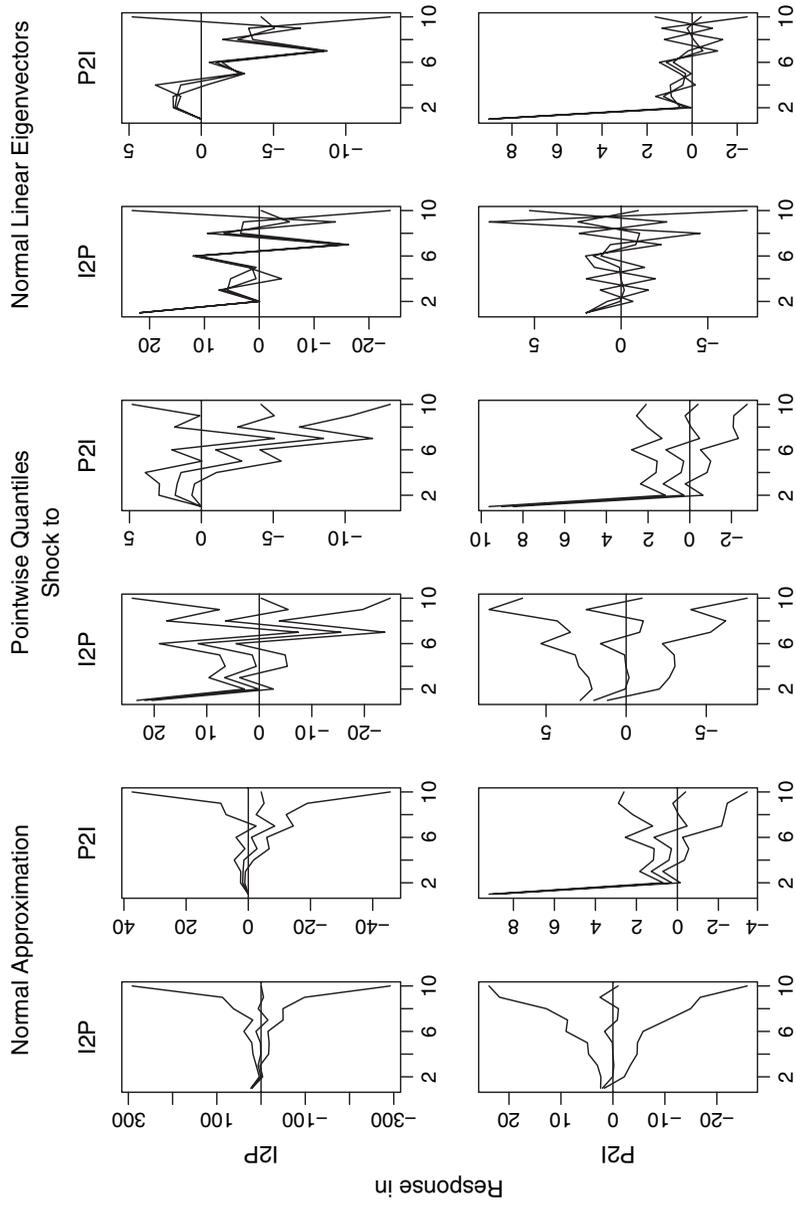
Fig. 1 Selected impulse responses for shocks to I2P and P2I from the six-variable system. Response variables are listed at the left. Columns are the variables that are shocked. Each 2 × 2 cluster of graphs illustrates the subset of four responses for these two variables. Error band computation method for each cluster is listed at the top. Posterior error band regions are 95% intervals. Responses are based on a flat prior BVAR model.

instance, the response of I2P for a shock to P2I appears to be little different from zero for the 12-week horizon. These error bands, however, do more clearly show the shape of the four impulse responses.

The "Normal Linear Eigenvector" decomposition bands are based on the first new method suggested by Sims and Zha (1999). In this case, we use the eigenvector decomposition of the impulse response variances but assume that the impulses are still joint normally distributed over the response horizon. The error bands for these responses are rather nonsensical, since at some points the posterior probability regions nearly collapse to the mean. In general, this is evidence that the normality approximation is a poor choice.

Figure 2 shows the preferred Bayesian shape error bands for impulse responses, that is, the likelihood-based eigenvector quantiles. This method of computing the error bands does not impose a normality assumption. It accounts for the main temporal correlation in the responses. We present the first three components of the eigenvector decomposition. Table 4 reports the percentage of the variance in the responses explained by each of them. The three components account for between 63% and 83% of the total variance in the responses, with the first component accounting for most of the variance.

Several interesting results about the posterior distribution of the responses emerge from these Bayesian shape error bands. The first eigenvector component explains the bulk of the variance in the overall shape of the responses. Here, unlike in the earlier sets of responses, we see that the impact of a positive shock in P2I on I2P is an immediate increase in cooperation, followed by additional hostility (the response of I2P is first positive, then negative). Further, the 95% posterior region for this pattern does not always include zero, thus lending credibility to this interpretation of the dynamic response of an innovation in P2I. We see from the second and third components of the variance of the response in I2P that there is a considerable amount of uncertainty about its symmetry and about the initial positive response of the Israelis toward the Palestinians. In the second component, the mean response of I2P appears to be closer to the lower edge of the 95% interval in the earlier period and closer to the upper edge when the I2P response becomes negative. In the third component, this same response in I2P for a positive shock to P2I appears no different from zero in the early weeks, but it is significantly skewed toward negative (hostile) values after about one month. The sum total of these responses then provides strong evidence for Israeli reciprocity toward the Palestinians in the first month after a surprise cooperative action by the Palestinians, but this reciprocity is short-lived.

