Forecasting in Dynamic Factor Models Using Bayesian Model Averaging

Gary Koop
Department of Economics
University of Leicester
Gary.Koop@leicester.ac.uk

and
Simon Potter
Federal Reserve Bank of New York
Simon.Potter@ny.frb.org

December 2003

Abstract: This paper considers the problem of forecasting in dynamic factor models using Bayesian model averaging. Theoretical justifications for averaging across models, as opposed to selecting a single model, are given. Practical methods for implementing Bayesian model averaging with factor models are described. These methods involve algorithms which simulate from the space defined by all possible models. We discuss how these simulation algorithms can also be used to select the model with the highest marginal likelihood (or highest value of an information criterion) in an efficient manner. We apply these methods to the problem of forecasting GDP and inflation using quarterly U.S. data on 162 time series. For both GDP and inflation, we find that the models which contain factors do out-forecast an AR(p), but only by a relatively small amount and only at short horizons. We attribute these findings to the presence of structural instability and the fact that lags of dependent variable seem to contain most of the information relevant for forecasting. Relative to the small forecasting gains provided by including factors, the gains provided by using Bayesian model averaging over forecasting methods based on a single model are appreciable.

1

1 Introduction

There has been a great deal of work in recent years which use factor models where the factors are constructed using large numbers of explanatory variables. The models used and the macroeconomic issues addressed differ across papers. However, the basic structure of these models is that there are one or a few variables of interest and a huge number of other variables which may have explanatory power for the variable(s) of interest. The information in these other variables is extracted using factor analysis.

The goal of the present paper is to show how Bayesian model averaging (BMA) can be used to address some of the problems associated with classical implementations of these models. In particular, BMA is ideally suited for the case where the number of potential explanatory variables is huge and classical approaches run into pre-test problems. Furthermore, prior information has often been used to improve forecast performance (e.g. Diebold and Pauly, 1990) and we show how BMA does this in an attractive and systemmatic manner. The ideas underlying BMA are quite general. In this paper, we develop methods for implementing BMA for the case of forecasting in dynamic factor models. Thus, one purpose of our paper is to motivate why BMA might be important and show how it can be implemented in dynamic factor models. A second purpose is to implement BMA in an empirical exercise involving forecasting two time series which are notoriously difficult to forecast: real GDP growth and inflation.

The literature on dynamic factor models is large and we will not attempt to survey it here. An important recent contribution is Stock and Watson (2002a), which carries out a forecasting exercise for a few key variables (i.e. industrial production, personal income, etc.) using up to 215 predictors. Most of the information in these predictors is extracted into a small number of factors which they call diffusion indexes. Another example is Bernanke, Boivin and Eliasz (2002) who take a VAR involving a standard set of variables (i.e. the log differences of prices and industrial production as well as the federal funds rate) and augment it with factors based on 120 other macroeconomic time series to create a so-called Factor Augmented VAR or FAVAR. This model is used to estimate the effects of monetary policy. The motivation for such a model rises from the fact that monetary shocks should reflect the actions of the Fed and Fed decisionmaking is based on information sets covering many variables other than those in a standard VAR. The fact that standard VARs often yield "price puzzles" (e.g. a finding that a contractionary monetary

shock causes prices to rise) is thought to arise from such omitted variables and FAVARs help surmount this problem.

These two examples illustrate how many of these things empirical macroeconomists want to do (e.g. forecast or identify monetary policy shocks) involve a large number of potential explanatory variables. A key empirical finding is that these potential explanatory variables tend to be highly correlated with one another and, hence, a few factors can extract most of the information contained in them (see, e.g., Giannone, Reichlin and Sala, 2002). A theoretical literature has emerged which discusses the statistical properties of various factor-based estimators (see, among many others, Bai and Ng, 2002, Boivin and Ng, 2002, Forni, Hallin, Lippi and Reichlin, 2000, Knox, Stock and Watson, 2002 and West, 2002). The empirical work we do involves macroeconomic forecasting, but we should stress that the basic methods described in this paper have relevance for any similar macroeconometric problem.

The standard approach in the relevant literature is to choose a single model and present empirical results based on this model. For instance, of the dozens or hundreds of factors created using the large number of available predictors, it is common to choose only a few using sequential testing procedures or information criteria. There are two potential problems with this approach. First, when the researcher selects a single model, statistical evidence from other plausible models is ignored. In the context of sequential hypothesis testing procedures, the pre-test problem is well-understood (see, e.g., Poirier, 1995, pages 519-523). Draper (1995) and Hodges (1987) are also important references which discuss the importance of proper treatment of model uncertainty for areas of policy analysis such as forecasting. We do not intend to survey this literature here. Suffice it to note that there are theoretical and practical reasons (some of which are discussed below) for basing inference not on a single model, but on averages across models. These have motivated an explosion of papers which use Bayesian model averaging in many fields of applied statistics (see the Bayesian Model Averaging website, http://www.research.att.com/~volinsky/bma.html, for links to many applications). However, there have been relatively few papers in econometrics which adopt Bayesian model averaging (Fernandez, Ley and Steel, 2001b and Jacobson and Karlsson, 2003 are notable exceptions). Hence, the main purpose of the present paper is to draw on ideas from the statistical literature on Bayesian model averaging and apply them to the problem of macroeconomic forecasting with factor-based models.

The second problem with the traditional approach is that evaluating an information criterion for every

3

possible model can be computationally prohibitive since, if K is the number of factors and models are defined by the inclusion or exclusion of each factor, then 2^K possible models exist. For $K > 20$ or so, direct computation of an information criterion for every model is cumbersome or impossible. Hence, it is common to order factors according to the size of their eigenvalues and just consider models where all of the first q factors are included. This reduces the size of the model space to K (and typically much less since the maximum value for information criteria often occurs for small values of q). However, in factor analysis, the size of eigenvalues is related to the amount of information extracted from the explanatory variables (not the dependent variable) and it is possible that some factors associated with large eigenvalues have no explanatory power while some with small eigenvalues do have explanatory power for the dependent variable. By searching over models defined by the first q factors, the researcher risks including irrelevant factors and missing important ones associated with small eigenvalues. Thus, a search over models which allow for non-sequential factors is potentially important. One contribution of the present paper is to adapt ideas from the Bayesian model selection literature to develop a simulation algorithm which efficiently searches over such high-dimensional model spaces to find the model with the highest marginal likelihood (or the highest information criteria). Intuitively, posterior simulation methods, which take draws of the parameters from the posterior density, are popular in Bayesian econometrics. Model selection can be done using analogous methods which simulate from model space instead of parameter space. The algorithm is constructed so as to focus on models of high probability. Thus, it is not necessary to evaluate the marginal likelihood (or an information criteria) for every model. This simulation algorithm can also be used to implement Bayesian model averaging.

We apply our Bayesian model averaging and selection methods to the problem of forecasting GDP and inflation using quarterly data on 162 time series from 1959Q1 through 2001Q1. We compare the real time forecasting performance of our methods to forecasts provided by an AR(p) and a model which simply includes the first q factors (where q is selected using marginal likelihoods). For both GDP and inflation, we find that the models which contain factors do out-forecast the AR(p), but only by a relatively small amount at short horizons. We attribute these findings to the presence of structural instability and the fact that lags of dependent variable seem to contain most of the information relevant for forecasting. Relative to the small forecasting gains provided by including factors, the gains provided by using Bayesian model

averaging over forecasting methods based on a single model are appreciable. However, we draw attention to some important issues regarding the issue of prior elicitation over model space.

2   The Model

The basic model considered in this paper is:

$$y_{t+1} = \alpha\left(L\right) y_t + \gamma\left(L\right) w_t + \varepsilon_{t+1}, \tag{2.1}$$

for $t = 1, .., T$ where y_t is a scalar dependent variable, w_t is a k_w vector of explanatory variables and $\alpha\left(L\right)$ and $\gamma\left(L\right)$ are polynomials in the lag operator of dimension p_1 and p_2. In many macroeconomic applications, standard methods for statistical inference in (2.1) are inappropriate since the number of explanatory variables is so high. In (2.1) we have $p_1 + p_2 \times k_w$ explanatory variables. In the application in Stock and Watson (2002a), $k_w = 215$, in the application in Boivin and Ng (2002) k_w is as high as 147 (although this latter paper shows how setting k_w as low as 40 actually leads to better forecasting performance). Other applications have a similarly high number of explanatory variables. In such cases, directly estimating (2.1) will yield very imprecise estimates since many of the explanatory variables will be insignificant. However, sequential testing procedures, designed to reduced the dimensionality of the problem, will be particularly unattractive due to the enormous number of tests which must be done. In frequentist statistical language, presenting results from one final model based on a series of preliminary tests will run into serious pre-testing problems and apparently significant empirical results may merely be due to data mining. In Bayesian language, such a strategy ignores model uncertainty and features such as posterior standard deviations will provide the researcher with a spurious over-confidence in empirical results. Put more informally, it is unwise to use sequential testing procedures to select a single model for two reasons. First, the model selected might not be the one which is best (where the definition of "best" depends on the metric chosen by the researcher). Second, even if the model selected is the best one, it is rarely optimal to ignore the evidence from other "not quite so good" models.

