

IMPLEMENTING BAYESIAN VECTOR AUTOREGRESSIONS*

RICHARD M. TODD

Research Department
Federal Reserve Bank of Minneapolis

Abstract:

This paper discusses how the Bayesian approach can be used to construct a type of multivariate forecasting model known as a Bayesian vector autoregression (BVAR). In doing so, we mainly explain Doan, Litterman, and Sims (1984) propositions on how to estimate a BVAR based on a certain family of prior probability distributions, indexed by a fairly small set of hyperparameters. There is also a discussion on how to specify a BVAR and set up a BVAR database. A 4-variable model is used to illustrate the BVAR approach.

Introduction

Economic forecasting models are based on a combination of human insight and statistical analysis, but the question of how best to combine these elements remains, in practice, unresolved. Their combination cannot be avoided. Given the limitations of human brains and economic databases, the potential linkages among economic variables are too numerous to be accurately gauged by either human reasoning or statistical analysis alone. The traditional solution to this degrees of freedom problem has been more or less, to use reasoning to select a small set of coefficients thought to represent the most important linkages among variables, to assume all other linkages can be ignored (by fixing their coefficients at zero), and then to estimate the chosen coefficients from the data. Bayesian statistical theory suggests a different approach: include as many linkages

* Presented to International Association of Business Forecasters, Annual Conference on Business Forecasting, Pittsburgh, Pennsylvania, December 4, 1987.

The views expressed here are those of the author and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

—and thus coefficients— as your computer can handle; then use human insight to overcome the degrees of freedom problem by specifying a set of prior beliefs that supplement the data. Prior beliefs and data then jointly determine the estimated values of all the coefficients.

This paper discusses how the Bayesian approach can be used to construct a type of multivariate forecasting model known as a Bayesian vector autoregression (BVAR).¹ Since the idea of specifying prior beliefs about the numerous coefficients of a multivariate forecasting model is rather daunting, the key to the BVAR approach is to simplify this task. To that end, Doan, Litterman, and Sims (1984) have proposed that a certain family of prior probability distributions, indexed by a fairly small set of "hyperparameters", can adequately represent modelers' prior beliefs in most cases. They have shown how to estimate a BVAR, either when the modeler picks a particular member of their family of prior distributions or in the not uncommon case where the modeler has no strong beliefs (i.e., flat priors) about which members of this family of possible prior distributions are most appropriate. The explanation of their suggestions, in Section 2, forms the core of this paper. However, the BVAR approach to economic forecasting has implications, as well as a set of conventional procedures, for how the data in a model are selected and transformed. For that reason, and for completeness, Section 1 discusses how to specify a BVAR and set up a BVAR database. A 4-variable model is used to illustrate the BVAR approach.

Section 1: Specifying a BVAR

In the simplest form of a BVAR, n lags of each variable in the model appear in each equation. In that case, model specification consists of picking n and a data series for each variable. Even more complicated BVARs generally consist of blocks of equations that have the simple form or slight variations of it, so the links between model specification and database preparation remain quite direct. For that reason, I will discuss the basics of specifying a BVAR through a step-by-step discussion of setting up a BVAR database. In many ways the considerations and conventions used by BVAR modelers are the same as those used by other forecasters, except that somewhat different criteria are used in deciding which variables to include in the model.

What is the Purpose of the Model?

The first step in specifying any forecasting model including a BVAR, is to state as precisely as possible the goal of the forecasting exercise. At a minimum, the forecaster presumably wishes to produce accurate forecasts of at least one variable, and often more. I will refer to these variables as the variables of interest. They should be identified at the outset, and a loss function over their forecast errors could also be specified.

Frequently the forecaster wants not only to forecast accurately but also to claim that his or her forecasts incorporate the effects of one variable on another. (Note that this desire may be independent of, or even conflict with, the desire to forecast a set of variables accurately.) After October 19, 1987, for example, forecasters had to indicate whether their forecasts reflected that day's crash in the stock market. Incorporating such cross variable linkages rules out one fairly effective class of pure forecasting models—those which forecast each variable of interest by means of a univariate ARIMA model. Instead the modeler will need a multivariate model to estimate the linkages among the variables of interest and/or between those variables and a set of other variables which

affect the variables of interest but are not of direct interest themselves. I will refer to these other variables as related variables.

The forecaster may be even more ambitious and wish not only to incorporate but also to separately identify the causal linkages among variables. For example, forecasters may wish to say not only that their forecasts reflect the stock market crash but also that they know by just how much the crash affected each variable's forecast. In this case the estimated multivariate model will have to be augmented by a set of identification restrictions on the model's coefficients and/or covariance matrices. These additional restrictions provide a unique relationship between the forecasting model and a model of the underlying structure of the economy. Meaningful what-if questions can then be posed in the structural form of the model and traced through to the corresponding forecasting model, producing what are known as conditional forecasts. To estimate the effects of the stock market crash, for instance, the forecaster could use the structural model to pose the question of what the current outlook would be if the stock market had held steady at its October 18th level. Without identifying restrictions, what-if questions can't be posed, and the forecasts are said to be unconditional. Modelers can say that their unconditional forecasts include the effects of the stock market crash², but they can't separately identify what those effects are.

For the purposes of this paper, I will assume that my goal is to accurately forecast, in as much detail as possible, economic activity in the state of Minnesota (USA). I will also assume that my clients, whom I take to be the general public, demand multivariate forecasts (whether or not these are more accurate than multiple univariate forecasts) that incorporate the linkages among Minnesota variables and between Minnesota and national variables. However, my clients are content with unconditional forecasts, allowing me to defer the complexities and controversies of how to impose identifying restrictions on BVARs.

How Many Variables and Linkages Can I Include in the Model?

Ideally, my clients would like forecasts of almost every Minnesota economic variable. Furthermore, they want each variable's forecast to incorporate the effects of all the other variables. In non-Bayesian modeling, this ideal cannot be met because of insufficient degrees of freedom. Because there are many data series, the number of potential linkages among them, and hence the potential number of model coefficients, is huge. But each data series is fairly short, so that the number of observations per coefficient in an equation is small, possibly less than one. To conserve degrees of freedom, some variables are left out of the model, and many potential linkages among the remaining variables are turned off by assuming that the coefficients that represent them are zero.

In BVAR modeling, concerns over degrees of freedom play a much less important role in deciding which variables and linkages among variables to include in the model. Instead modelers must worry more about the availability of computing power, support staff, and their own time and energy in deciding whether to enlarge their models.

BVAR models can easily demand more memory than many computers and software packages allow. The prior information a BVAR modeler supplies to the estimation program includes at least a variance-covariance (VCV) matrix for each equation. Since BVARs typically include several lags of many variables in each equation, the number of coefficients, k , in a given equation can be large by non-Bayesian standards. Since the VCV matrices, in turn, are $k \times k$, the memory required to store them may exceed the limits of many computer systems (microcomputers, for example), or of software oriented to non-Bayesian methods. For example, although the RATS software available on the

mainframe at the Federal Reserve Bank of Minneapolis is designed for BVARs and accommodates a fairly generous number of coefficients per equation—about 200—this limit is binding at times and has dictated a block recursive structure for some of the models in use there. On a personal computer, RAIS can handle about 50 coefficients per equation. For a quarterly model with 6 lags of each variable in each equation, this imposes a limit of about 9 variables.

BVARs can also put large demands on support staff and on the modeler's time and energy. Whenever a large model is specified, someone has to assemble and maintain the database. This takes a lot of time, or money, or both. In addition, estimating a BVAR requires some programming and generates a lot of output that needs to be analyzed. My impression is that, variable for variable, BVARs are probably much easier to estimate and maintain than many econometric alternatives. Nonetheless, they do require support, and in practice limits on these resources may also constrain the size of BVARs.