The response of the Palestinians to a surprise shock of cooperation by the Israelis toward them is very uncertain in Fig. 1. But in Fig. 2, the Bayesian shape bands' first component lends support to the central "zig-zag" pattern of this response. Substantively, it appears that the initial reaction (first four weeks) of the Palestinians to a surprise cooperative action by the Israelis is quite flat, but more volatile in the later weeks. But this eigenvector component only accounts for 53% of the total variance in the response. An additional 26% of the variance is accounted for by the second and third components. In these components, there is much more uncertainty about the overall response. The second component shows that there is an asymmetry in the response where the mean response is close to the upper edge of the posterior region. It is more likely that as we move further from the surprise in Israeli cooperation, the Palestinians are more favorably disposed toward the Israelis.

In contrast, similar interpretations are hard to support using any of the error bands in Fig. 1. The "Normal Approximation," "Pointwise Quantile," and "Normal Linear Eigenvector" error bands all have the general shape of the bands in Fig. 2. However, the bands in Fig. 1 misrepresent the uncertainty about the shape of the response likelihood. They miss the asymmetry in the likelihood of the responses insofar as they overstate the degree
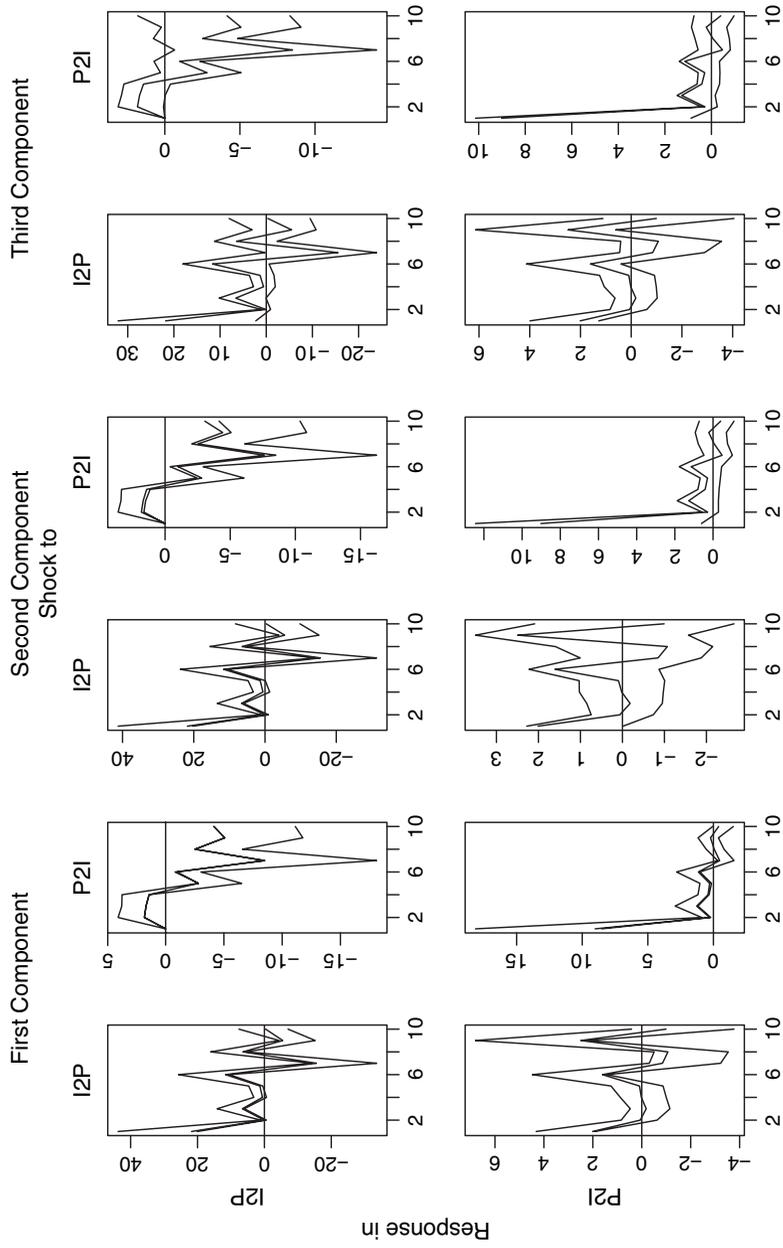
Fig. 2 Impulse responses for shocks to I2P and P2I from the six-variable system. Response variables are listed at the left. Columns are the variables that are shocked. Each 2×2 cluster of graphs illustrates the subset of four responses for these two variables. Posterior error band regions are 95% intervals. Responses are based on a flat prior BVAR model.

Table 4 Percentage of variance in impulse responses explained by each eigenvector
using the likelihood-based method

Shock	Response	Component 1	Component 2	Component 3	Total
I2P	I2P	61	13	10	83
P2I	I2P	50	15	10	75
I2P	P2I	53	15	11	79
P2I	P2I	30	19	13	63

Note. The first two columns define the variable shocked in the system and the observed response. The Total column is the percentage of the variance explained by the first three eigenvectors.

of conflict directed by the Israelis toward the Palestinians in response to a positive shock by the Palestinians toward the Israelis.

4.2 *Forecasting and Counterfactuals*

Forecasting is *the* common standard used in time series modeling. The *fit* of time series models is judged by the in-sample forecasts generated by the model (via one-step error minimization). As such, it seems natural to propose forecast-based methods for assessing model fit and performance. In addition, we show how (B)VAR models can be used for policy evaluation and counterfactual analysis.