The standard approach in the macroeconometric literature (whether Bayesian or non-Bayesian) is to replace the K explanatory variables in (2.1) with a much smaller set of factors. For instance, Stock and

Watson (2002a) replace (2.1) with:

$$y_{t+1} = \alpha\left(L\right)y_t + \beta\left(L\right)f_t + \varepsilon_{t+1}, \tag{2.2}$$

where f_t is an $q-$vector of factors generate according to:

$$w_{it} = \lambda_i\left(L\right)f_t + v_{it} \tag{2.3}$$

where w_{it} is the ith element of w_t (for $i = 1,..,k_w$) and $\lambda_i\left(L\right)$ is a polynomial in the lag operator. Applications with such a specification show that some increases in forecasting performance can be achieved over ARs or low-dimensional VARs even if only a very small number of factors are used.

In this paper we consider BMA based on the the dynamic factor approach described by (2.2) and (2.3). It addresses the pre-test criticisms in a sensible manner using Bayesian model averaging (or BMA, see, e.g., Hoeting, Madigan, Raftery and Volinsky, 1999). In the following sections we describe BMA and show how it can be implemented in the context of macroeconomic forecast using dynamic factor models.

3   Bayesian Model Averaging and Selection

The literature on Bayesian model averaging and selection has burgeoned in recent years (see Hoeting et al, 1999, or Chipman, George and McCulloch, 2001, for recent surveys). The basic idea behind Bayesian model averaging can be explained quite simply: Suppose the researcher is entertaining R possible models, denoted by $M_1,...,M_R$, to learn about a quantity to be forecast y_{T+h}. If we treat y_{T+h} and M_r as random variables, the rules of conditional expectation imply that:

$$E\left(y_{T+h}|Data\right) = \sum_{r=1}^{R} p\left(M_r|Data\right) E\left(y_{T+h}|Data, M_r\right). \tag{3.1}$$

This same logic applies to functions of y_{T+h} so, for instance, we can use:

$$E\left(y_{T+h}{}^2|Data\right) = \sum_{r=1}^{R} p\left(M_r|Data\right) E\left(y_{T+h}^2|Data, M_r\right) \tag{3.2}$$

to help us calculate the posterior variance of y_{T+h}, which can then be used to calculate predictive standard deviations and quantify uncertainty about y_{T+h}. For the models considered in this paper, $p\left(M_r|Data\right)$

can be calculated for $r = 1, .., R$. As we shall see, it is also straightforward to calculate $E\left(y_{T+h}|Data, M_r\right)$ in every model.

An alternative to Bayesian model averaging is Bayesian model selection which involves simply choosing M_* which has the maximum value for $p\left(M_r|Data\right)$. Point forecasts can then be based on $E\left(y_{T+h}|Data, M_*\right)$. This is analogous to the standard frequentist approach of selecting a single model using, e.g., an information criterion and then forecasting using this model.

Bayesian model averaging has been frequently used with the linear regression model where there are a large number of potential explanatory variables. The researcher expects that only a few of these are important, but does not know which these few are. This is precisely the sort of issue which arises in macroeconomic forecasting with large panels. Roughly speaking, papers such as those mentioned in the introduction include virtually every variable which has available data and could conceivably be relevant. Stock and Watson (2002a), for instance, use 26 different real output and income variables, 28 relating to employment, 9 variables relating to retail and manufacturing sales, 22 variables relating to housing starts or sales, etc.. Most of these variables are likely unimportant and/or are highly correlated with other variables in the model. However, the researcher does not know beforehand which of these potential variables is unimportant.

In theory, if you treat models as random variables, model averaging is the correct thing to do in the sense that (3.1) follows from the rules of probability. In addition, papers such as Min and Zellner (1993) and Raftery, Madigan and Hoeting (1997) show how model averaging is optimal for forecasting in decision theory problems. In practice, Bayesian model averaging does not suffer from the criticisms associated with sequential testing procedures, since it formally includes model uncertainty in the statistical procedure. Furthermore, by putting little weight on implausible models it surmounts the problems that arise when enormous numbers of explanatory variables are directly used in a regression.

Two important issues arise when implementing Bayesian model averaging. First, the prior for the parameters must be a valid probability density function in order to yield meaningful values for $p\left(M_r|Data\right)$. This rules out the use of many noninformative priors which tend to be improper. Hence, several papers (see, among many others, Fernandez, Ley and Steel, 2001a) describe proper priors which do not require substantive amounts of subjective prior elicitation by the researcher. In this paper we use such so-called

objective or benchmark priors as well as an empirical Bayesian approach which uses the data to estimate a key prior hyperparameter. Second, it is often the case that the number of models, R, is enormous and, hence, it is not possible to evaluate $p(M_r|Data)$ and $E(y_{T+h}|Data, M_r)$ for every model. In the present application, we will define models based on the inclusion/exclusion of each explanatory variable. Thus, we have $R = 2^K$ models where K is the number of potential explanatory variables (including lags). For instance, if $K = 30$ then we have $2^{30} > 10^9$ models. Even if the computer could analyze each model in 0.001 of a second, it would take almost two years to analyze all the models. In response to this problem, a literature has developed that devises various ways of overcoming this problem through simulation over the space defined by the various models. In this paper we use the algorithm described in Clyde (1999) (which is similar to the importance sampling algorithm described in Clyde, DeSimone and Parmigiani, 1996) which should be particularly well-suited to the problem at hand. Intuitively, this is an algorithm which draws models from $p(M_r|Data)$. In this way, the algorithm attaches more weight to the models with high probability (which are drawn more often) and less weight to implausible models. Even this algorithm is computationally quite intensive and, hence, it is important to stay within a class of models where the marginal likelihood can be calculated analytically.[1] For this reason, in our empirical work we stay within the framework of the Normal linear regression model with natural conjugate prior. The remainder of this section describes the details involved when adopting this approach.

As an aside, we should mention that the class of dynamic factor models does not allow for the non-linearities or structural instabilities which some have found to be present in many macroeconomic time series.[2] Bayesian analysis of these models typically involves computationally demanding posterior simulation, precluding BMA over a model space of the dimension considered in this paper. Thus, in practice, the researcher is often faced with the choice of giving a proper treatment of the problems caused by the huge number of potential explanatory variables (within a linear framework which may be restrictive) or selecting a greatly reduced set of models within a potentially more realistic framework which allows for

---

[1] The marginal likelihood, which is $p(Data|M_r)$, is a key component of $p(M_r|Data)$ since the rules of probability imply

$$p(M_r|Data) \propto p(M_r)\, p(Data|M_r).$$

$p(M_r)$ is the prior model probability discussed below.

[2] Koop and Potter (2001) contains the authors' thoughts on nonlinearities and structural instabilities and provides citations to some of the voluminous literature on nonlinear time series and structural breaks.

nonlinearities or structural instabilities. In this paper, we investigate the former strategy. Koop and Potter (1999) investigates the latter strategy (in a different empirical context not involving forecasting). Our previous paper develops a nonlinear time series model involving different regimes. There are several ways of defining these regimes and we average over these definitions using Bayesian methods. Insofar as one interprets the different regimes as defining different models, our previous work is a BMA piece involving a small set of nonlinear time series models.

In this paper, we always condition on the first $\max(p_1, p_2)$ observations which we denote by $0, -1, .., 1 - \max(p_1, p_2)$. With this convention, we define $y = (y_2, .., y_{T+1})'$, $\varepsilon = (\varepsilon_1, .., \varepsilon_T)'$, X to be the $T \times K$ matrix containing all potential lagged explanatory variables and write (2.1) as:

$$y = X\beta + \varepsilon. \tag{3.3}$$

We stress that the tth row of y contains data available at time $t + 1$ while the tth row of X contains data available at time t. Following standard practice (see, e.g., Stock and Watson, 2002a) all variables are transformed to stationarity (see the Data section and Data Appendix for more details). When forecasting h periods in the future, y is redefined as $y = (y_{1+h}, .., y_{T+h})'$.