Degrees of freedom do play a role in choosing variables for a BVAR. Although they ease the degrees of freedom problem, the prior beliefs imposed on the coefficients of BVAR models express a degree of uncertainty about the value of the coefficient. The more data are available, the more this uncertainty can be reduced. In cases where the prior uncertainty is large and the data series are short, some variables may have to be dropped from a BVAR to conserve degrees of freedom. I have estimated 6-variable, 6-lag BVARs with as little as 10 years of monthly economic data, but where longer data series can be substituted I normally exclude business cycle variables with less than 15 or 20 years of available data.

For my Minnesota model, let me begin with a generous "wish list" of variables of interest: gross state product, nonfarm employment, nonfarm earned income, farm income, an index of farm prices, farm land values, nonearned income, the unemployment rate, retail sales, the Minneapolis-St. Paul CPI, and possibly some further disaggregations of these series. I'll tentatively assume that the model will have the simple BVAR form, with n lags of each variable in each equation (n to be determined later).

How Shall I Transform the Data?

The BVAR framework does not have strong implications for whether data series should be modeled as seasonally adjusted or not, logged or not, deflated or real, etc. Nonetheless, some conventions appear to have been followed in the majority of BVAR applications. In addition, the BVAR framework itself allows certain new ways of dealing with these issues, especially seasonality.

Seasonal adjustment. Most BVARs to date have been estimated with seasonally adjusted data. I suspect that this is because most BVAR applications so far have been for general-purpose macro or regional forecasting, where success consists of beating your competitors' forecasts of the seasonally adjusted versions of variables like GNP, CPI, and unemployment. To keep things simpler, I will assume that my clients also want me to forecast the seasonally adjusted versions of my Minnesota variables.

As BVARs spread to other applications—such as predicting government revenues or corporate cash flows—forecasts of the nonseasonally adjusted versions of variables will be more relevant. The question then arises: Is it more accurate to directly model the unadjusted series, or should I first adjust, then model, then unadjust? Some research now underway touches on that question, but I don't think there's an answer yet. BVARs do open up one interesting possibility, which is to model the unadjusted data directly and introduce seasonal patterns into the prior beliefs about the coefficients (Canova, 1987; Ballabriga, 1987).

Logging. The general practice in BVAR models has been to log most variables, except those in ratio form or those taking on negative values. Many economic data series—GNP and Minnesota nonfarm employment, for example—seem to grow nearly exponentially with proportional disturbances. Logging them is convenient because it allows their behavior to be better approximated by a linear model with constant-variance additive disturbances. Even series that show little evidence of exponential growth or proportional disturbances may be logged, if their effects on other variables are believed to be more nearly multiplicative than additive. In U.S. models the exchange rate is often logged partly for these reasons, and in my Minnesota model I would log the index of ag prices for this reason.

Some series are usually not logged, however. Series in ratio or proportional form—such as interest rates or unemployment rates—are usually not logged. Their ratio form already captures some multiplicative effects, and their U.S. time series may show little evidence of exponential growth³. For these reasons, I will not log the Minnesota unemployment data. Series that have negative values cannot, of course, be logged. For this reason, I will work with levels rather than logs of Minnesota net farm income.

Deflating. With regard to deflating nominal variables, BVAR modelers seem to follow the convention of convention. That is, if the users of the forecast are primarily interested in the deflated version of a variable, then that form is directly incorporated in the model (rather than being computed from the forecasts of the nominal version and a deflator). Thus GNP is generally modeled in real terms (despite the occasional interest in nominal GNP). If the nominal version is of more direct interest—as is often true of stock price indices of the money supply—it is included in the model.

Sometimes this rule of thumb is not workable. For example, I don't know whether the users of my Minnesota forecasts are more interested in real or nominal retail sales. Also, my previous experience with this series has indicated that modeling it in nominal form may lead to implicit forecasts of real growth that seem unreasonable. Therefore, I would deflate retail sales.

Detrending. Although many data series, even those that have been logged, seem to grow over time, BVAR modelers rarely detrend their data series. Trend effects are allowed to enter into the estimated coefficients of the many distributed lags in the model. Frequently this results in explosive roots in the estimated model.

Are Appropriate Data Series Available?

At this point, I know which variables I want in my model, and in which forms. Now I have to see if the necessary raw data series are available or can be created.

Some series are just not available at all. They must be dropped or replaced by proxies. In my case, gross state product figures for Minnesota are not available now. To some extent, they can be proxied by employment and income.

Some raw series are available, but not frequently enough. The available series on my wish list are reported at a variety of intervals—monthly, quarterly, and annually. I could put them on a common basis with an annual model, but my clients aren't satisfied with this option. An alternative is use interpolation procedures to estimate quarterly values for series, such as farm land values, that are available only annually. The model could then be standardized on a quarterly calendar. My clients approve, and I proceed on this basis⁴. To interpolate quarterly land values, I need to find a set of quarterly data series that provide information about the likely intrayear movements in land values. I can use some of the series I have already chosen, such as farm prices and incomes. I may also wish to

gather some additional series—such as farm mortgage interest rates—just for this purpose⁵.

Some raw series are too short or inconsistent. My Minnesota employment figures go back, in one form or another, to 1939, but income data being in 1958, retail sales data in 1964, and unemployment data in 1970. I will pick 1958 as my starting date. I throw away the earlier data on employment (except that I've already used it to seasonally adjust) and estimate regression equations for retail sales (using post-1964 data) and unemployment (using post-1970 data) that allow me to backcast them to 1958. In these regressions, as in my interpolations, I may use a combination of variables in the model and variables gathered solely for this purpose. Similar techniques can be used to splice a more recent to a less recent version of a data series that is not available in a statistically consistent form for the whole data period. Mendesh (1987) describes one not entirely successful application of this approach to M1 data in the United States.

Finally, some series are released with an unusually long delay. If this delay is not too much longer than for the other variables in the model, the data may still be useful. This is the case for the quarterly data on Minnesota earned income, which are released about four months after the end of the quarter they cover. Most of my other series are available by one month after a quarter, but the one-quarter lag of the income data does not prevent them from contributing to my forecasts of the other variables. This will probably not be true of the new data series on state gross product that may be released soon. As I understand it, these series will be annual and available about 13-15 months after the fact. Even if I could get a good historical gross state product series and interpolate it to a quarterly level, the delayed release of this series would probably make it useless in my forecasting model.

How Should Related Series Be Handled?

Recall that related series, though not of direct interest themselves, may belong in the model because they affect the forecasts of the variables of interest. The usual BVAR methodology requires that forecasting equations for these related series appear somewhere in the model, so that the model is self-contained in the sense that it can produce forecasts autonomously. However, the modeler has wide latitude in choosing exactly how to model the related series.

The essential feature of a related series is that it contributes to the forecasts of the variables of interest, either by making them more accurate or by allowing modelers to claim that the forecasts incorporate all information their clients believe is relevant. Note that to improve the forecasts of the variables of interest, the related series could be either causally related or merely correlated with the variables of interest. In my Minnesota model, for example, national output, income, and employment variables can be viewed as shifting the demand for Minnesota products, thereby causing changes in Minnesota's employment and income. The role of stock market indices, such as the S&P 500, is less clear. They seem to contribute significantly to making both U.S. and Minnesota growth forecasts more accurate, but economists disagree over whether this is because they cause or merely prefigure changes in the real sector. A sharper example was related to me by Thomas Cargill, who found that state gambling revenues in Nevada could be used to improve forecasts of GNP. Probably no one thinks that Nevada gambling revenues cause big changes in GNP, but it is easy to believe that they might be correlated with—and thus serve as a proxy for—some unmeasured dimension of consumer confidence.