We begin our presentation of forecast performance by looking at the benefits of using the Sims-Zha form of a BVAR prior. We forecasted the six data series in our analysis for the periods from 1988:51 to 1989:10 using the sample data from 1979:15 to 1988:50. We used two different models for constructing our forecasts. Both models include six lags. In the first, we employ a flat prior implicit in the maximum likelihood VAR model use by Goldstein et al. (2001). In our second model, we employ a reference prior using the Sims-Zha specification outlined earlier with the following hyperparameters: $\lambda_0 = 0.6$, $\lambda_1 = 0.1$, $\lambda_3 = 2$, $\lambda_4 = 0.5$, and $\mu_5 = \mu_6 = 0$.

The choice of these hyperparameters comes both from "experience" and theory. The selection of the parameters for the prior cannot and should not depend on the data alone—although it should be informed by the properties of the data and their dynamics. If the prior is derived from the data, the resulting forecasts will too closely mirror the sample data rather than the population. However, the prior must be consistent with the data such that it reflects the general beliefs analysts have about the data's variation, dynamic properties, and the general interrelationships of this dyadic conflict. As such, this prior may "work" for forecasting the Israeli-Palestinian data, but it will likely need to be modified when applied to other cases.[41]

We base our design of the prior on several considerations. The first is practical and reflects the properties of event data. We choose to discount the overall scale of the error covariance and the standard deviation of the intercept because we believe that the sample error covariance will overstate the true error covariance. For example, the former puts too much weight on extreme events. In addition, setting the standard deviation of the intercept to be 0.5 reflects the belief that there is a long-run fixed level for the conflict series.[42] Our second consideration concerns the dynamics and the lag structure. Even with six lags, we expect that the effect of events six weeks prior should be rather diffuse. Thus we select a rather rapid lag decay factor of $\lambda_3 = 2$. This means that the variance of the parameters

---

[41]We will be analyzing other international conflicts in future work.
[42]Hence we set $\mu_5 = \mu_6 = 0$. We thank Phil Schrodt for his advice on this aspect of the specification.
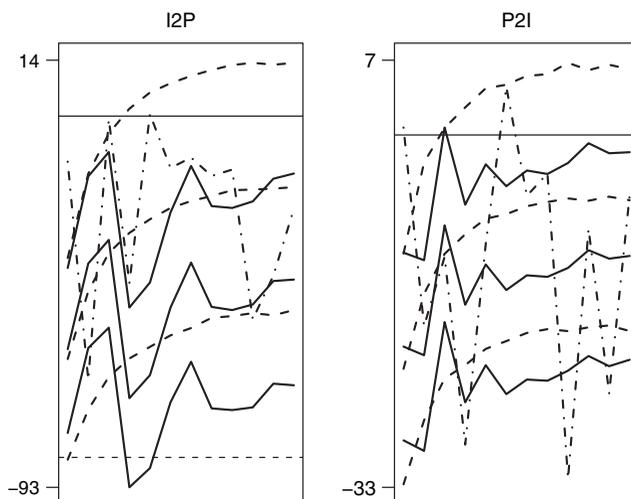
Fig. 3 Comparison of flat and Sims-Zha prior unconditional forecasts for I2P and P2I, 1988:51–
1989:10. Results are based on the six variable VAR models described in the text. Solid lines are the
flat prior forecasts, dashed lines are the reference prior forecast, dotted lines are the actual series.

around lag j are approximately proportionate to j^{-2}. We also choose to place a tighter prior
on the first lag coefficients because we believe that more proximate events are highly
predictive of the conflict events today. Finally, the Sims-Zha prior allows our beliefs about
the model parameters to be correlated across the equations. Thus, if there is correlation in
the residuals of the I2P and P2I equations, the beliefs about the parameters in these two
equations will be similarly correlated. In our case the estimated correlation of the residuals
is 0.21, reflecting our belief in reciprocity.

We believe that these hyperparameters are also roughly consistent with the data. This is
confirmed by a search of the hyperparameter space using the marginal log-likelihood and
log-posterior of the data as measures of fit. The reason we choose not to use a measure
such as the value of the log-posterior pdf of the data or the marginal log-likelihood to
select the prior is that this puts too much weight on the prior. Designing the prior on these
bases only reproduces the density of the sample data. Evidence of this fact is that the
values of the hyperparameters that maximize these measures of posterior fit are all very
"tight." This would be fine for making inferences in-sample, but they do not reflect the
uncertainty we expect to see out-of-sample.[43]

Our illustrative forecast is a challenging one. This is because the week before, Yasser
Arafat proposed a major policy shift for the PLO, renouncing terrorism by the PLO and
accepting U.N. resolutions 242 and 383. As such, this could be a period of structural
change in Israeli-Palestinian-U.S. relations. We return to this possibility in the conclusion.

Figure 3 presents the two sets of forecasts and the actual data for the 12 weeks after
1988:50. Here we present 68% pointwise error bands (approximately one standard
deviation).[44] As can be seen in these bands and forecasts, the forecast of Israeli actions
toward the Palestinians (I2P) indicates more peaceful (more positive) relations after

---

[43]Details of this hyperparameter specification search and the rankings of the hyperparameters by the posterior fit
measures are available in Brandt and Freeman (2002).