In our application, we wish to include variables which are common to every model (i.e. an intercept and lags of the dependent variable). An attractive way of treating such variables (see, e.g., Chipman, George and McCulloch, 2001) is to integrate them out using a noninformative prior. This is equivalent to removing the linear effect of these common variables on the dependent and other explanatory variables. To be precise, if y^* and X^* are the original dependent and potential explanatory variables and X^{**} contains a set of explanatory variables common to all models, then in (3.3) we work with $y = \left[I - X^{**} \left(X^{**\prime} X^{**} \right)^{-1} X^{**\prime} \right] y^*$ and $X_i = \left[I - X^{**} \left(X^{**\prime} X^{**} \right)^{-1} X^{**\prime} \right] X_i^*$ where X_i and X_i^* are the ith columns of X and X^*, respectively.

We assume

$$\varepsilon \sim N\left(0, \sigma^2 I_T\right). \tag{3.4}$$

We use a natural conjugate prior:

$$\beta|\sigma^{-2} \sim N\left(\underline{\beta}, \sigma^2 \underline{B}\right) \tag{3.5}$$

and

$$\sigma^{-2} \sim G\left(\underline{s}^{-2}, \underline{\nu}\right) \tag{3.6}$$

where $G\left(\underline{s}^{-2}, \underline{\nu}\right)$ denotes the Gamma distribution with mean $\underline{s}^{-2}$ and degrees of freedom $\underline{\nu}$ (see Poirier, 1995, page 100). The prior hyperparameters $\underline{\beta}, \underline{B}, \underline{s}^{-2}, \underline{\nu}$ will be discussed below.

The algorithm described in Clyde (1999, section 3) requires the explanatory variables to be orthogonal to one another. Accordingly , we use an orthogonal transformation of (3.3) of the sort used in the factor analysis literature. That is, if we define $Z = XW$ where W is a nonsingular $K \times K$ matrix chosen so that the columns of Z are orthogonal we can write (3.3) as

$$y = Z\alpha + \varepsilon, \tag{3.7}$$

where $\alpha = W^{-1}\beta$.

There are many ways of choosing a nonsingular W such that $Z = XW$ and the columns of Z are orthogonal. However, in the present application, a very logical choice suggests itself to us. We construct W using principal components, implying a factor structure similar to that used by others in the field (e.g. Stock and Watson, 2002). That is, W is simply the matrix of eigenvectors of $X'X$. With macroeconomic panels, it is common for $K \geq T$ and the last $K - T + 1$ columns of Z are equal to zero. If this occurs, then we delete these last columns of Z. In our empirical work, the condition $K \geq T$ usually holds and, hence, Z is usually a $(T-1) \times (T-1)$ matrix. In the remainder of the paper we use K to denote the number of potentially explanatory variables actually included in the model and, thus, it is typically the case that $K = T - 1$.

The prior for σ^{-2} is unaffected by the transformation from (3.3) to (3.7) and the prior for the regression coefficients becomes

$$\alpha|\sigma^{-2} \sim N\left(\underline{\alpha}, \sigma^2 \underline{A}\right) \tag{3.8}$$

where $\underline{\alpha} = W^{-1}\underline{\beta}$ and $\underline{A} = W^{-1}\underline{B}\left(W^{-1}\right)'$. Formulae for the posterior and marginal likelihood for this model are available in any standard Bayesian textbook (e.g. Poirier, 1995, pages 526 and 543). The predictive density is discussed in the Technical Appendix.

Different models are defined through a $K \times 1$ vector γ with all elements equalling either zero or one. If the jth element of γ is zero, then the jth column of Z is deleted from the model (as is the jth element of $\underline{\alpha}$ and the jth row and column of $\underline{A}$). There are 2^K possible configurations for γ and this defines the set of all possible models. Clyde (1999) shows how $p\left(\gamma|y\right)$ takes a simple form and, thus, Monte Carlo methods used to take posterior draws of γ if σ^2 is known. When σ^2 is unknown a Gibbs sampler can be set up which sequentially draws from $p\left(\gamma|y, \sigma^2\right)$ and $p\left(\sigma^2|y, \gamma\right)$. This algorithm (which hinges on the fact that the explanatory variables are orthogonal), is very efficient in that $p\left(\gamma|y, \sigma^2\right)$ is directly drawn from. This contrasts with other common algorithms, such as Markov Chain Monte Carlo Model Composition (MC^3) which draw from $p\left(\gamma_j|y, \sigma^2.\gamma^{(-j)}\right)$ where $\gamma^{(-j)} = \left(\gamma_1, .., \gamma_{j-1}, \gamma_{j+1}, .., \gamma_K\right)'$ and typically produce a highly correlated sequence of drawn models.

The output from our algorithm can be used either to carry out Bayesian model averaging or model selection. In the latter case, the model with the highest value for $p\left(M_r|y\right)$ (or, equivalently, the highest value for $p\left(\gamma|y\right)$) is selected. For every drawn model, the predictive mean and second moment are obtained using the formulae in the Technical Appendix. These can then be averaged in the same way as posterior simulator output. For instance, if $E\left(y_{\tau+1}|Data, M^{(s)}\right)$ is the predictive mean obtained when model $M^{(s)}$ is drawn (for $s = 1, .., S$) based on $Data$ containing data through period τ, then

$$\frac{1}{S}\sum_{s=1}^{S} E\left(y_{\tau+1}|Data, M^{(s)}\right)$$

will converge to $E\left(y_{\tau+1}|Data\right)$ as S goes to infinity.

We do not provide all the technical details here since our implementation is the same as that described in Clyde (1999, section 3). Briefly, an analytical form for $p\left(\gamma|y, \sigma^2\right)$ can be derived given a prior, $p\left(\gamma\right)$, using the fact that $p\left(y|\gamma, \sigma^2\right)$ is simply the marginal likelihood for the Normal linear regression model defined by γ. $p\left(\sigma^2|y, \gamma\right)$ takes the usual inverted-Gamma form for the Normal linear regression model defined by γ.

It remains to specify the prior model probability, $p(M_r)$ (or, equivalently, a prior for γ) and choose $\underline{\beta}, \underline{B}, \underline{s}^{-2}, \underline{\nu}$ and W. A standard choice is

$$p(\gamma) = \prod_{j=1}^{K} \theta_j^{\gamma_j} (1 - \theta_j)^{(1-\gamma_j)} \tag{3.9}$$

where θ_j for $j = 1, .., K$ is the prior probability that each potential explanatory variable enters the model. A common noninformative benchmark case sets $\theta_j = \frac{1}{2}$, a value which implies $p(M_r) = \frac{1}{R}$ for $r = 1, .., R$. However, other priors are possible. For instance, one might expect factors corresponding to higher eigenvalues of $X'X$ to be more relevant than factors with low eigenvalues. We can incorporate this by allowing θ_j to depend on v_j, the jth largest eigenvalue of $X'X$. Thus, some of the empirical work we do assumes:

$$\theta_j = \frac{v_j}{v_1}. \tag{3.10}$$

We also work with what we call a 99.9% prior which sets $\theta_j = \frac{1}{2}$ for $j = 1, .., K_{99.9}$, where 99.9% of the variation in X is contained in the first $K_{99.9}$ factors. Thus, this third prior discards the factors associated with very small eigenvalues but is otherwise noninformative.

The remaining prior hyperparameters are chosen using a strategy suggested in, among other places, Fernandez, Ley and Steel (2001a). This is based on the insight that using a noninformative, improper, prior over parameters common to all models is an acceptable practice (see, e.g., Kass and Raftery, 1995, page 783). Hence, we choose a noninformative prior for σ^{-2} (i.e. $\underline{\nu} = 0$ and, with this choice, $\underline{s}^{-2}$ does not enter the marginal likelihood or posterior). For $\underline{\beta}$ and $\underline{B}$ it is not acceptable to make noninformative choices since Bayes factors are either degenerate or depend on arbitrary normalizing constants. Following Fernandez, Ley and Steel (2001a), we center the prior for these regression coefficients over zero (i.e. $\underline{\beta} = 0_K$) and use a g-prior form for $\underline{B}$ (see Zellner, 1986). The g-prior reduces the choice of the $K \times K$ prior covariance matrix $\underline{B}$ to a single scalar hyperparameter g by setting:

$$\underline{B} = (gX'X)^{-1}. \tag{3.11}$$

The use of a g-prior is common in Bayesian model averaging and justification for a prior of this choice is

given in many papers (e.g. Fernandez, Ley and Steel, 2001a or Zellner, 1986). See also Diebold and Pauly (1990) which uses the g-prior in a forecasting exercise not involving BMA. In essence, this allows prior information to have the same scale as likelihood information. In the present context, the g-prior cannot be used if the original number of explanatory variables exceeds T since then $X'X$ is singular. We get around this problem by working directly with a g-prior for (3.7) and, thus,

$$\underline{A} = (gZ'Z)^{-1}. \tag{3.12}$$

We should digress for a moment to discuss some issues relating to the g-prior. In the Normal linear regression model, having the prior depend on explanatory variables is acceptable. That is, the likelihood function and posterior are defined conditionally upon Z (and, hence, X), so having X in the prior does not violate the rules of conditional probability. Having the prior depend on y would violate the rules of conditional probability, but our prior does not have this property. Furthermore, the structure of the prior covariance matrix in (3.10) is similar to the observed information matrix (for stationary data). It is common to elicit priors related to the information matrix. Thus, there is a sense in which the g-prior incorporates the time series structure implied by inclusion of lagged explanatory variables. We use a noninformative prior for the lagged dependent variables, which are included in all models.