For my Minnesota model, I will choose my related series from a list of important national economic variables. These include the national counterparts of all my Min-

nesota variables of interest, plus GNP and selected GNP components, the interest rate on 3 month T-Bills, the money supply, the exchange rate, the S&P 500 index of stock prices, and an index of commodity prices.

The usual BVAR estimation method requires that forecasting equations for the related series be included in the model. The method relies on computing realistic simulations of the out-of-sample forecasting performance of the model. That, for example, means estimating the model through the first quarter of 1964, forecasting the next eight quarters, computing the resulting eight forecast errors, and then repeating the process for each successive quarter. If the modeler tries to avoid modeling the related series by simply plugging in someone else's forecasts of these variables, it will probably be impossible to find out what that other forecaster forecasted, or would have forecasted, as of the first quarter of 1964. As a result, it will be impossible to compute realistic statistics on the model's out-of-sample forecasting performance. Statistics computed by using the actual rather than forecasted values of the related series—in effect, under the assumption of perfect foresight of the related series—will generally be too optimistic. This can lead to the estimation of a model which is suboptimal in computing current forecasts, when perfect foresight of the related series is impossible.

Some important BVAR applications also require that the related series be included in the model. For example, BVAR forecasters frequently use stochastic simulation to compute confidence bands around forecasts or probabilities of events like a recession. Such probability assessments cannot be accurately computed without a stochastic model for the related series.

Although the related series need to be included in the model, they do not have to be modeled in the BVAR style. The modeler is free to make liberal use of exclusion restrictions in the equations for the related variables, since by definition the clients don't care whether the forecasts of the related series incorporate feedback from other variables. At one extreme, then, the modeler may choose to model each related series as a univariate autoregression. At the other extreme, the modeler may simply enlarge the BVAR and treat the related series as though they were variables of interest. Each equation would contain n lags of the variables of interest and of the related variables. Even committed BVAR modelers may stop short of this extreme, however, especially if it generates more coefficients than can be estimated with the modeler's computer package. An intermediate possibility is to follow a block recursive pattern: one BVAR with equations for the related variables only, with no variables of interest on the right-hand side, and a second BVAR with equations for the variables of interest, with some or all of the related variables on the right-hand side. A variety of other forms is also possible, and the out-of-sample forecasting statistics used in estimating the model can guide the selection of a final model. I will use these statistics in deciding whether to estimate a single BVAR for the Minnesota and U.S. series or to adopt a block recursive structure by excluding Minnesota variables from the U.S. variable's equations.

Is the Model Manageable?

I have now drawn up a fairly long wish list of variables of interest and related series. It's time to check whether my staff, my computer, and I can handle so big a model.

My computing resources are probably adequate. They allow about 200 coefficients per equation. Since I am working with quarterly data, I can probably get by with 4 lags of each variable. (I recommend at least enough lags to cover half a year, and preferably a year's worth or more.) This means I can put about 50 variables in the model and still not have to exclude any variable from any equation. If I want to experiment with 8

lags—which might be a good idea—I will have to cut down to about 25 variables. Even that is not very restrictive.

In this case my human resources are binding, however. I have only a limited amount of time to devote to building the model. In addition, I am not clever enough to clearly and concisely present a large model as an illustration of the BVAR technique. As a result, I will prune my model way back. My only variable of interest will be Minnesota nonfarm employment. My related series will be the S&P 500 index, the interest rate on 3-month T-bills, and GNP. All will be modeled as quarterly averages, with data since 1958. Only GNP will be deflated. GNP and Minnesota employment will be seasonally adjusted, but the other two series are already virtually nonseasonal in their raw form. All except interest rates will be logged.

Section 2: Estimating a BVAR

As its name suggests, the Bayesian vector autoregression procedure is motivated by Bayesian statistical theory. However, BVAR methods do not rigorously follow Bayesian guidelines. They can instead be interpreted as approximations to exact Bayesian methods.

Without Bayesian-like restrictions, my proposed Minnesota model would probably be subject to degrees of freedom problems, leading to what is known as overfitting of the coefficients. With four lags of each of my four variables in each of my four equations, plus a constant term, I would have an unrestricted VAR (UVAR) system with 68 coefficients, or 17 per equation. Because the right-hand-side variables are the same for each equation, the theory of seemingly unrelated regression (SURE) implies that I can optimally estimate my entire UVAR by applying OLS to each equation separately. I have 119 quarters of data for each equation, or 7 for each of the 17 coefficients in the equation. Those 17 coefficients are probably more than enough to fit quite well the in-sample data on Minnesota employment. In fact, the fitted coefficients may fit the data too well, in the sense that meaningless accidental patterns in the historical time series may strongly influence the estimated coefficients (see Todd (1984) for a nontechnical discussion of this idea). This is called overfitting.

A comparison of simulated out-of-sample forecasts of this UVAR and a set of four 4-lag univariate autoregression suggests that overfitting is at least a mild problem in the UVAR (see Table 1). One step ahead, its forecasts for all variables are less accurate than those of the system of univariate equations. For the national variables, this remains true at longer horizons, and the IVAR GNP forecasts are clearly inferior. For Minnesota employment, however, the two models' forecasts are nearly equally accurate, and the UVAR does better at long horizons. This reinforces the notion that national variables are important for Minnesota forecasts.

In larger UVAR models, where the number of coefficients grows rapidly, overfitting is more of a problem. The number of coefficients may even exceed the number of observations.

The BVAR method is designed to offset overfitting by imposing Bayesian prior restrictions on the coefficients. In effect, the information that OLS extracts from the data is supplemented by the modeler's beliefs about the likely values of the coefficients. For practical reasons, these prior beliefs have so far been imposed by methods that only approximate true Bayesian procedures.

One reason that BVAR methods are only approximately Bayesian is that the coefficients of BVARs are generally estimated with equation-by-equation estimators rather than system estimators. Although optimal for BVARs, single-equation methods are sub-optimal for BVARs unless the prior variance-covariance matrices of the coefficients are identical for each equation (up to a scale factor). It is unlikely that a modeler's prior

TABLE 1a
FORECASTING PERFORMANCE OF A SYSTEM
OF 4-LAG UNIVARIATE EQUATIONS

Quarters Ahead	Mean Error	Mean Absolute Error	Root Mean Squared Error	Theil U	Numbers of Observations
Forecast Statistics for Stock Prices					
1	0.44129199E-02	0.42502138E-01	0.56529307E-01	0.91949525	94
2	0.90060410E-02	0.76055179E-01	0.97561945E-01	0.95152243	93
4	0.16141438E-01	0.12081456	0.15058548	0.96612738	91
8	0.24718777E-01	0.16682311	0.21227678	0.97025187	87
12	0.28204854E-01	0.18362790	0.23081953	0.92956019	83
Forecast Statistics for T-Bill Interest Rates					
1	-0.10744089E-01	0.636635907	0.99507047	0.97783979	94
2	-0.28423123E-01	1.1173874	1.5945973	1.0338853	93
4	-0.70139866E-01	1.5833610	2.0769228	1.0178494	91
8	-0.12470743	2.5288659	3.0916833	1.0372513	87
12	-0.14503393	2.9584100	3.6232216	1.0444660	83
Forecast Statistics for GNP					
1	0.17570410E-03	0.77680328E-02	0.98777363E-02	0.79686875	94
2	0.39077426E-03	0.12310705E-01	0.15663106E-01	0.72366887	93
4	0.80566419E-03	0.20828869E-01	0.26260504E-01	0.67236302	91
8	0.71383881E-03	0.34583572E-01	0.40548729E-01	0.58997202	87
12	-0.92070822E-03	0.40445501E-01	0.46587041E-01	0.49970576	83
Forecast Statistics for Minnesota Employment					
1	0.10791717E-02	0.53575996E-02	0.64582705E-02	0.62954228	94
2	0.24459644E-02	0.10848400E-01	0.13116183E-01	0.66134598	93
4	0.52632735E-02	0.21999314E-01	0.25701500E-01	0.67459082	91
8	0.10109570E-01	0.38996145E-01	0.44899233E-01	0.664370795	87
12	0.14222564E-01	0.45364438E-01	0.55285882E-01	0.57681050	83

Current value of the system likelihood is 592.06.