[44]We do not use the eigenvector decomposition methods in this example because we want to highlight the
benefits of the Sims-Zha prior itself and not confound the presentation with the Bayesian error band method.
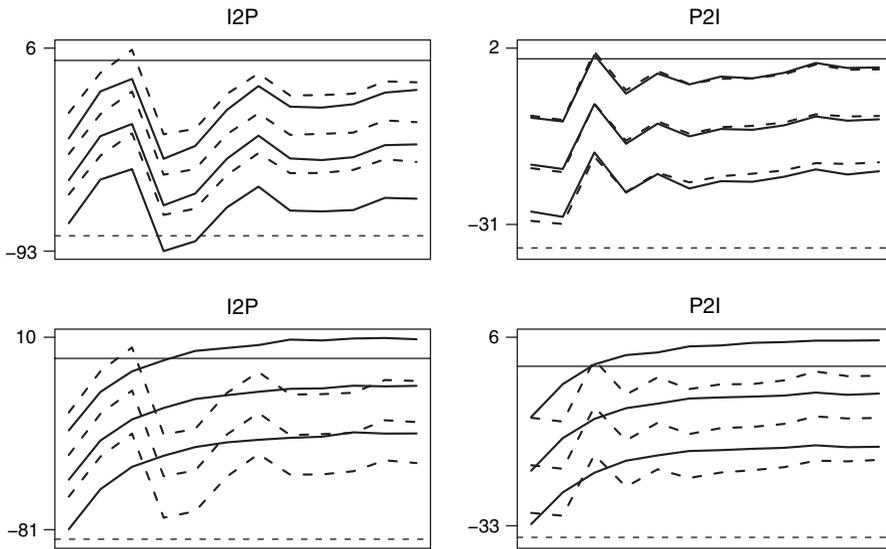
Fig. 4 U.S. policy counterfactual for A2I in the 12 weeks following Arafat's agreement to U.N. Resolutions 242 and 338. Conditional forecasts using flat and Sims-Zha priors, 1988:51–1989:10. Results are based on the six variable BVAR models described in the text. The first row of graphs compares the 12 period conditional (dashed) and unconditional (solid) forecasts using the flat prior. The second row compares the conditional forecasts with the reference prior (solid) to the conditional forecasts with the flat prior (dashed). Confidence regions are the 0.68 probability region, computed pointwise. Dashed lines indicate the value of each series on 1988:50 (last period of the estimation sample).

1988:50. Further, the error bands for the Sims-Zha prior forecast are well above those of the flat prior model. In fact, the flat prior model forecasts tend to be too pessimistic, with many of the actual data points falling above the flat prior forecast confidence region. In contrast, the Bayesian Sims-Zha prior model tends to correctly capture the central tendency over this 12-week horizon. A less clear result is seen for the Palestinian actions toward the Israelis (P2I). Here the reference or Sims-Zha prior model provides superior forecasts in the early weeks. However, in the later weeks, the flat prior model performs better. In this illustration, then, the benefits of the Sims-Zha prior accrue in short- to medium-term forecasts.

To understand the implications of U.S. policy toward the Palestinian-Israeli conflict, we construct counterfactual forecasts. At the time of Arafat's announcement, the Goldstein score for U.S. action toward the Israelis is 9.4, indicating cooperation. Here we consider what would have happened had, for the next 12 weeks, the United States sustained a level of cooperation toward the Israelis that is one standard deviation above the mean of A2I in the forecast period (Goldstein score for A2I = 7.566 for 1988:51–1989:10).

To analyze this policy counterfactual, we employ the two different BVAR models for our system of equations. One is based on a flat prior and one is based on the Sims-Zha prior, with the selection of hyperparameters discussed earlier. Figure 4 compares the conditional and unconditional forecast results for the flat and Sims-Zha prior VAR models. These conditional forecasts and their density summaries were generated using the Gibbs sampling algorithm in the appendix. The summaries are based on a burn-in of 3000 iterations and a final posterior of 5000 values for each series forecasted. There are two

important comparisons: the effects of the prior and the effects of conditioning the forecast of A2I. The first row of graphs presents the I2P and P2I forecasts based on the conditioning of A2I versus no conditioning—both with a flat prior. Here we can see that the forecast condition leads to a modest decrease in the level of conflict between the Israelis and the Palestinians. However, the results are rather diffuse and the confidence regions heavily overlap. Note also that the impact of the "hard" A2I condition has a larger impact on the Israeli actions toward the Palestinians than on the Palestinian actions toward the Israelis.

The second row compares the conditional forecasts with and without the Sims-Zha prior. The first thing of note is that the Sims-Zha prior smoothes out the forecasts considerably (as we would expect from a shrinkage prior like the Sims-Zha prior). In addition, the confidence region for I2P variable includes much more of the positive (cooperative) region when the reference prior is used. Further, after the initial forecast periods, the mean forecast for I2P using the prior is more positive (cooperative) than that without the prior. The failure to employ the reference prior leads one to understate the policy impact of the U.S. policy change.

One counterclaim is that the prior effectively biases the forecasts. In general this could be the case, since the prior is centered near the mean or equilibrium level of the data. However, this alleged bias in the I2P series is in the wrong direction, since the mean value of I2P over the sample period is much lower than the forecasted values. Therefore, we should take the results here as strong evidence that a U.S. policy change in the last weeks of 1988 and early weeks of 1989 could have had a sizable impact on the level of cooperation between the Israelis and Palestinians.

Another way to analyze these forecasts and see the impact of the prior is to look at the conditional distribution of the I2P and P2I series at a specific time point. Here we choose the twelfth or final forecast period, 1989:10. Figure 5 presents several views of the joint distribution of the conditional forecasts of the I2P and P2I series on this date. We refer to this collection of plots as a "mountain plot" because it compares the two bivariate conditional densities (mountains) produced by the flat and Sims-Zha priors. Starting with the bottom right plot and working counterclockwise, we see four views of the densities from the two models. The three-dimensional plot shows that the conditional forecast density for the flat prior model (red hill) sits to the back and right of the Sims-Zha prior conditional forecast (black hill). Since this plot has been rotated so that more pacific Goldstein scores are at the front edges, this plot indicates that the reference prior model forecasts a more pacifying effect for U.S. intervention than the flat prior model.