It remains to specify g. Fernandez, Ley and Steel (2001a) investigate the properties of many possible choices for g and show that some of them yield posterior model probabilities which have properties similar to commonly-used information criteria. In an objective Bayesian spirit, we focus on such values for g. In particular, we consider three different values:

$$g = \frac{1}{T}, \tag{3.13}$$

$$g = \frac{1}{[\ln(T)]^3}, \tag{3.14}$$

and

$$g = \frac{1}{K^2}. \tag{3.15}$$

13

Fernandez, Ley and Steel (2001a) show that the values for g in (3.13) and (3.14) yield consistent model selection criteria (i.e. asymptotically the model which maximizes $p(M_r|y)$ will be the correct one, insofar as there is a single correct model). Furthermore, logs of Bayes factors obtained using (3.13) and (3.14) behave asymptotically like the Schwarz and Hannan-Quinn (with $C_{HQ} = 3$) criteria, respectively. The value for g given in (3.15) implies the Risk Inflation Criterion (RIC) of Foster and George (1994) who also provide motivation for this choice. Further motivation for (3.13) can be found in Kass and Wasserman (1995) who refer to a similar prior as a "unit information prior" and relate it to the intrinsic and fractional Bayes factors literature (see Berger and Pericchi, 1996). It is worth stressing that Bayesian model selection done using these values for g is comparable to traditional model selection approaches using information criteria. In fact, a sensible, albeit ad hoc, non-Bayesian model averaging procedure could be done based on information criteria with $p(M_r|Data)$ in (3.1) replaced by an information criteria (normalized so that the weights used in the model averaging procedure sum to one).

In addition to such information criteria-based choices for g, we also consider choosing g in a data based manner. Such an empirical Bayesian methodology can be criticized on the grounds that allowing a prior to depend on y violates the rules of conditional probability. Nevertheless, empirical Bayesian methods are popular with many practical econometricians (see, e.g., Knox, Stock and Watson, 2002). There are various approaches which use the label "empirical Bayes". The most common empirical Bayesian methods choose the value of a single prior hyperparameter (here this is g) which maximizes the marginal likelihood. Here we adopt such an approach.

4 Data

We use 162 U.S. time series from 1959Q1 through 2001Q1 which are essentially the same as those listed in Stock and Watson (2002b). Our paper differs with related work (e.g. Stock and Watson, 2002a) in that we use quarterly data which allows us to forecast the series of most popular interest, GDP. Most previous work has used monthly data and focussed on series such as industrial production or personal income (partly because GDP is difficult to forecast). In addition to forecasting GDP, we also forecast inflation. Note that other researchers have found that it is harder to forecast price variables than real variables, so that the two

series we have chosen should be ones that are difficult to forecast. We do this to ensure that our results are not perceived as due to our searching over 162 variables in order to find some which support our case.

The complete list of variables, along with brief descriptions and DRI acronyms, are given in the Data Appendix. Here suffice it to note that our quarterly GDP measure is GDPQ and we use PUNEW, the CPI (all items), as our price measure. In order to avoid modeling issues relating to unit roots and cointegration (and following standard practice see, e.g., Stock and Watson, 2002a), we induce stationarity in all our variables. The exact transformation used for every series is given in the Data Appendix. For our forecasted series, we take first and second differences, respectively, of the logs of GDPQ and PUNEW. Thus, in our empirical work, including the forecasting exercise, we are working with the GDP growth and the growth of inflation.

5 Empirical Results

We investigate three types of econometric procedure: Bayesian model averaging, Bayesian model selection using the model which maximizes $p\left(M_r | Data\right)$ and a conventional model selection procedure where we simply consider models with the first q factors included and choose the value of q which maximizes the marginal likelihood (we search over values of q up to 20). Given the relationship between marginal likelihoods and information criteria for the benchmark priors used in this paper (see the discussion after equation 3.15), our two model selection procedures are comparable to traditional approaches using information criteria. Note, in particular, that even for the reader uninterested in model averaging, our computational algorithm provides for a very efficient search over models with non-sequential factors (i.e. there will be 2^K such models and evaluating an information criteria for each will be computationally infeasible if K is at all large).

With regards to the prior for the parameters of the models, we use the g-prior of (3.12) with the three values for g given in (3.13) through (3.15). In addition, we choose g using empirical Bayesian methods. We use three priors over model space (see equation 3.9). One is noninformative over all factors (i.e. $\theta_j = \frac{1}{2}$ for $j = 1,..,K$), one is noninformative over the first q factors, where q is chosen such that 99.9% of the variation in X is included (we refer to this as the 99.9% prior) and the other is informative with probability

allocated to the coefficient on each factor proportional to the size of its eigenvalue (i.e. θ_j given in equation 3.10 for $j = 1, .., K$).

Throughout we compare our results to an AR(p) base case (using the standard noninformative prior). We choose the value of p which yields the lowest forecast root mean squared error. In this way, we are allowing the benchmark AR(p) to perform in the most favorable manner. For both GDPQ and PUNEW we find $p = 2$. We also use two lags of all potential explanatory variables when we construct the factors. In the following tables, results relating to GDPQ are presented in tables with "a" suffixes, while results for PUNEW have "b" suffixes.

5.1 Estimation and Model Comparison using the Entire Sample

Before carrying out a real time forecasting exercise, it is useful to present estimation and model comparison results using the full sample. Tables 1a and 1b present the logs of the marginal likelihoods from the Bayesian model selection and averaging exercises as well as for models where the first q factors are included and the AR(2). The BMA marginal likelihood is a weighted average of all the marginal likelihoods where the weight associated with the model r is $p(M_r|Data)$.

A word of caution is called for when interpreting the BMA marginal likelihood. When comparing two single models (e.g. a model with two factors to one with three factors), a ratio of marginal likelihoods can be used in the standard way as reflecting the relative plausibility of the two models. It is more difficult to compare an average across models to a single model. If we define M to be the "BMA model" which is the super-model which averages over all models, then the rules of probability imply the "BMA marginal likelihood" should be:

$$p(Data|M) = \sum_{r=1}^{R} p(Data|M_r, M) p(M_r|M).$$

In other words, theory suggests that we average across models using the prior, $p(M_r|M)$. However, given the enormous number of models under consideration, this prior is very small (e.g. $\frac{1}{2^K}$) and our "BMA marginal likelihoods" would be massively smaller than the marginal likelihoods of single models. Using such a measure to compare BMA to the model selection procedure would be unfair since the selected model was chosen on the basis of a pre-test procedure. One possibility would be to weight the marginal likelihood

16

of the selected model by its prior model probability (and the reader may do this if desired). Given that model comparison is not the focus of this paper (and, indeed, as we will see, the marginal likelihoods do not shed great light on predictive performance), we do not delve into these issues deeply here. Rather we prefer to interpret the BMA marginal likelihood presented in Tables 1a and 1b in a more informal measure as reflecting the plausibility of the models drawn in our simulation algorithm. It is also worthwhile to keep in mind that, by definition, the marginal likelihood presented for our model selection procedure will always be greater than the comparable BMA marginal likelihood (i.e. the maximum value of $p\left(Data|M_r\right)$ will always be bigger than a weighted average across models if the weights sum to one).