TABLE 1b
FORECASTING PERFORMANCE OF A
4-LAG UNRESTRICTED VAR

Quarters Ahead	Mean Error	Mean Absolute Error	Root Mean Squared Error	Theil U	Numbers of Observations
Forecast Statistics for Stock Prices					
1	0.64694356E-02	0.56751922E-01	0.69585327E-01	1.1318620	94
2	0.17736743E-01	0.10303019	0.12422263	1.2115443	93
4	0.31752133E-01	0.15047637	0.18640531	1.1939405	92
8	0.64125931E-01	0.20885106	0.27720111	1.2669414	87
12	0.65056825E-01	0.27150705	0.38216381	1.5390564	83
Forecast Statistics for T-Bill Interest Rates					
1	-0.70387902E-01	0.72140918	1.0716148	1.0530586	94
2	-0.13341794	1.2586961	1.7903711	1.1608187	93
4	-0.14276735	1.5966581	2.2257944	1.0908077	91
8	-0.85740554E-01	2.3709643	3.1562491	1.0589130	87
12	0.53176012E-01	2.5657213	3.3852398	0.97386300	83
Forecast Statistics for GNP					
1	0.81345505E-03	0.92385880E-02	0.12096472E-01	0.97586126	94
2	0.15835237E-02	0.14849571E-01	0.19575944E-01	0.90445033	93
4	0.41698089E-02	0.27198716E-01	0.35305520E-01	0.90394785	91
8	0.10053677E-01	0.51197739E-01	0.71922229E-01	1.0464472	87
12	0.99343588E-02	0.82881626E-01	0.13764736	1.4764445	83
Forecast Statistics for Minnesota Employment					
1	0.14234481E-03	0.52966308E-02	0.66608800E-02	0.64929233	94
2	0.41966464E-03	0.94588151E-02	0.11549874E-01	0.58236933	93
4	0.13690653E-02	0.18894005E-01	0.23195708E-01	0.60882095	91
8	0.34885120E-02	0.41518127E-01	0.55188549E-01	0.79122304	87
12	0.12864247E-02	0.70509216E-01	0.11052038	1.1530849	83

Current value of the system likelihood is -655.93.

beliefs would satisfy this condition exactly, so the typical BVAR single-equation estimation procedure is inefficient.

Full system estimation of a BVAR, though efficient, is computationally demanding. In my Minnesota model, it would involve the 68×68 variance-covariance matrix of all the coefficients, instead of the four 17×17 variance-covariance matrices of the separate equations. Some efforts are being made to find efficient algorithms for system estimation of BVARs, but I am not aware of any successful applications yet⁶.

For each BVAR equation, the prior information takes the form of a "known" variance of the disturbance term and a normal distribution for the coefficients.

Although treated as known, the disturbance term variances in the prior are usually computed from the data. [Alternatively, Doan, Litterman, and Sims (1984) propose treating this variance as an additional hyperparameter.] For each BVAR variable, Doan, Litterman, and Sims (1984) suggest setting the standard deviation of its disturbance term to 0.9 times the standard error of the residuals in a regression of the variable on six lags of itself. The factor 0.9 allows for a 10-percent reduction in the standard deviation when other variables are added to the equation. All of these standard deviations are also used in scaling the prior covariance matrix of the coefficients of each equation. Therefore a standard deviation must be computed even for related series that are not modeled with BVAR equations.

The mean of the normal distribution of each equation's coefficients is typically set according to the so-called random walk (or random walk with drift) prior. The intuition behind this prior is that most economic series are reasonably well approximated by a random walk around a trend.

To help capture the trend, a constant is included in each equation. For logged variables, the constant can pick up the trend growth rate. (As it happens, the variables that typically have not been logged also typically exhibit little trend.) Usually, the prior means of all the constant terms are zero. Their prior variances are usually scaled by a hyperparameter dedicated to this purpose. Alternatively, their variances can be set at a large multiple (100,000 for example) of the disturbance term variance. This expresses prior ignorance about the constant term and lets the data determine its value.

To represent a discrete-time random walk, the prior mean of the coefficient on the first lag of the dependent variable is set to 1.0. All other coefficients are given a prior mean of 0.0.

It may be more plausible that a variable behaves like a random walk in continuous, not discrete, time. In this case the implications of the random walk prior for the BVAR coefficients are somewhat different. BVARs are generally fit to time-averaged discrete data series. If the underlying continuous time version of the variable follows a random walk, the time-averaged version will show a more complicated pattern of lag coefficients. [See Christiano and Eichenbaum (1987), pp. 75-77, and Working (1960).] The coefficient of the j th lag of the dependent variable will have a nonzero mean given by $(1-\alpha)^j - 1$, where $\alpha = \sqrt{3}/2$. The prior means of the other coefficients in the equation remain zero, however.

Specifying the prior variance-covariance matrices for the coefficients of each equation is the heart of the BVAR method and the core of a typical BVAR computer program. Usually the prior VCVs are chosen from a standard family of VCVs. This family, which is indexed by about ten so-called hyperparameters, has gradually emerged from the work of Sims, Litterman, and Doan. Each setting of the hyperparameters identifies a particular VCV in this family and thus completes the specification of the prior normal distribution of the coefficients. Viewed another way, the prior VCV can be written as a

function of the hyperparameters (and of the disturbance term variances). Doan, Litterman, and Sims (DLS) have thus considerably simplified the task of specifying prior VCVs, reducing the number of decisions from hundreds or thousands to about ten.

The DLS family of VCVs is based on two simple beliefs about the coefficients of a multivariate time series forecasting model. One is the belief that, in forecasting a given variable, its own past values are likely to be more useful than the past values of other variables. The other is that, for both the forecasted variable and the other variables used to forecast it, more recent values are likely to be more useful than more distantly lagged values. In other words, our confidence that a coefficient is nearly zero increases if the coefficient applies to a more distant lag or to a variable other than the dependent variable, as indicated in Figure 1.

These two simple beliefs are expressed by means of two or three hyperparameters. One, which I will call DAMP, controls the rate at which the prior variances shrink toward zero as lag length increases. This is generally done by making the prior variances of coefficients on j th lags proportional to either j -DAMP or DAMP^j . Although the most prominent BVAR software package (RATS, by VAR Econometrics) has a built-in facility for varying DAMP (which it refers to as the lag decay parameter), in many applications the formula $j-1$ is used.

Two other basic hyperparameters govern the relative prior variances of coefficients of own lags (lags of the dependent variable) versus coefficients of cross lags (lags of other variables). Own-lag variances are all scaled by the hyperparameter DLS call π_1 , and cross-lag variances are all scaled by the hyperparameter they call π_2 . π_1 and π_2 can strongly affect the forecasting performance of a BVAR.