The two plots on the left show the projection of the forecast densities. The P2I (I2P) figure compares the Sims-Zha prior (black) and flat prior (red) conditional forecasts on the P2I (I2P) dimension. We see that the effect of U.S. intervention is asymmetric insofar as the impact of sustained cooperation from the United States to Israel appears to be greater on I2P than on P2I.[45] For the I2P directed actions, the mean forecasted Goldstein score for the twelfth week is -13 for the reference prior model and -31 for the flat prior model. For the P2I directed dyad, the mean forecasted Goldstein score for the twelfth week is -6 for the (black) Sims-Zha prior model and -11 for the (red) flat prior model.

Finally, the upper right plot shows the contours of the densities. Here we see that the conditional forecast density based on the flat prior model indicates more conflict than that based on the reference prior model because it is lower and slightly more to the left. The

---

[45]Think of these two figures as the projections created by shining a light on one side of the three-dimensional densities in the lower right graph of Fig. 5.
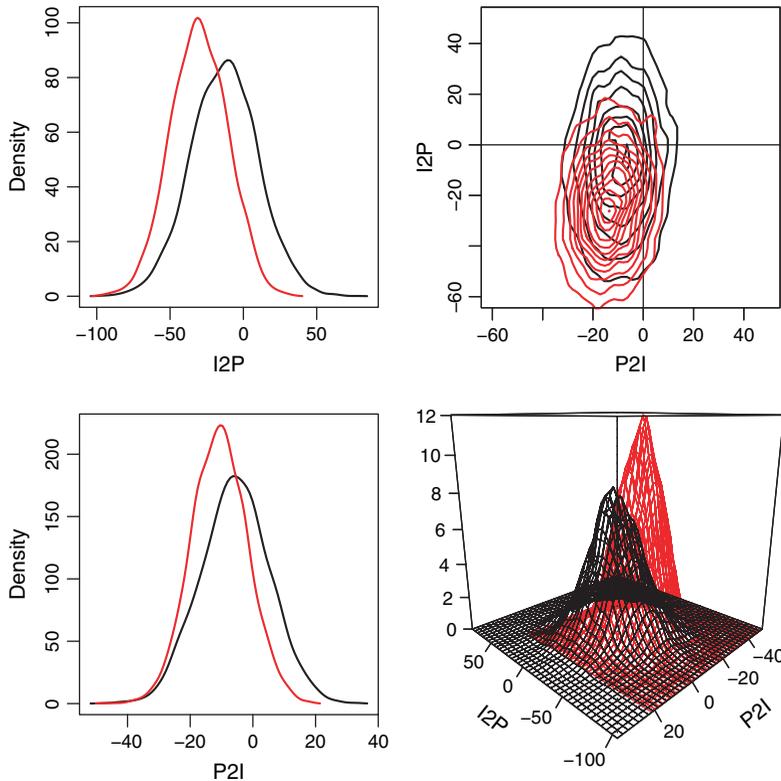
Fig. 5 Mountain plot of the conditional forecast densities for I2P and P2I for 1989:8. The black lines/densities are for the reference prior model. The red lines/densities are for the flat prior model. The red (black) hill is the bivariate density for the flat (reference) prior model. Variables are labeled on the respective axes.

reference prior model shows that the conditional forecasts are nonspherical in the sense that most of the variance in the joint forecasts of I2P and P2I is in the I2P dimension. In contrast, the choice of the prior has little impact on the estimated amount of variation in the P2I variable.

5 Conclusion

Multi-equation time series models have become a staple in political science. With the tools we present here, the Bayesians among us can use these models much more effectively. The Bayesian shape (eigenvector decomposition) method for constructing error bands for our impulse responses gives us a means, for the first time, to gauge the serial correlation over time of uncertainty about our inferences. The modified Sims-Zha prior we outline here is a first step toward developing informed priors for short- and medium-term political forecasting in international relations. The use of such priors will help analysts anticipate outbreaks of violence in places like the Middle East. Finally, we reviewed why, because of politics, policy counterfactuals can be meaningfully evaluated. And we showed how a Bayesian multi-equation model with a modified Sims-Zha prior can be used to gauge the potential impact of third-party intervention in an important international conflict. When further developed, such demonstrations should be of much interest to government

agencies and international (non)governmental organizations. Software to facilitate these methodological innovations will be available.[46]

There are important topics for future research in each of the three areas. Unit roots and cointegration, as we have noted, pose major challenges for causal inference in both the frequentist and Bayesian frameworks. The Sims-Zha prior gives us a starting point for addressing these challenges. We need to explore its usefulness in models that contain variables that we know are first-order integrated either because of theory or our experience analyzing the relevant series. This is part of the focus in the sequel to this article (Brandt and Freeman 2005). In it, we use a macro political economy example to discuss the problem of overfitting in more detail and apply a Sims-Zha reference prior with provisions for unit roots and cointegration. Among the important issues regarding forecasting is the measurement of fit. Econometricians have developed for this purpose concepts like generalized mean square error (Clements and Hendry 1998) and probability integral transform goodness-of-fit tests (Diebold et al. 1998; Clements 2004). The latter, for example, are used to determine if entire forecast densities could have been produced by the respective data generating process. In addition, decision theory needs to be incorporated in evaluations of the kind of Bayesian forecasts of political time series we have illustrated here (see Ni and Sun 2003; Clements 2004).