Table 1a: Log Marginal Likelihoods for Different Priors, GDPQ					
	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g	Value of Optimal g
BMA 99.9% prior	374.6	374.8	374.9	378.3	0.14
Model Selection 99.9% prior	376.5	376.8	376.9	380.6	0.13
BMA $\theta_i = \frac{1}{2}$	375.9	388.7	364.3	412.0	0.03
Model Selection $\theta_i = \frac{1}{2}$	409.8	415.1	373.6	430.8	0.02
BMA θ_i in (3.10)	375.7	375.9	370.4	377.9	0.08
Model Selection θ_i in (3.10)	378.6	378.4	370.7	382.9	0.08
One Factor	365.7	365.8	367.0	367.1	0.37
Two Factors	365.5	368.2	368.2	368.3	0.30
Three Factors	365.4	365.8	369.0	369.4	0.28
Four Factors	376.2	376.6	379.6	379.7	0.10
Five Factors	374.7	375.2	378.6	379.5	0.11
Six Factors	373.8	374.4	377.8	379.8	0.13
Seven Factors	373.2	373.9	377.0	380.5	0.14
Eight Factors	370.8	371.7	374.3	379.6	0.17
Nine Factors	368.3	369.2	371.3	378.7	0.19
Ten Factors	365.8	366.8	368.2	374.8	0.22
Twenty Factors	353.3	355.4	344.6	379.5	0.30
AR(2)	358.2	358.4	354.4	358.8	0.02

	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g	Value of Optimal g
Table 1b: Log Marginal Likelihoods for Different Priors, PUNEW					
BMA 99.9% prior	517.7	518.0	518.2	520.3	0.07
Model Selection 99.9% prior	519.5	519.9	520.1	522.1	0.07
BMA $\theta_i = \frac{1}{2}$	488.3	498.7	502.1	546.2	0.02
Model Selection $\theta_i = \frac{1}{2}$	544.7	556.4	510.8	567.1	0.02
BMA θ_i in (3.10)	515.0	515.5	502.1	518.5	0.07
Model Selection θ_i in (3.10)	518.8	520.8	506.4	524.6	0.07
One Factor	496.5	496.6	498.5	498.6	1.00
Two Factors	504.1	504.3	505.4	505.8	0.10
Three Factors	509.8	510.1	512.0	512.1	0.08
Four Factors	507.3	507.6	510.7	510.9	0.11
Five Factors	508.2	508.7	512.0	512.8	0.11
Six Factors	515.1	515.7	518.8	520.1	0.09
Seven Factors	513.0	513.7	516.6	519.2	0.10
Eight Factors	510.4	511.2	513.8	518.0	0.12
Nine Factors	508.2	509.1	511.1	517.2	0.13
Ten Factors	505.7	506.7	508.0	516.2	0.15
Twenty Factors	495.7	497.8	487.1	519.0	0.21
AR(2)	490.8	491.1	493.3	493.4	0.10

Tables 1a and 1b show how Bayesian model selection as described in the previous section can yield models with much higher marginal likelihoods than simply choosing q lags of factors (marginal likelihoods become smaller for $q > 20$ so we only present results up to $q = 20$). For both of our time series, for each of the four different choices for g (and all three of the priors used for γ), the difference between the log marginal likelihood for the model chosen using Bayesian model selection and a traditional model with q factors is usually substantial. For instance, when we use a flat prior over model space and empirical Bayes methods to estimate g, the relevant Bayes factor in favor of the model chosen Bayesian model selection is at least e^{60} for GDP and e^{40} for inflation (and usually much more than this). The log marginal likelihoods obtained using BMA are not quite as high, but this is to be expected as BMA averages across models as opposed to selecting the model with highest marginal likelihood. For this reason, it is hard to compare the

BMA log marginal likelihoods with the other numbers in the table. Nevertheless, results for BMA indicate that the algorithm is, as it should, selecting the model with the highest marginal likelihood and averaging across models with high marginal likelihoods.

Note that the results for the case $g = \frac{1}{K^2}$ are a little hard to interpret since K is so different across models (i.e. $K = T - 1$ for two of the Bayesian model averaging cases, whereas $K = q$ for the models where the first q factors are chosen). Note also that, especially for PUNEW and regardless of which prior we use, the relationship between q and the marginal likelihood is quite nonlinear. For PUNEW, the marginal likelihoods increase by a large amount when the third factor is added, then fall, then show another big jump when the sixth factor is added. This indicates that the third and sixth factors are quite important, but that some of the other factors are not very important (findings supported by the Bayesian model selection results). The conventional strategy of simply selecting the first q factors thus involves including some irrelevant factors in this case.

Tables 2a and 2b contain information relating to the number of factors included. In the case of Bayesian model selection this is simply the number of factors included in the selected model, while for BMA Tables 2a and 2b presents the comparable quantity, $E\left(\sum_{j=1}^{K} \gamma_j | Data\right)$. These tables shed a great deal of light on the results of Tables 1a and 1b. When we use a flat prior over the model space (i.e. $\theta_i = \frac{1}{2}$), then, a priori, it is just as likely that a factor associated with a small eigenvalue enters as one with a large eigenvalue. Using either BMA or Bayesian model selection, very many factors are chosen to enter the model. The one exception to this is when $g = \frac{1}{K^2}$. Since $K = T - 1$, this value implies very large prior variances on the coefficients. It is well-known that (see, e.g., Poirier, 1995), in the Normal linear regression model with natural conjugate prior, as the prior variance of a coefficient goes to infinity, the Bayes factor in favor of the coefficient equalling zero goes to infinity. In other words, by increasing the prior variance, the reward for parsimony is increased. This accounts for the fact that so few factors are chosen when $g = \frac{1}{K^2}$. In case the reader is puzzled by the fact that 100 or more factors are usually chosen when the prior is flat over model space, it is worth mentioning that non-Bayesian results indicate a similar pattern. For instance, a simple OLS regression of GDP growth on the first 100 factors yields t-statistics on 46 coefficients which are greater than one and 15 which are greater than two (in absolute value). A regression containing the first 10 factors has $R^2 = 0.232$, while the regression containing the first 100 factors has $R^2 = 0.777$ (even

19

though the ratio of the 100th to the first eigenvalue is 1.04×10^{-7}). Thus, it appears that, although most of the information in X is contained in the first few factors, many of the factors associated with lower eigenvalues do have significant in-sample explanatory power for real GDP growth. Presumably, there is so much information contained in X that even a very small amount of it can have useful explanatory power. Similar statements can be made for PUNEW.

The use of the priors over model space given in (3.10) and the 99.9% prior effectively rule out most of the factors associated with small eigenvalues and, hence, the marginal likelihood results of Tables 1a and 1b are more similar to those for a traditional factor model with q selected using, e.g., an information criteria. For instance, for PUNEW with the 99.9% prior and g chosen to maximize the marginal likelihood, the Bayesian model selection procedure chooses six non-sequential factors (the factors associated with the eigenvalues ranked 1, 2, 3, 5, 6 and 11th). The results in Table 1b indicate this model is to be preferred over a model which simply chooses, e.g., the first 6 factors.

Table 2a: Number of Factors in Model, GDPQ				
	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g
BMA 99.9% prior	3.7	3.9	3.9	6.12
Model Selection 99.9% prior	3	3	3	6
BMA $\theta_i = \frac{1}{2}$	90.5	101.5	4.1	95.4
Model Selection $\theta_i = \frac{1}{2}$	118	122	1	98
BMA θ_i in (3.10)	2.2	2.3	2.0	2.6
Model Selection θ_i in (3.10)	5	4	2	5

Table 2b: Number of Factors in Model, PUNEW				
	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g
BMA 99.9% prior	4.6	4.7	4.8	5.9
Model Selection 99.9% prior	4	4	4	5
BMA $\theta_i = \frac{1}{2}$	38.3	62.9	5.9	100.5
Model Selection $\theta_i = \frac{1}{2}$	118	125	3	102
BMA θ_i in (3.10)	3.9	3.9	2.9	4.0
Model Selection θ_i in (3.10)	6	6	4	7

In this sub-section, we have shown how the Bayesian model selection procedure can be used to find models with much higher marginal likelihoods than a traditional strategy of simply choosing q. Furthermore, BMA is averaging over models which include many more factors than are traditionally chosen in comparable forecasting exercises. However, in-sample performance does not necessary imply good forecasting performance and it is to this we now turn.

5.2 Forecasting Exercise

The forecasting exercise we carry out has a similar structure to that done by others in the literature (e.g. Stock and Watson, 2002a). That is, we do simulated real time forecasting from 1970Q1 through 2000Q4 of $y_{\tau+h}$ for horizons $h = 1, 4$ and 8. To be precise, we do Bayesian model averaging and selection using the model given in (3.7) for every period from 1970Q1 to the end of the sample. At any point forecast time, τ, the matrix of factors contains data through the time the forecast is made. We will denote this matrix of factors as Z_τ where τ is the time the forecast is being made (i.e. the last row of Z_τ contains data from period τ). The tth row of y contains the value of GDP growth in period $t + h$. Thus, for each variable, we carry out nearly 500 BMA exercises for each choice of g (i.e. roughly 160 choices for τ and three choices for h). When calculating the optimal g, we do a grid search at each τ. Although computationally demanding, this is well within the power of modern personal computers. Note that in the rows labelled "Model with First q Factors Selected" we select a potentially different value for q for each value of τ and g.