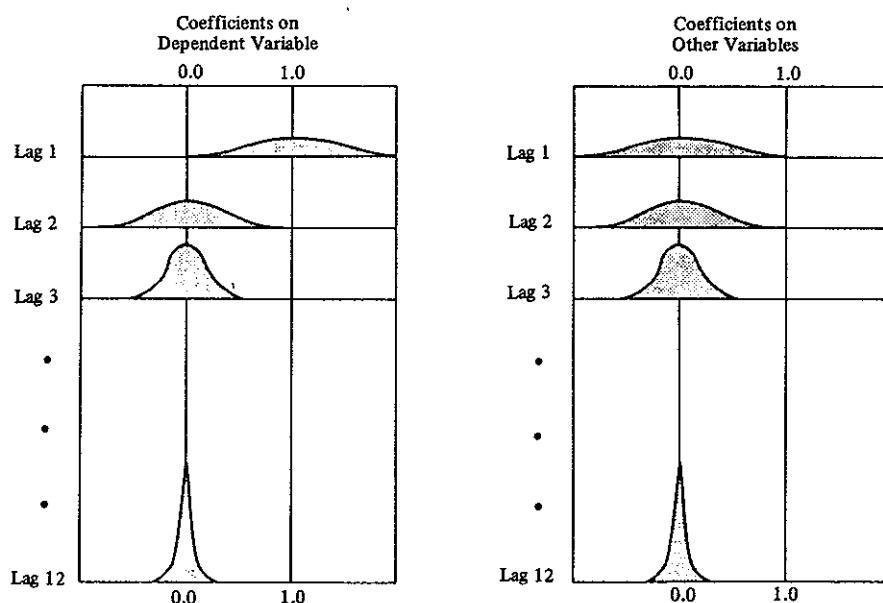
Over the years, the DLS family of VCVs has been refined by the addition of more hyperparameters. The prior variances of the constant terms, for example, are scaled by a hyperparameter they call π_3 . Other hyperparameters influence the time variation of the coefficients (π_7, π_8), the sum of the coefficients (π_6), and relative variances among the cross variables (π_4).

Although DLS have taken us a long way toward specifying a prior VCV for our coefficients, it is usually the case that even experienced BVAR modelers cannot pick a setting of the hyperparameter that adequately represents their beliefs. In fact, most of the easily agreed to characteristics of a prior VCV are common to all members of the DLS family, and DLS focus attention on the hyperparameters partly because they index the remaining uncertainty about the nature of the prior. In Bayesian terms, the prior distribution of the coefficients is a mixture of normal distributions, and the hyperparameters index the family of normal distributions in the mixture. (More correctly, the distribution of the coefficients is a mixture of conditional normal distributions, each conditional on the assumed variance of the equation disturbance term.) A Bayesian could proceed by expressing a fairly flat prior distribution over the hyperparameters, computing the implied mixture of coefficient distributions, and using that mixture to specify the prior VCV of the coefficients.

This fully Bayesian approach has never, as far as I know, been implemented. Instead DLS have suggested a procedure for using the historical data to select a particular set of hyperparameters (and thus also a particular VCV for each equation). This procedure appears to be contradictory. A prior is the distribution the modeler believes before examining the data, so how can the data be used to pick the prior? Strictly speaking, they cannot. However, under certain conditions, this data-based method for selecting a prior is justified as an approximation to the Bayesian mixture-of-distributions procedure outlined above (see the Appendix).

FIGURE 1

A SCHEMATIC REPRESENTATION OF THE MEANS AND VARIANCES OF THE COEFFICIENTS IN THE DISCRETE-TIME RANDOM WALK PRIOR DISTRIBUTION



The key ingredient in the DLS procedure for selecting a set of hyperparameters is realistic simulation of the model's out-of-sample forecasting performance. A set of hyperparameters is chosen. Then the model is estimated, using Kalman filter techniques to incorporate the prior associated with those hyperparameters, through a startup date. A forecast is made from that date and compared to the actual data to compute its forecast errors. The Kalman filter is applied again to update the estimates of the coefficients with data through date startup plus one, and the forecasts and forecast errors for that date are computed. This continues until the end of the historical data is reached. The whole procedure is then repeated for many other settings of the hyperparameters, so that each hyperparameter setting has a simulated forecasting track record. The setting with the best track record wins. That is, its prior is singled out and used to estimate the final forecasting model.

A critical aspect of this procedure, obviously, is deciding which forecasting track record is best. Several criteria, or metrics, have been proposed and used, including minimization of the log determinant of the VCV of the simulated one-step-ahead forecast errors and minimization of weighted averages of root-mean-squared forecast errors one or more steps ahead. The choice of a criterion is to some extent problem-specific and could be based on the modeler's forecast error loss function. In my Minnesota model, for example, I could focus on my variable of interest, using as my criterion the root-mean-squared error of one-step-ahead forecasts of Minnesota nonfarm employment.

However, the criterion that is most consistent with the notion that BVAR methods are approximately Bayesian is maximization of the quasiliikelihood statistic proposed by Dean, Litterman, and Sims (1984). For any given equation, maximizing this statistic is equivalent (under certain assumptions) to minimizing a weighted sum of the out-of-sample one-step-ahead forecast errors, with weights that are inversely proportional to the conditional forecast error variances. An equation's forecasting performance thus depends both on its model's accuracy and on its model's ability to detect when forecasts are subject to abnormal uncertainty. An analogous statistic could be computed for the model as a whole if efficient system Kalman filtering algorithms were available. Since these algorithms are not widely available yet, a model's forecasting performance (that is, its performance under a given set of hyperparameters) is usually ranked by summing the quasiliikelihoods of its individual equations. This is the criterion I will use to select hyperparameters for my Minnesota model.

The remaining task—searching for a set of hyperparameters that optimize the criterion—is computationally intensive. Even if the final model selected will be small enough to estimate and use on a microcomputer, it is helpful to use a mainframe at this stage. This is especially true if the number of hyperparameters is large, say 5 or more.

The computational burden of hyperparameter search depends on two factors, the number of hyperparameters to be picked and the computational complexity of computing out-of-sample forecasts for a given set of hyperparameters. The latter factor depends primarily on the number of coefficients per equation and the length of the historical period over which out-of-sample forecasts will be computed. The former factor determines the dimension of the space to be searched. This is important because in can rule out a nearly surefire method of optimizing the criterion—checking all hyperparameter settings on a fine grid that encompasses all reasonable values. For example, five hyperparameters and 5 grid points for each implies a grid with 5^5 , or 3125, settings of the hyperparameters. Except for small models with short out-of-sample forecasting periods, checking all of those hyperparameter settings would require hours—possibly days—of mainframe CPU time. Similarly, computational burdens also limit the possibilities of computing the numerical derivatives needed for hill-climbing algorithms.

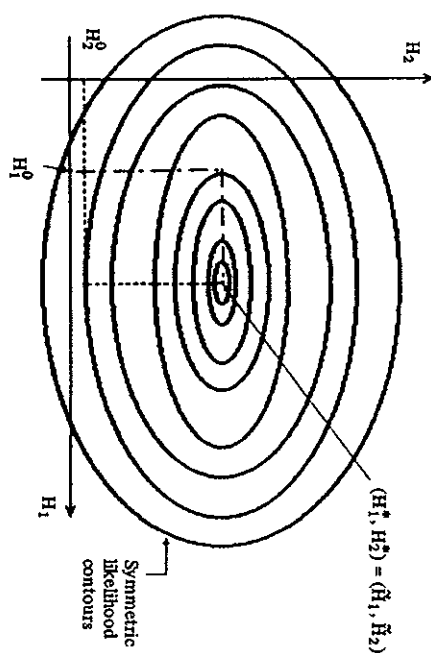
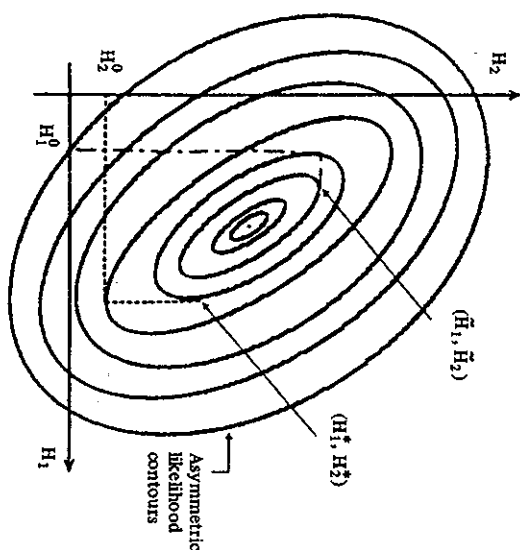
With grid search and hill climbing effectively ruled out, the most common alternative is a procedure sometimes referred to as axial search. The modeler takes an initial guess at the optimal setting of the hyperparameters. Then a grid of settings of the first hyperparameter is searched, holding the remaining hyperparameters fixed. The best point in this grid used to pick the optimal value of the first hyperparameter. The procedure is repeated for each hyperparameter in succession, always with the other hyperparameters held constant. A grid of 5 points for each of 5 hyperparameters thus implies 25 points under axial search, rather than 3125 under a multidimensional grid search. And if the criterion is a fairly smooth, symmetrical (when bisected by lines parallel to the axes) function of the hyperparameters, axial search will get close to the optimal setting of the hyperparameters. (See Figure 2a.)