As for the models themselves, they can be enriched in several ways. Allowing for cointegration leads naturally to Bayesian vector error correction models. And, as suggested by Williams's (1993) original work on the subject, parameters might be time varying. In fact, the I.M.F. is exploring this possibility in its analyses of the impacts of the European Monetary Union (Ciccarelli and Rebucci 2003). When combined with theoretically informed identification of the contemporaneous correlation matrix (A_0), Bayesian time series methods facilitate modeling large-scale systems. In fact, Leeper et al. (1996) show how systems of 13 and 18 variables can be used to study the nature and impact of U.S. monetary policy. Leeper et al.'s approach could prove useful for studying large-scale international conflicts like those in the Levant and Bosnia. Model scale is also discussed in the sequel (Brandt and Freeman 2005). Finally, there is the possibility that, because of recurring changes in the decision rules employed by agents, parameters switch in values between different "regimes." Bayesian Markov switching multi-equation time series models have been developed to account for this possibility (Sims and Zha 2004). Such models may be able to capture the conflict phase sequences and conflict phase shifts international relations scholars have uncovered. If so, we could produce Bayesian conflict phase-contingent impulse responses, forecasts, and contingent forecasts. Work is underway to develop and apply Bayesian Markov switching multi-equation models to the Israeli-Palestinian and several other important international conflicts.[47]

Appendix: Gibbs Sampling Algorithm for Constructing Forecasts

Here we describe the algorithm for calculating conditional forecasts under hard policy counterfactuals. This parallels the discussion in Waggoner and Zha (1999) but with slightly more detail about the steps and the computations for BVAR models with the Sims-Zha prior. We then detail how this algorithm can be used to construct unconditional forecast densities.

---

[46]Details about the software can be found on the *Political Analysis* Web site or by contacting the first author.
[47]Evidence of such switching has been found in the analyses of the impact of politics on currency markets by Freeman et al. (2000) and Hays et al. (2003).

Waggoner and Zha (1999) show that conditional on Eq. (33) in the text and the parameter vector of the VAR (a), the joint conditional h-step forecast distribution is Gaussian with

$$p(y_{T+n} \mid a, \mathbf{Y}_{T+n-1}) = \phi\left(c + \sum_{l=1}^{p} y_{T+n-l}B_l + \mathbf{M}(\epsilon_{T+n})A_0^{-1}; \; A_0^{-1'}\mathbf{V}(\epsilon_{T+n})A_0^{-1}\right) \quad (A1)$$

where $\mathbf{Y}_{T+n-1}$ is the data matrix up to $T + n - 1$. $\mathbf{M}(\epsilon_{T+n})$ and $\mathbf{V}(\epsilon_{T+n})$ are the mean and variance of the constrained innovations under the conditional forecast:

$$p(\epsilon_t \mid a, R(a)'\epsilon = r) = \phi(R(a)(R(a)'R(a))^{-1}r(a); \; I - R(a)(R(a)'R(a))^{-1}R(a)') \quad (A2)$$

With these distributions, the Gibbs sampling algorithm of Waggoner and Zha (1999) becomes:

Let N_1 be the number of burn-in draws and N_2 the number of Gibbs samples after the burn-in. Then:

1. Initialize the values of a_0 and a_+ for the VAR, as defined in Eq. (3). This can be done using either a BVAR or other estimator. These values should come from the peak of $p(a \mid \mathbf{Y_T})$.
2. Generate an *unconditional forecast* $y_{T+1} \ldots y_{T+h}$ based on the draw of a_0 and a_+.
3. For this unconditional forecast, compute the related impulse responses for the coefficients in Eq. (1). These provide the C_i impulse responses.
4. Using the impulse responses that correspond to the unconditional forecast, compute the mean and variance of the constrained innovations and sample the *constrained* or conditional forecast innovations sequence from the density in Eq. (A2). Note that at each iteration one must recompute the value of the mean of ϵ, which depends on r, which in turn depends on a, which is sampled in the Gibbs iterations.
5. Using these constrained innovations, construct the constrained forecasts using the unconditional forecasts according to the reduced form representation in Eq. (31) in the text.
6. Update estimates of a_0 and a_+ for the sample augmented by the h forecast periods. This ensures that the joint density of the (B)VAR parameters reflects the forecast uncertainty. The same estimator should be used at this stage as is used to initialize the sequence of VAR parameters.
7. Repeat the previous steps until the sequence

$$\left\{a^1, y_{T+1}^1, \ldots, y_{T+h}^1, \ldots, a^{N_1+N_2}, y_{T+1}^{N_1+N_2}, \ldots, y_{T+h}^{N_1+N_2}\right\}$$

   is simulated.
8. Keep the last N_2 draws.

As Waggoner and Zha note, the crucial part of the computation is updating the VAR parameters to account for the forecast uncertainty. This then accounts for both the *parameter uncertainty* and the *structural shocks* that are constrained for a conditional forecast. Most existing forecasting inference and forecasting procedures (particularly those that are non-Bayesian) ignore this critical step and therefore take the innovations as the only source of uncertainty.

This same algorithm can be modified to produce unconditional or unconstrained forecasts. that account for both forecast and parameter uncertainty. To construct unconditional forecasts, replace steps 3–5 with a draw from the unconstrained forecast innovations, $\epsilon_t \sim N(0, \Sigma)$. These innovations are used to construct the unconstrained forecasts. The remainder of the algorithm proceeds in the same manner.

Convergence of this Gibbs sampler for these forecasts can be evaluated using standard Markov chain Monte Carlo (MCMC) convergence diagnostics. In particular, we applied the Geweke convergence test for the means in the Markov chain to each forecast period for each variable (Geweke 1992). These results indicated that the Markov chain had converged. Similar conclusions were produced using the Heidelberger and Welch run length control diagnostic test for MCMC convergence (Heidelberger and Welch 1981, 1983).

References

Beck, Nathaniel, Gary King, and Langche Zeng. 2000. "Improving Quantitative Studies of International Conflict: A Conjecture." *American Political Science Review* 94:21–36.

Beck, Nathaniel, Gary King, and Langche Zeng. 2004. "Theory and Evidence in International Conflict: A Response to de Marchi, Gelpi, and Grynaviski." *American Political Science Review* 98(2):379–389.

Box, George E. P., and George C. Tiao. 1973. *Bayesian Inference in Statistical Analysis.* New York: John Wiley and Sons.