Tables 3a and 3b present some diagnostics on the point estimates of the forecasts (i.e. the predictive means) using the root mean squared forecast error as a metric:

$$RMSE = \sqrt{\sum_{\tau=1969:4}^{2001:1-h} \left[y_{\tau+h} - E\left(y_{\tau+h}|Z_\tau\right)\right]^2}.$$

To aid in interpretation, we normalize the sum of squared forecast errors relative to that provided by an AR(2). Thus, e.g., an entry in Table 3a of 95.0 implies that the RMSE of the relevant model is 95% as large as in an AR(2).

Tables 4a and 4b present some evidence relating to coverage of forecast intervals. The numbers in this table are the percentage of times the actual value of the time series lies within the interval containing the predictive mean plus/minus two predictive standard deviations. Note that, when a single model is selected, the predictive follows a t-distribution with τ degrees of freedom. Since τ is 38 or more, the predictive should be roughly Normal and, hence, a two predictive standard deviation interval should approximate a 95% highest predictive density interval. When Bayesian model averaging is done, the predictive will be a mixture of t-distributions. Tables 5a and 5b present information on the number of factors found to be important (see the discussion of Tables 2a and 2b for additional explanation).

	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g
Table 3a: RMSE relative to AR(2), percentage, GDPQ				
Bayesian Model Averaging (equal prior weights to all models)				
$h = 1$	171.1	172.7	102.6	173.7
$h = 4$	188.2	190.9	99.4	186.1
$h = 8$	246.3	248.3	131.0	248.5
Bayesian Model Selection (equal prior weights to all models)				
$h = 1$	181.7	184.0	103.7	176.9
$h = 4$	194.1	194.1	109.5	188.6
$h = 8$	254.2	252.5	123.4	249.2
Bayesian Model Averaging (model prior using 3.10)				
$h = 1$	99.5	99.5	99.9	98.2
$h = 4$	100.7	100.8	101.5	100.1
$h = 8$	100.2	100.2	100.1	100.1
Bayesian Model Selection (model prior using 3.10)				
$h = 1$	97.6	99.3	100.4	97.0
$h = 4$	101.4	103.1	101.6	110.1
$h = 8$	107.5	107.7	105.4	114.8
Bayesian Model Averaging (99.9% prior)				
$h = 1$	94.1	94.1	94.2	93.0
$h = 4$	100.1	100.2	100.1	99.6
$h = 8$	99.0	99.1	99.1	99.2
Bayesian Model Selection (99.9% prior)				
$h = 1$	96.5	96.1	95.6	93.1
$h = 4$	101.5	101.8	101.4	99.7
$h = 8$	100.5	100.7	100.5	101.8
Model with First q Factors Selected				
$h = 1$	94.9	94.8	94.3	94.6
$h = 4$	99.4	99.4	97.9	100.7
$h = 8$	100.4	100.4	100.5	100.5

	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g
Table 3b: RMSE relative to AR(2), PUNEW				
Bayesian Model Averaging (equal prior weights to all models)				
$h = 1$	120.6	121.9	92.4	121.4
$h = 4$	141.7	142.6	104.0	143.5
$h = 8$	150.9	154.8	102.9	158.5
Bayesian Model Selection (equal prior weights to all models)				
$h = 1$	131.5	131.0	95.2	130.6
$h = 4$	144.6	142.7	106.3	145.1
$h = 8$	159.8	162.1	106.2	163.3
Bayesian Model Averaging (model prior using 3.10)				
$h = 1$	95.1	95.1	97.1	94.8
$h = 4$	99.7	99.7	99.7	99.8
$h = 8$	100.0	100.0	100.0	100.1
Bayesian Model Selection (model prior using 3.10)				
$h = 1$	91.2	91.9	95.1	93.3
$h = 4$	105.0	105.7	102.7	107.5
$h = 8$	103.2	102.2	100.8	103.7
Bayesian Model Averaging (99.9% prior)				
$h = 1$	91.2	91.3	91.4	88.2
$h = 4$	100.8	100.8	100.8	101.1
$h = 8$	100.9	100.9	100.9	100.7
Bayesian Model Selection (99.9% prior)				
$h = 1$	93.5	94.6	94.7	90.0
$h = 4$	102.4	100.9	100.8	103.4
$h = 8$	100.6	100.6	100.6	101.4
Model with First q Factors Selected				
$h = 1$	92.7	93.4	94.1	89.2
$h = 4$	99.6	99.6	101.3	101.3
$h = 8$	100.0	100.0	100.0	100.0

	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g
Table 4a: Percentage of Predictive Means within 2 Standard Deviations of Actual Value, GDPQ				
Bayesian Model Averaging (equal prior weights to all models)				
$h = 1$	58.3	57.5	96.1	52.0
$h = 4$	33.1	28.2	95.2	37.1
$h = 8$	37.5	32.5	91.7	33.3
Bayesian Model Selection (equal prior weights to all models)				
$h = 1$	39.4	33.1	93.7	40.9
$h = 4$	21.8	20.2	87.1	29.0
$h = 8$	21.7	19.2	85.8	23.3
Bayesian Model Averaging (model prior using 3.10)				
$h = 1$	95.3	95.3	96.1	95.3
$h = 4$	91.1	91.1	91.1	92.7
$h = 8$	94.2	94.2	94.2	94.2
Bayesian Model Selection (model prior using 3.10)				
$h = 1$	93.7	94.5	94.5	92.9
$h = 4$	91.1	91.9	91.3	91.1
$h = 8$	90.8	90.0	90.8	90.8
Bayesian Model Averaging (99.9% prior)				
$h = 1$	96.1	96.1	96.1	96.1
$h = 4$	91.9	91.9	91.9	91.9
$h = 8$	95.0	95.0	95.0	95.0
Bayesian Model Selection (99.9% prior)				
$h = 1$	95.3	95.3	95.3	95.3
$h = 4$	91.1	91.1	91.9	91.9
$h = 8$	95.0	95.0	95.0	94.2
Model with First q Factors Selected				
$h = 1$	96.1	96.1	95.3	95.2
$h = 4$	91.1	91.1	91.9	91.1
$h = 8$	93.3	93.3	94.2	91.7

	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g
Table 4b: Percentage of Predictive Means within 2 Standard Deviations of Actual Value, PUNEW				
Bayesian Model Averaging (equal prior weights to all models)				
$h = 1$	63.0	58.3	92.1	58.3
$h = 4$	44.4	40.3	90.3	41.9
$h = 8$	31.7	32.5	89.2	46.7
Bayesian Model Selection (equal prior weights to all models)				
$h = 1$	26.8	22.8	86.6	36.2
$h = 4$	33.9	33.1	85.5	34.7
$h = 8$	25.8	23.3	83.3	35.0
Bayesian Model Averaging (model prior using 3.10)				
$h = 1$	89.0	89.0	90.4	89.0
$h = 4$	87.1	87.1	87.1	87.9
$h = 8$	86.7	86.7	86.7	86.7
Bayesian Model Selection (model prior using 3.10)				
$h = 1$	88.2	87.4	85.0	87.4
$h = 4$	86.3	85.5	86.3	83.1
$h = 8$	84.2	85.0	86.7	84.2
Bayesian Model Averaging (99.9% prior)				
$h = 1$	88.2	88.2	87.4	89.2
$h = 4$	87.9	87.9	87.9	87.1
$h = 8$	89.2	87.5	87.5	87.5
Bayesian Model Selection (99.9% prior)				
$h = 1$	86.6	86.6	86.6	87.4
$h = 4$	87.9	87.9	87.9	86.3
$h = 8$	87.5	87.5	87.5	87.5
Model with First q Factors Selected				
$h = 1$	87.2	87.2	85.8	88.8
$h = 4$	87.1	87.1	87.9	87.1
$h = 8$	86.7	86.7	86.7	86.7