In general, however, axial search can go astray, as illustrated in Figure 2b. As the figure suggests, it may be valuable to do several axial searches, varying the order in which the hyperparameters are searched. Table 2 shows the differences between 2 axial searches over 10 hyperparameters for the 4-lag version of my Minnesota model. The first search, leading to Model 1, ends with fairly typical optimal hyperparameter values (although the hyperparameter on the variance of the constant term is somewhat low). The search leading to Model 3, over a different ordering of the hyperparameters, leads to a lower likelihood and somewhat unusual degrees of time variation, decay of coefficients, and tightness on sums of coefficients.⁷

Tables 3a–3c show three BVAR models that achieve better performance statistics for Minnesota employment than the univariate and unrestricted VAR systems at least one year ahead. Results for the national variables are less clear. One step ahead, the BVAR models generally forecasted more accurately than the univariate and unrestricted VAR models. At longer horizons, however, BVAR Model 1 (4 lags, first axial search order) is dominated by the univariate model and, for TBILL, by the unrestricted VAR. Model 2 (8 lags), which had the highest likelihood, fared better. It almost uniformly forecasted better than the univariate and unrestricted BVAR models. (I also estimated 8-lag univariate and unrestricted VAR models. The former performed about the same as its 4-lag counterpart. The latter performed much worse than any model shown here.) Based on its higher likelihood and overall forecasting performance, Model 2 is my choice as the best forecasting model. However, the forecasting performance of Model 3 (4 lags, second axial search order) was generally nearly as good and was better for Minnesota employment.

The optimal hyperparameter settings from the axial search reported in Table 2 were used to estimate the Minnesota BVAR models 1 and 2. I am now ready to use them, especially model 2, for forecasting my variable of interest. Table 4 shows the results, under two assumptions. With data only through the third quarter of 1987, both models forecast strong GNP and Minnesota employment growth ahead. When preliminary estimates of fourth quarter average stock prices and interest rates are added to the data (to reflect October's stock market crash), the models forecast an immediate but brief downturn in GNP and a slowdown in Minnesota employment growth.

My model is up and running, and my discussion of implementing Bayesian vector autoregressions would now have to turn to applications. Perhaps another day.

FIGURE 2A
SUCCESSFUL AXIAL SEARCHFIGURE 2B
UNSUCCESSFUL AXIAL SEARCH

Explanation: Each graph shows two searches. Each search begins at the initial guess (H_1^0, H_2^0) . The search denoted by and * first optimizes H_1 while fixing H_2 at H_2^0 . It then optimizes H_2 while fixing H_1 and H_1^* . The search denoted by - - - - - or ~ first optimizes H_2 (with $H_1 = H_1^0$) and then optimizes H_1 (with $H_2 = H_2^*$).

TABLE 2

OPTIMAL HYPERPARAMETERS AND LIKELIHOOD FOR
3 SPECIFICATIONS AND SEARCHES

(DLS notation)	OWN (π_1)	CROSS (π_2)	TVAR (π_7)	SUM (π_6)	TITE (π_5)	WT (π_4)	CONST (π_3)	DECAY (π_8)	BEGWT (*)	MNWT (**)	Likelihood
Orderings											
First order of search	1	2	9	8	6	10	5	7	3	4	
Second order of search	9	8	3	1	4	2	7	6	10	5	
Results											
First order 4 lags (Model 1)	0.01	0.01	0.000001	10.0	5.0	1.0	0.001	1.0	0.2	0.25	769.38
First order 8 lags (Model 2)	0.01	0.01	0.0001	10,000.0	5.0	0.01	0.001	0.999	1.5	10.0	789.55
Second order 4 lags (Model 3)	0.05	0.01	0.0001	10,000.0	5.0	0.01	100.0	0.999	0.2	10.0	760.73

(*) Extends no-change forecast in prior to coefficient covariances.

(**) Larger values increasingly constrain Minnesota coefficients in U.S. equations to be near zero. A value of 10 implies very little feedback from Minnesota to U.S.

TABLE 3a

FORECASTING PERFORMANCE OF
BVAR MODEL 1

Quarters Ahead	Mean Error	Mean Absolute Error	Root Mean Squared Error	Theil U	Numbers of Observations
-------------------	---------------	---------------------------	-------------------------------	---------	-------------------------------

Forecast Statistics for Stock Prices

1	0.96457942E-02	0.41427879E-01	0.55086468E-01	0.89602629	94
2	0.22915027E-01	0.73341238E-01	0.94585965E-01	0.92249767	93
4	0.49051095E-01	0.11945865	0.14578003	0.93522655	91
8	0.88026369E-01	0.17331010	0.21617859	0.98803932	87
12	0.11542238	0.19562545	0.25357343	1.0211951	83

Forecast Statistics for T-Bill Interest Rates

1	-0.66237272E-01	0.64885951	1.0024518	0.98509333	94
2	-0.13288265	1.1374443	1.6338256	1.0593197	93
4	-0.18117597	1.5773181	2.0916440	1.0250639	91
8	-0.15550373	2.5676163	3.3218537	1.1144729	87
12	-0.76784764E-01	3.1334149	4.0116873	1.1564490	83

Forecast Statistics for GNP

1	-0.64374095E-03	0.79779010E-02	0.98313782E-02	0.79312890	94
2	-0.12042653E-02	0.12271489E-01	0.15022276E-01	0.69406116	93
4	-0.97842509E-03	0.22186438E-01	0.26025275E-01	0.66634032	91
8	0.22722351E-02	0.39164592E-01	0.47713133E-01	0.66421198	87
12	0.67937170E-02	0.60463561E-01	0.73629265E-01	0.78976829	83

Forecast Statistics for Minnesota Employment

1	-0.49030206E-03	0.38385829E-02	0.47857412E-02	0.46650669	94
2	-0.11091415E-02	0.74289833E-02	0.91212427E-02	0.45991254	93
4	-0.17157645E-02	0.15542168E-01	0.18689575E-01	0.49054787	91
8	-0.42042340E-03	0.35660881E-01	0.40454940E-01	0.57999134	87
12	0.37508124E-02	0.51762176E-01	0.60092712E-01	0.62696127	83

TABLE 3b

FORECASTING PERFORMANCES OF
BVAR MODEL 2

Quarters Ahead	Mean Error	Mean Absolute Error	Root Mean Squared Error	Theil U	Numbers of Observations
-------------------	---------------	---------------------------	-------------------------------	---------	-------------------------------

Forecast Statistics for Stock Prices

1	0.22856762E-02	0.42061257E-01	0.55767352E-01	0.90710142	94
2	0.31050548E-02	0.75705907E-01	0.95092680E-01	0.92743967	93
4	-0.45228548E-02	0.11010722	0.14087597	0.90383305	91
8	-0.36636097E-01	0.15387294	0.19664602	0.89876613	87
12	-0.75970975E-01	0.17690725	0.22369056	0.90085028	83