Box-Steffensmeier, Janet, and Renee Smith. 1996. "The Dynamics of Aggregate Partisanship." *American Political Science Review* 90(3):567–580.

Brandt, Patrick T., and John R. Freeman. 2002. "Moving Mountains: Bayesian Forecasting As Policy Evaluation." Presented at the 2002 Meeting of the Midwest Political Science Association, Chicago, Illinois.

Brandt, Patrick T., and John R. Freeman. 2005. "Modeling Macropolitical Dynamics." Presented at the Annual Meeting of the American Political Science Association, Washington, DC.

Brandt, Patrick T., and John T. Williams. 2001. "A Linear Poisson Autoregressive Model: The Poisson AR(p) Model." *Political Analysis* 9(2):164–184.

Brandt, Patrick T., and John T. Williams. Forthcoming. *Multiple Time Series Models.* Beverly Hills: Sage.

Brandt, Patrick T., John T. Williams, Benjamin O. Fordham, and Brian Pollins. 2000. "Dynamic Modeling for Persistent Event Count Time Series." *American Journal of Political Science* 44(4):823–843.

Buckley, Jack. 2002. "Taking Time Seriously: The Dynamic Linear Model and Bayesian Time Series Analysis." Unpublished manuscript, SUNY Stony Brook.

Ciccarelli, Matteo, and Alessandro Rebucci. 2003. "Bayesian VARs: A Survey of the Recent Literature with an Application to the European Monetary System." Technical report IMF Working Paper WP/03/102 Washington, DC: International Monetary Fund.

Clements, Michael. 2004. "Evaluating the Bank of England Density Forecasts of Inflation." *Economic Journal* 114:844–866.

Clements, Michael, and David Hendry. 1998. *Forecasting Economic Time Series.* New York: Cambridge University Press.

Cooley, Thomas F., Stephen F. LeRoy, and Neil Raymon. 1984. "Econometric Policy Evaluation: A Note." *American Economic Review* 3:467–470.

DeBoef, Suzanna, and James Granato. 1997. "Near Integrated Data and the Analysis of Political Relationships." *American Journal of Political Science* 41(2):619–640.

deMarchi, Scott, Christopher Gelpi, and Jeffery D. Grynaviski. 2004. "Untangling Neural Nets." *American Political Science Review* 98(2):371–378.

Diebold, F. X., T. A. Gunther, and A. S. Tsay. 1998. "Evaluating Density Forecasts with an Application to Financial Risk Management." *International Economic Review* 39:863–883.

Doan, Thomas, Robert Litterman, and Christopher Sims. 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions." *Econometric Reviews* 3:1–100.

Edwards, George C., and B. Dan Wood. 1999. "Who Influences Whom? The President and the Public Agenda." *American Political Science Review* 93(2):327–344.

Fair, Ray C., and Robert J. Shiller. 1990. "Comparing Information in Forecasts from Economic Models." *American Economic Review* 80(3):375–390.

Fearon, James. 1991. "Counterfactuals and Hypothesis Testing in Political Science." *World Politics* 43:161–195.

Freeman, John R., and James E. Alt. 1994. "The Politics of Public and Private Investment in Britain." In *The Comparative Political Economy of the Welfare State*, eds. Thomas Janoski and Alexander M. Hicks. New York: Cambridge University Press, pp. 136–168.

Freeman, John R., Jude C. Hays, and Helmut Stix. 2000. "Democracy and Markets: The Case of Exchange Rates." *American Journal of Political Science* 44(3):449–468.

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.

Freeman, John R., John T. Williams, and Tse-Min Lin. 1989. "Vector Autoregression and the Study of Politics." *American Journal of Political Science* 33:842–77.

Gerner, Deborah J., Philip A. Schrodt, Ronald A. Francisco, and Judith L. Weddle. 1994. "Machine Coding of Event Data Using Regional and International Sources." *International Studies Quarterly* 38:91–119.

Geweke, John. 1992. "Evaluating the Accuracy of Sampling-Based Approaches to Calculating Posterior Moments." In *Bayesian Statistics*, eds. J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith. Vol. 4. Oxford: Clarendon, pp. 169–194.

Geyer, C. J. 1992. "Practical Markov Chain Monte Carlo." *Statistical Science* 7:473–511.

Gill, Jeffrey. 2002. *Bayesian Methods: A Social and Behavioral Sciences Approach.* Boca Raton, FL: Chapman and Hall.

Gill, Jeffrey. 2004. "Introduction to the Special Issue." *Political Analysis* 12(4):323–337.

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. 1992. "A Conflict-Cooperation Scale for WEIS Event Data." *Journal of Conflict Resolution* 36:369–385.

Goldstein, Joshua S., and John R. Freeman. 1990. *Three-Way Street.* Chicago: University of Chicago Press.

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. 1999. *Empirical Modeling in Economics: Specification and Evaluation.* Cambridge: Cambridge University Press.

Hamilton, James D. 1994. *Time Series Analysis.* Princeton, NJ: Princeton University Press.

Hays, Jude C., John R. Freeman, and Hans Nesseth. 2003. "Exchange Rate Volatility and Democratization in Emerging Market Countries." *International Studies Quarterly* 47:203–228.

Heidelberger, P., and P. D. Welch. 1981. "A Spectral Method for Confidence Interval Generation and Run Length Control in Simulations." *Communications of the A.C.M.* 24:233–245.

Heidelberger, P., and P. D. Welch. 1983. "Simulation Run Length Control in the Presence of an Initial Transient." *Operations Research* 31:1109–1144.

Jackman, Simon. 2000. "Estimation and Inference via Bayesian Simulation: An Introduction to Markov Chain Monte Carlo." *American Journal of Political Science* 44(2):375–405.