	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g
Table 5a: Number of Factors in Model (average over all τ), GDPQ				
$E\left(\sum_{j=1}^{K} \gamma_j \middle| Data\right)$ for Bayesian Model Averaging (equal prior weights to all models)				
$h = 1$	47.8	50.6	3.9	63.7
$h = 4$	67.2	70.5	3.9	64.0
$h = 8$	63.5	67.0	3.1	66.0
$\sum_{j=1}^{K} \gamma_j$ for Selected Model for Bayesian Model Selection (equal prior weights to all models)				
$h = 1$	55.0	60.8	1.6	66.9
$h = 4$	70.9	72.3	2.0	62.9
$h = 8$	65.2	67.6	1.1	66.4
$E\left(\sum_{j=1}^{K} \gamma_j \middle| Data\right)$ for Bayesian Model Averaging (model prior using 3.10)				
$h = 1$	1.8	1.8	1.8	2.1
$h = 4$	1.8	1.8	1.4	1.9
$h = 8$	1.1	1.1	1.0	1.2
$\sum_{j=1}^{K} \gamma_j$ for Selected Model for Bayesian Model Selection (model prior using 3.10)				
$h = 1$	3.5	3.6	2.0	4.5
$h = 4$	3.0	3.0	1.9	3.6
$h = 8$	1.5	1.6	1.2	2.9
$E\left(\sum_{j=1}^{K} \gamma_j \middle| Data\right)$ for Bayesian Model Averaging (99.9% prior)				
$h = 1$	3.4	3.4	3.4	5.2
$h = 4$	3.4	3.4	3.4	5.2
$h = 8$	1.9	1.9	1.9	4.0
$\sum_{j=1}^{K} \gamma_j$ for Selected Model for Bayesian Model Selection (99.9% prior)				
$h = 1$	2.3	2.3	2.3	4.7
$h = 4$	3.1	3.0	3.0	4.2
$h = 8$	0.4	0.4	0.4	2.6

	$g = \frac{1}{T}$	$g = \frac{1}{[\ln(T)]^3}$	$g = \frac{1}{K^2}$	Optimal g
Table 5b: Number of Factors in Model (average over all τ), PUNEW				
$E\left(\sum_{j=1}^{K} \gamma_j \| Data\right)$ for Bayesian Model Averaging (equal prior weights to all models)				
$h = 1$	48.7	52.8	4.5	59.3
$h = 4$	62.5	67.8	3.0	65.9
$h = 8$	65.9	69.1	3.0	63.9
$\sum_{j=1}^{K} \gamma_j$ for Selected Model for Bayesian Model Selection (equal prior weights to all models)				
$h = 1$	68.1	72.6	2.3	62.6
$h = 4$	64.8	68.9	0.4	67.5
$h = 8$	66.3	70.5	0.4	64.9
$E\left(\sum_{j=1}^{K} \gamma_j \| Data\right)$ for Bayesian Model Averaging (model prior using 3.10)				
$h = 1$	2.3	2.3	1.8	2.5
$h = 4$	1.1	1.1	1.0	1.2
$h = 8$	1.0	1.0	1.0	1.1
$\sum_{j=1}^{K} \gamma_j$ for Selected Model for Bayesian Model Selection (model prior using 3.10)				
$h = 1$	4.3	4.3	2.8	4.9
$h = 4$	1.7	1.7	1.2	2.4
$h = 8$	1.7	1.7	1.0	2.1
$E\left(\sum_{j=1}^{K} \gamma_j \| Data\right)$ for Bayesian Model Averaging (99.9% prior)				
$h = 1$	4.4	4.4	4.4	5.8
$h = 4$	1.7	1.7	1.6	3.5
$h = 8$	1.4	1.4	1.4	1.8
$\sum_{j=1}^{K} \gamma_j$ for Selected Model for Bayesian Model Selection (99.9% prior)				
$h = 1$	4.0	4.0	4.0	5.2
$h = 4$	0.1	0.04	0.1	1.7
$h = 8$	0.4	0.4	0.4	1.0

Overall, the forecasting results are less clearly favorable to the Bayesian model averaging and model selection methods that are the focus of this paper than were those obtained using the full sample. In general, none of the models which augment an AR(p) with factors exhibits an enormous improvement in forecasting over the AR(p) for either GDPQ or PUNEW. At medium to long forecast horizons (i.e. $h = 4$ or 8), there is virtually no evidence that a factor augmented model can beat an AR(2). As we shall discuss

in more detail below, this is likely due to the facts that lags of the dependent variable contain most of the information relevant for forecasting our variables and that substantial intertemporal variation in the regression coefficients occurs. Thus, at longer forecasting horizons, this instability in coefficients comes to dominate the small amount of information in the factors and the factor-augmented models cannot beat the AR(2). In the short run (i.e. $h = 1$), however, there is evidence that incorporation of the information in the factors allows for some moderate improvement in forecasting performance. For this reason, we will focus most of the following discussion on results for $h = 1$.

If we focus on short run forecasting and the 99.9% prior, Table 3a indicates that incorporating the information in the factors can reduce RMSEs for GDPQ by roughly 5% while Table 3b indicates roughly 10% RMSE reductions for PUNEW. Given the difficulties of macroeconomic forecasting, especially for quarterly data, and the fact that we have chosen two variables which, although important, are notably difficult to forecast, we consider such moderate reductions in RMSE to be quite important. Just as important for our purposes, we find that Bayesian model averaging is forecasting slightly better than either of the two model selection strategies. This can be seen in Tables 3a and 3b where BMA yields RMSEs which tend to be at least 1% lower than the model selection methods (a decrease which is substantive when one considers that all the information in the factors is only improving RMSEs by 5% or 10%). It can also be seen in Tables 4a and 4b where BMA predictive intervals (i.e. the predictive mean +/- two predictive standard deviations) exhibit slightly better coverage than predictive intervals based on a single model. The reason for this is that, by incorporating model uncertainty as well as parameter uncertainty, predictive standard deviations are slightly larger for BMA than with model selection methods.

It is also worth noting that forecasting results are fairly insensitive to the choice of g with the computationally demanding empirical Bayesian methodology performing roughly as well as simpler approaches. Hence, at least for $h = 1$ and the 99.9% prior, our findings are quite encouraging for Bayesian model averaging. It produces point forecasts which are more accurate than those produced using model selection methods and predictive intervals that have better coverage. However, all these findings hold for the prior over model space which is noninformative over the subset of the factors which contain 99.9% of the information in X. The other priors over model space yield worse forecasting performance. The prior given in (3.10) performs only somewhat worse than the 99.9% prior, however the forecasting performance of the

completely noninformative can be dire. It is to this issue of sensitivity to prior over model space that we now turn.

Tables 5a and 5b shed light on the poor forecasting performance which occurs when noninformative priors are used. Many factors associated with very small eigenvalues are included when the priors with $\theta_j = \frac{1}{2}$ and $g = \frac{1}{T}$ or $g = \frac{1}{[\ln(T)]^3}$ are used, and these have dire consequences for forecasting. Results in Tables 1a and 1b show that including many factors greatly improves measures of in-sample fit (such as the marginal likelihood). Before seeing the forecast results, we had thought this plausible. That is a few factors associated with small eigenvalues might be useful in explaining rare events (e.g. business cycle turning points). However, this good in-sample performance and plausible story do not translate into good forecasting performance. The only models which forecast better than an AR(2) are those which rule out many of the factors. Note that the factors associated with smaller eigenvalues can either be ruled out by directly attaching low prior weight to their entering the model (i.e. through the 99.9% prior or a prior such as 3.10), or by using a relatively flat prior for the regression coefficients (i.e. through choosing $g = \frac{1}{K^2}$, a value which implies a strong reward for parsimony) or by simply not including these factors (i.e. in the conventional approach where we simply include the first few factors and ignore the rest).

The use of empirical Bayesian methods which, for every forecast horizon, choose the prior which maximizes the marginal likelihood does not substantively improve forecast performance. This finding supports the story that successful in-sample performance (as measured by marginal likelihoods) does not map into successful out-of-sample forecasting performance.

Our results are consistent with those of Knox, Stock and Watson (2002) who consider forecasting using a large number of monthly time series using various estimation methods, including empirical Bayesian methods of a different sort from those used in our paper. These methods include many factors as explanatory variables and a data-based prior to carry out Bayesian inference. Knox, Stock and Watson (2002) find very good in-sample performance of their empirical Bayesian methods, but relatively poor forecasting performance in a simulated real-time forecasting exercise similar to that carried out in this paper. Despite the fact that they are using very different data from us, this pattern of good in-sample performance and bad out-of-sample performance holds in both of our exercises. Knox, Stock and Watson (2002) argue that parameter instability is probably the reason for this.