Forecast Statistics for T-Bill Interest Rates

1	-0.15492653E-02	0.63847701	0.96049839	0.94386636	94
2	0.17478035E-01	1.1022750	1.5295039	0.99168087	93
4	0.12194619	1.4434553	1.9036699	0.93294236	91
8	0.40882801	2.2568759	2.9055395	0.97480058	87
12	0.72548312	2.7045816	3.3794420	0.97419167	83

Forecast Statistics for GNP

1	0.1196314E-03	0.75846517E-02	0.96080025E-02	0.77510846	94
2	0.22709227E-03	0.11573013E-01	0.14943110E-01	0.69040351	93
4	0.50809497E-03	0.20209497E-01	0.25525459E-01	0.65354325	91
8	-0.12175093E-02	0.33500818E-01	0.39440695E-01	0.57383045	87
12	-0.53656837E-02	0.39503988E-01	0.45586761E-01	0.48897648	83

Forecast Statistics for Minnesota Employment

1	0.17109189E-02	0.43507330E-02	0.53161864E-02	0.51821378	94
2	0.39960907E-02	0.87830836E-01	0.10600508E-01	0.53450025	93
4	0.10024067E-01	0.19389729E-01	0.22983272E-01	0.60324510	91
8	0.23136702E-01	0.41841318E-01	0.48283716E-01	0.66222035	87
12	0.35664753E-01	0.55455294E-01	0.64764457E-01	0.67570268	83

TABLE 3c

FORECASTING PERFORMANCE OF
BVAR MODEL 3

Quarters Ahead	Mean Error	Mean Absolute Error	Root Mean Squared Error	Theil U	Numbers of Observations
Forecast Statistics for Stock Prices					
1	0.35541865E-02	0.43619641E-01	0.55569037E-01	0.90387568	94
2	0.70145339E-02	0.78414110E-01	0.95721289E-01	0.93357049	93
4	0.95860105E-02	0.11990585	0.14613505	0.93757428	91
8	0.11074612E-01	0.16263327	0.20623418	0.94258861	87
12	0.36510444E-02	0.17841622	0.22663072	0.91269092	83
Forecast Statistics for T-Bill Interest Rates					
1	-0.41453112E-01	0.64017480	0.99358674	0.97638176	94
2	-0.81996515E-01	1.1394534	1.5930433	1.0328778	93
4	-0.11252442	1.5891316	2.0644306	1.0117273	91
8	-0.55075583E-01	2.3617689	2.9809486	1.0001001	87
12	0.93502966E-01	2.5722453	3.2281564	0.93058057	83
Forecast Statistics for GNP					
1	-0.60704319E-01	0.77246191E-02	0.95961864E-02	0.77415521	94
2	-0.13451009E-02	0.1133701E-01	0.14273644E-01	0.65947277	93
4	-0.2516580E-02	0.18371969E-01	0.23179269E-01	0.59347237	91
8	-0.52181970E-02	0.29454862E-01	0.37036705E-01	0.53887311	87
12	-0.10336068E-01	0.39937586E-01	0.50844328E-01	0.54537063	83
Forecast Statistics for Minnesota Employment					
1	0.54594528E-03	0.38864228E-02	0.48762372E-02	0.47532811	94
2	0.12777054E-02	0.76168731E-02	0.92663490E-02	0.46722911	93
4	0.32155998E-02	0.15305952E-01	0.18582637E-01	0.48774104	91
8	0.70654911E-02	0.32061078E-01	0.37841634E-01	0.54252509	87
12	0.10233763E-01	0.44356588E-01	0.53757773E-01	0.56086738	83

TABLE 4

FORECASTS OF GROWTH RATES FOR GNP AND MINNESOTA EMPLOYMENT

	GNP				MINNESOTA NONFARM EMPLOYMENT			
	With Data Through 87:3 only		With Additional Data on Stock Prices and T-Bill Rates of 87:4		With Data Through 87:3 only		With Additional Data on Stock Prices and T-Bill Rates for 87:4	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
87:4	4.7	4.5	-1.1	-2.0	3.6	3.5	4.5	4.0
88:1	4.4	4.2	2.4	0.9	3.6	3.5	2.1	1.7
88:2	3.6	2.9	2.5	0.4	3.7	3.1	1.4	0.7
88:3	3.0	1.7	3.7	1.1	3.5	2.2	1.7	-0.1
88:4	2.7	1.1	5.2	1.9	3.1	1.5	2.8	0.2
89:1	2.5	1.3	5.5	3.6	2.8	1.2	3.7	1.5
89:2	2.4	1.5	5.0	4.5	2.6	0.9	4.0	2.5
89:3	2.3	1.8	4.3	4.4	2.4	0.9	4.1	3.1
89:4	2.3	2.2	3.8	3.9	2.3	1.0	4.0	3.1

Notes

- 1 This paper does not focus on other topics related to BVARs, such as their conceptual motivation [see Doan, Litterman, and Sims (1984) or Todd (1984)], their forecasting performance [see Litterman (1986a) and McNees (1986)], or their use in resolving questions concerning economic structure or policy [see Sims (1982, 1986a, 1987), Litterman (1985), Cooley and LeRoy (1985), Runkle (1987), and the references cited therein].
- 2 This assumes that the unknown underlying structure of the economy will not change through the forecasting period (or, in the case of a time-varying coefficients model, will continue to evolve as it has in the past). This may not be a good assumption in the case of the 1987 stock market crash, which may very well induce new private sector strategies and public sector regulations. In addition, the mere size of the crash itself may be inconsistent with the stochastic structure of many models.
- 3 However, in countries where interest rates have taken on more extreme values than in the United States, the log transformation may be more useful and has the advantage of capturing the multiplicative relationship between interest rates and asset values.
- 4 There is a long literature on optimal interpolation of economic time series. One early BVAR modeler, Robert Litterman, made extensive use of interpolated national income and product account data to build a monthly model of the U.S. economy (Litterman, 1984). His interpolation method (Litterman, 1983) generalizes the earlier methods of Chow and Lin (1971) and Fernandez (1981).
- 5 Amiri-Zadeh (1985) describes many examples in which series both inside and outside the model are used to interpolate data from a quarterly to a monthly level.
- 6 Valentin Carril (1988) has experimented with system estimation procedures for a Chilean macro-economic BVAR. His procedure is expensive and has so far not led to improved forecasts.
- 7 Sims (1986b) has recently proposed an alternative search procedure. First the criterion is evaluated on a small grid (up to 50 points) of the hyperparameter space. Then Bayesian procedures and some assumptions about the rate of change of the criterion function are used to interpolate the value of the function for any other hyperparameter setting. These interpolations are not computationally expensive, so a hill climbing routine can be used to find optimal hyperparameter settings for the interpolated function.

Appendix:

A Bayesian Interpretation of Searching over Hyperparameters

The following is patterned after the discussion in Doan, Litterman, and Sims (1984). I am responsible for any errors.

Suppose the data y are conditionally distributed with density function $p(y|\theta)$, where θ is a vector of coefficients. If we can then specify a prior density $q(\theta)$ for these coefficients, we can form the joint density $p(y|\theta)q(\theta)$ of the data and coefficients, calculate the likelihood function by plugging observed values of y into p , and apply Bayes's Rule to get a posterior pdf for θ . This is the standard Bayesian approach.

Suppose now, however, that our prior for θ can be written as a function of a smaller set of parameters π , so that $q_1(\theta|\pi)$ represents our prior pdf for θ , conditional on the vector π . Suppose further that we are unsure of our beliefs about π , holding the prior pdf $q_2(\pi)$. Then our prior pdf for θ is actually a mixture of pdfs of the form $q_1(\theta|\pi)$, where the mixture is formed by weighting $q_1(\theta|\pi)$ by $q_2(\pi)$ and integrating, or

$$q(\theta) = \int q_1(\theta|\pi)q_2(\pi)d\pi.$$

We are free to formulate our prior for θ in this way and then proceed as above.