Jackman, Simon. 2004. "Bayesian Analysis for Political Research." *Annual Review of Political Science* 7: 483–505.

Kadiyala, K. Rao, and Sune Karlsson. 1997. "Numerical Methods For Estimation and Inference in Bayesian VAR-Model." *Journal of Applied Econometrics* 12:99–132.

Kilian, Lutz. 1998. "Small-Sample Confidence Intervals for Impulse Response Functions." *Review of Economics and Statistics* 80:186–201.

King, Gary, and Langche Zeng. 2004. "When Can History Be Our Guide? The Pitfalls of Counterfactual Inference." Unpublished manuscript, Harvard University.

Leeper, Eric M., Christopher A. Sims, and Tao Zha. 1996. "What Does Monetary Policy Do?" *Brookings Papers on Economic Activity* 1996(2):1–63.

Litterman, Robert B. 1986. "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience." *Journal of Business, Economics and Statistics* 4:25–38.

Lutkepohl, H. 1990. "Asymptotic Distributions of Impulse Repsonse Functions and Forecast Error Variance Decompositions in Vector Autoregressive Models." *Review of Economics and Statistics* 72:53–78.

Martin, Andrew, and Kevin Quinn. 2002. "Dynamic Ideal Point Estimation via Markov Chain Monte Carlo for the U.S. Supreme Court." *Political Analysis* 10(2):134–153.

McGinnis, Michael, and John T. Williams. 1989. "Change and Stability in Superpower Rivalry." *American Political Science Review* 83(4):1101–1123.

Mittnik, S., and P. A. Zadrozny. 1993. "Asymptotic Distributions of Impulse Reponses, Step Responses, and Variance Decompoistions of Estimated Linear Dynamic Models." *Econometrica* 20:832–854.

Morris, Benny. 2001. *Righteous Victims: A History of the Zionist-Arab Conflict 1881–2001*. New York: Vintage Books.

Ni, Shawn, and Dongchu Sun. 2003. "Noninformative Priors and Frequentist Risks of Bayesian Estimators in Vector Autoregressive Models." *Journal of Econometrics* 115:159–197.

Ostrom, Charles, and Renee Smith. 1993. "Error Correction, Attitude Persistence, and Executive Rewards and Punishments: A Behavioral Theory of Presidential Approval." *Political Analysis* 3:127–184.

Robertson, John C., and Ellis W. Tallman. 1999. "Vector Autoregressions: Forecasting and Reality." *Economic Review (Atlanta Federal Reserve Bank)* 84(1):4–18.

Runkle, David E. 1987. "Vector Autoregressions and Reality." *Journal of Business and Economic Statistics* 5:437–442.

Schrodt, Philip A., Deborah J. Gerner, Rajaa Abu-Jabr, Oemeur Yilmaz, and Erin M. Simpson. 2001. "Analyzing the Dynamics of International Mediation Processes in the Middle East and Balkans." Presented at the Annual Meeting of the American Political Science Association, San Francisco.

Sims, Christopher A. 1980. "Macroeconomics and Reality." *Econometrica* 48(1):1–48.

Sims, Christopher A. 1987a. "Comment [on Runkle]." *Journal of Business and Economic Statistics* 5(4):443–449.

Sims, Christopher A. 1987b. A Rational Expectations Framework for Short-Run Policy Analysis. In *New Approaches to Monetary Economics*, eds. William Barnett and Kenneth Singleton. New York: Cambridge University Press, pp. 293–310.

Sims, Christopher A., and Tao A. Zha. 1995. "Error Bands for Impulse Responses." (Available from http://sims.princeton.edu/yftp/ier/.)

Sims, Christopher A., and Tao A. Zha. 1998. "Bayesian Methods for Dynamic Multivariate Models." *International Economic Review* 39(4):949–968.

Sims, Christopher A., and Tao A. Zha. 1999. "Error Bands for Impulse Responses." *Econometrica* 67(5):1113–1156.

Sims, Christopher A., and Tao A. Zha. 2004. "Were There Regime Switches in U.S. Monetary Policy?" (Available from http://www.princeton.edu/sims.)

Theil, Henri. 1963. "On the Use of Incomplete Prior Information in Regression Analysis." *Journal of the American Statistical Association* 58(302):401–414.

Waggoner, Daniel F., and Tao Zha. 1999. "Conditional Forecasts in Dynamic Multivariate Models." *Review of Economics and Statistics* 81(4):639–651.

Waggoner, Daniel F., and Tao Zha. 2000. "A Gibbs Simulator for Restricted VAR Models." Working Paper 2000-3, Federal Reserve Bank of Atlanta.

West, Mike, and Jeff Harrison. 1997. *Bayesian Forecasting and Dynamic Models*, 2nd ed. New York: Springer-Verlag.

Western, Bruce, and Meredith Kleykamp. 2004. "A Bayesian Change Point Analysis for Historical Time Series Analysis." *Political Analysis* 12(4):354–374.

Williams, John T. 1990. "The Political Manipulation of Macroeconomic Policy." *American Political Science Review* 84(3):767–795.

Williams, John T. 1993. "Dynamic Change, Specification Uncertainty, and Bayesian Vector Autoregression Analysis." *Political Analysis* 4:97–125.

Williams, John T., and Brian K. Collins. 1997. "The Political Economy of Corporate Taxation." *American Journal of Political Science* 41(1):208–244.

Zellner, Arnold. 1971. *An Introduction to Bayesian Inference in Econometrics.* New York: Wiley Interscience.

Zha, Tao A. 1998. "A Dynamic Multivariate Model for the Use of Formulating Policy." *Economic Review (Federal Reserve Bank of Atlanta)* First Quarter:16–29.