Alternatively, suppose we leave q_2 unspecified and regard y and θ as conditionally (on π) jointly distributed as $p(y|\theta)q_1(\theta|\pi)$. If we integrate θ out, we obtain $m(y|\pi)$, the

marginal distribution for y conditional on π . For a fixed data vector y , $m(y|\pi)$ plays the role of a likelihood function for π . And, if we know our prior distribution for π is relatively flat in the region where $m(y|\pi)$ is large, we can say, even without specifying anything more about our prior for π , that our posterior for π is roughly proportional to the "likelihood" $m(y|\pi)$. Our posterior for θ , in turn, is a weighted average of those obtained (by the usual Bayesian methods) conditional on values of π , with weights given by the posterior on π and hence also roughly proportional to $m(y|\pi)$.

This process of weighting the conditional posterior for θ by $m(y|\pi)$ has never, to my knowledge, been used by a BVAR modeler. Instead, the value π^* that maximizes $m(y|\pi)$ has been found (at least approximately), and the posterior value θ^* for θ conditional on π^* has been computed and used as the fitted coefficient vector. On the surface, this is a procedure for using the data to select a prior and is thus not a Bayesian procedure. However, if the likelihood for π , $m(y|\pi)$, is large only within some region P (and falls off rapidly elsewhere), and if the posterior for θ conditional on π is not very sensitive to variations of π within P , then the difference between θ^* and the true posterior value of θ formed by weighting conditional posteriors for θ by $m(y|\pi)$ is small. In this sense, picking π^* and computing θ^* can approximate the true Bayesian flat- π -prior estimate of θ . At least in some well-known examples of BVAR models [Doan, Litterman, and Sims (1984)], the conditions for approximate optimality of θ^* appear to be met. Estimating BVARs by picking optimal hyperparameter in this way to some extent implies a belief that these conditions are not too seriously violated.

References

- AMIRIZADEH, HOSSAIN (1985). "Technical Appendix for the Bayesian Vector Autoregressive Model of the U.S. Economy". Research Department, Federal Reserve Bank of Minneapolis.
- BALLABRIGA, FERNANDO (1987). "The Relative Importance of International Shocks for the Spanish Economy". University of Minnesota, manuscript.
- CANOVA, FABIO (1987). "Forecasting Seasonal Series with Bayesian Methods". University of Minnesota, manuscript (May).
- CARRIL, VALENTIN (1988). Oral communication, based on his research toward a Ph.D. in Economics at the University of Minnesota.
- CHOW, GREGORY C., and AN-LOH, LIN (1971). "Best Linear Unbiased Interpolation, Distribution and Extrapolation of Time Series by Related Series". *Review of Economics and Statistics* 53, 372-75.
- CHRISTIANO, LAWRENCE J., and EICHENBAUM, MARTIN (1987). "Temporal Aggregation and Structural Inference in Macroeconomics". *Carnegie-Rochester Conference Series on Public Policy* 26, 63-130.
- COOLEY, THOMAS F., and LEROY, STEPHEN F. (1985). "A theoretical Macroeconomics: A Critique". *Journal of Monetary Economics* 16, 283-308.
- DOAN, THOMAS A., and LITTERMAN, ROBERT B. (1986). *User's Manual, RATS Version 2.00*. VAR Econometrics.
- , and SIMS, CHRISTOPHER (1984). "Forecasting and Conditional Projection Using Realistic Prior Distributions". *Econometric Review* 3, 1-100.
- FERNANDEZ, ROQUE B. (1981). "A Methodological Note on the Estimation of Time Series". *Review of Economics and Statistics* 63, 471-76.
- LITTERMAN, ROBERT B. (1983). "A Random Walk, Markov Model for the Distributions of Time Series". *Journal of Business and Economic Statistics* 1, 169-73.
- (1984). "Forecasting and Policy Analysis with Bayesian Vector Autoregression Models". *Federal Reserve Bank of Minneapolis Quarterly Review*, Fall, 30-41.
- (1985). "How Monetary Policy in 1985 Affects the Outlook". *Federal Reserve Bank of Minneapolis Quarterly Review*, Fall, 2-13.
- (1986a). "Forecasting with Bayesian Vector Autoregressions—Five Years of Experience". *Journal of Business and Economic Statistics* 4, 25-38.

- (1986b). "A Statistical Approach to Economic Forecasting". *Journal of Business and Economic Statistics* 4, 1-4.
- (1986c). "Specifying Vector Autoregressions for Macroeconomic Forecasting". Chapter 6 in P. Goel and A. Zellner, ed. (1986), *Bayesian Inference and Decision Techniques*. Elsevier Science Publishers.
- MCNEES, STEPHEN K. (1986). "Forecasting Accuracy of Alternative Techniques: A Comparison of U.S. Macroeconomic Forecasts". *Journal of Business and Economic Statistics* 4, 5-15.
- MENDES, SUSAN (1987). "A Warning About M1". Federal Reserve Bank of Minneapolis, memo (August 25).
- RUNKLE, DAVID E. (1987). "Vector Autoregressions and Reality". *Journal of Business and Economic Statistics* 5, 437-42.
- SIMS, CHRISTOPHER A. (1982). "Policy Analysis with Economic Models". *Brookings Papers on Economic Activity*, 107-64.
- (1986a). "Are Forecasting Models Usable for Policy Analysis?". *Federal Reserve Bank of Minneapolis Quarterly Review*, Winter, pp. 2-16.
- (1986b). "BAYESMITH: A Program for Multivariate Bayesian Interpolation". University of Minnesota, Center for Economic Research, Discussion Paper 234 (September).
- (1987). "A Rational Expectations Framework for Short-Run Policy Analysis". In William A. Barnett and Kenneth J. Singleton, eds. (1987), *New Approaches to Monetary Economics*. Cambridge University Press.
- TODD, RICHARD M. (1984). "Improving Economic Forecasting with Bayesian Vector Autoregression". *Federal Reserve Bank of Minneapolis Quarterly Review*, Fall, 18-29.
- WORKING, HOLBROOK (1960). "Note on the Correlation of First Difference of Averages in a Random Chain". *Econometrica* 28, 916-18.

A BVAR FORECASTING MODEL FOR THE CHILEAN ECONOMY*

RICHARD M. TODD

Research Department
Federal Reserve Bank of Minneapolis

FELIPE G. MORANDE**

Graduate Program in Economics
ILADES/Georgetown University

Abstract:

Doan, Litterman, and Sims have described a method for estimating Bayesian vector autoregressive (BVAR) forecasting models. The method has been successfully applied to the U.S. macroeconomic dataset, which is relatively long and stable. Despite the brevity and volatility of the post-1976 Chilean macroeconomic dataset, this paper shows that a straightforward application of the DLS method to this dataset, with simple modification to allow for delays in the release of data, also appears to satisfy at least one criterion of relative forecasting accuracy suggested by Doan, Litterman, and Sims. However, the forecast errors of the Chilean BVARs are still large in absolute terms. Also, the model's coefficients change sharply in periods marked by policy shifts, such as the floating of the peso in 1982.

Introduction

The task of forecasting of main Chilean macro variables has been a hard one after decades of significant structural changes and policy shifts. However, no clear understanding on the way these changes affect forecasting modeling can be achieved without laying out the basic of a forecasting structure.

* Comments and encouragement from Christopher Sims, Felipe Montt and Valentin Carril were quite helpful. The authors, naturally, take responsibility for the paper's errors and shortcomings. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis, the Federal Reserve System, or any other organization.

** Support from FONDECYT under grant 1436/86 is gratefully acknowledged.