Another way of linking our findings to others is through the concept of shrinkage. There are many ways of shrinking forecasts (see, e.g., Giacomini, 2002, who finds so-called "Bayesian shrinkage estimators" to perform best in a forecasting exercise). In our framework, shrinkage can either occur through priors on the parameters (i.e. through g) or through priors on model space. In the latter case, omitting an explanatory variable is the ultimate way of shrinking its effect. An advantage of Bayesian model averaging is that it is much more flexible in the way it can handle this second shrinkage effect. That is, BMA can put weights on factors between zero and one (see equation 3.1) which effectively shrinks their effect. Hence, one way of looking at the poor results using the noninformative prior over model space, is that it simply does not shrink forecasts enough.

The one case where the noninformative prior does yield reasonable forecasts is when $g = \frac{1}{K^2}$ (i.e. the most noninformative value we consider). This case illustrates an important point regarding the role of g in shrinkage. In the context of a single model, decreasing g will make the prior for the coefficients less informative and, thus, decrease shrinkage of the forecasts. However, in a BMA exercise, decreasing g will also increase shrinkage due to the reward for parsimony associated with less informative priors. In Tables 5a and 5b, when $g = \frac{1}{K^2}$ it can be seen that very few factors are included (which shrinks forecasts relative to the case where more factors are included). However, the fact that g is so small means that the coefficients are not shrunk and will be very close to OLS estimates. Thus, decreasing g can either increase or decrease shrinkage. The general point we draw from this is that, unlike Bayesian estimation in the context of a single model, priors can matter a great deal and must be carefully selected. The specific point we draw is that it is best to choose the prior on model space (i.e. choose θ_j) to reflect prior information about the likely number of factors in the model and g to reflect prior information about the degree of shrinkage of regression coefficients (or use empirical Bayesian methods to estimate g). To attempt to use a single hyperparameter, g, to control both types of shrinkage is impossible.

Results with the model-space prior given in 3.10 are much better than with the completely flat prior, but worse than the 99.9% prior. This is due to the former prior requiring the first factor to always be included and downweighting some factors associated with very small eigenvalues. Since including the first factor does not always improve forecasts, while including some factors with very small eigenvalues sometimes does, the prior given in 3.10 does not do quite as well as the 99.9% prior.

The previous discussion has focussed on short forecasting horizons since parameter instability precludes accurate forecasting at longer horizons. However, it is worth briefly noting that, when we use an informative prior over model space, the average number of factors included in each forecasting model tends to be very small. For instance, with the 99% prior the models selected using our Bayesian model selection procedure tend to have, on average, much less than one factor included. In other words, at many or most of forecasting points, no factors at all are included. This is getting very close to simply providing a trivial forecast of zero at every point in time (remember, our data have been de-meaned and stationarity induced). We take this as additional evidence that past information in our data set is not very useful in forecasting, especially when $h = 4$ or 8. The most likely reason for this is parameter instability.

6   Conclusions and Directions for Future Research

In this paper, we have used Bayesian model averaging to address the problem of forecasting in dynamic factor models. We have provided both theoretical and empirical justifications for such an approach, as opposed to selecting a single model, are given. We have shown how BMA can be implemented in factor models using algorithms which simulate from the space defined by all possible models. Such simulation algorithms can be used either to do Bayesian model averaging or Bayesian model selection in an efficient manner. We applied these methods to the problem of forecasting GDP and inflation using quarterly U.S. data on 162 time series. For both GDP and inflation, we found that the models which contain factors do out-forecast an AR(p), but only by a relatively small amount and only at short horizons. These findings can be attributed to the presence of structural instability and the fact that lags of dependent variable seem to contain most of the information relevant for forecasting. Relative to the small forecasting gains provided by including factors, the gains provided by using Bayesian model averaging over forecasting methods based on a single model are appreciable.

Our findings suggest several avenues for future research. We have found strong evidence of structural instability and/or nonlinearity in the time series we have considered. One way that this problem can partially be addressed is by using rolling forecast windows. The methods of Bayesian model averaging and selection described in this paper can be used directly to handle this modification. Furthermore,

simple extensions of our framework could be done which would allow us to move towards addressing some of these data features (e.g. including dummy variables for structural breaks or including nonlinear transformations of the factors). However, it would be desirable to construct models which allow for the structural instabilities and nonlinearities which may exist. State space models provide a natural framework to deal with such issues. For instance, the class of latent factor regression models described in West (2002) is an attractive one when working with large macroeconomic panels. Indeed, the use of a latent factor representation (instead of using the actual factors as we have done), may help mitigate some of the problems we have found when doing Bayesian model averaging with the noninformative prior over model space (i.e. the latent factor representation removes some of the idiosyncratic variation in the factors and, hence, should reduce over-fitting problems). Extending the latent factor regression model to allow for structural instabilities and/or nonlinearities is relatively straightforward. However, in such a class of models, it would be difficult or impossible to carry out Bayesian model averaging over 2^K possible models since analytical expressions for predictive moments and marginal likelihoods do not exist. For this reason, in the present paper we have stayed within the framework of the Normal linear regression model with natural conjugate prior, despite its failure to properly model some key data features.

7   References

Bai, J. and Ng, S. (2002). "Determining the number of factors in approximate factor models," *Econometrica*, forthcoming.

Berger, J. and Pericchi, L. (1996). "The intrinsic Bayes factor for model selection and prediction," *Journal of the American Statistical Association*, 91, 109-122.

Bernanke, B., Boivin, J. and Eliasz, P. (2002). "Measuring the effects of monetary policy: A Factor-augmented vector autoregressive (FAVAR) approach," manuscript, Princeton University.

Boivin, J. and Ng, S. (2002). "Are more data always better for factor analysis?" manuscript, Columbia University.

Chipman, H., George, E. and McCulloch, R. (2001). "The practical implementation of Bayesian model selection," manuscript available at http://gsbwww.uchicago.edu/fac/robert.mcculloch/research/papers/index.html.

Clyde, M., Desimone, H. and Parmigiani, G. (1996). "Prediction via orthogonalized model mixing," *Journal of the American Statistical Association*, 91, 1197-1208.

Diebold, F. and Pauly, P. (1990). "The use of prior information in forecast combination," *International Journal of Forecasting*, 6, 503-508.

Draper, D. (1995). "Assessment and propagation of model uncertainty (with discussion)," *Journal of the Royal Statistical Society, Series B*, 56, 45-98.

Fernandez, C., Ley, E. and Steel, M. (2001a). "Benchmark priors for Bayesian model averaging," *Journal of Econometrics*, 100, 381-427.

Fernandez, C., Ley, E. and Steel, M. (2001b). Model uncertainty in cross-country growth regressions. *Journal of Applied Econometrics*, 16, 563-576.

Forni, M., Hallin, M., Lippi, M. and Reichlin, L. (2000). "The generalized dynamic factor model: Identification and estimation," *Review of Economics and Statistics*, 82, 540-554.

Foster, D. and George, E. (1994). "The risk inflation criterion for multiple regression," *The Annals of Statistics*, 22, 1947-1975.

George, E. and McCulloch, R. (1993). "Variable Selection via Gibbs Sampling," *Journal of the American Statistical Association*, 88, 881-889.

George, E. and McCulloch, R. (1997). "Approaches for Bayesian Variable Selection," *Statistica Sinica, 7, 339-373.*

Giacomini, R. (2002). "Tests of conditional predictive ability," University of California at San Diego, manuscript.

Giannone, D., Reichlin, L. and Sala, L. (2002). "Tracking Greenspan: Systematic and unsystematic monetary policy revisited," manuscript available at http://www.dynfactors.org/.

Hodges, J. (1987). "Uncertainty, policy analysis and statistics," *Statistical Science*, 2, 259-291.

Hoeting, J., Madigan, D., Raftery, A. and Volinsky, C. (1999). "Bayesian Model Averaging: A Tutorial," *Statistical Science*, 14, 382-417.

Jacobson, T. and Karlsson, S. (2003). "Finding Good Predictors for Inflation: A Bayesian Model Averaging Approach', *Journal of Forecasting*, forthcoming.

Kass, R. and Wasserman, L. (1995). "A reference Bayesian test for nested hypotheses and its relation-ship to the Schwarz criterion," *Journal of the American Statistical Association*, 90, 928-934.

Knox, T., Stock, J. and Watson, M. (2002). "Empirical Bayes forecasts of one time series using many predictors," manuscript available at http://www.wws.princeton.edu/~mwatson/wp.html.

Koop, G. and Potter, S. (2001). "Are apparent findings of nonlinearity due to structural instability in economic time series?" *The Econometrics Journal*, 4, 37-55.

Koop, G. and Potter, S. (1999). "Dynamic asymmetries in U.S. unemployment," *Journal of Business and Economic Statistics*, 17, 298-312.

Min, C. and Zellner, A. (1993). "Bayesian and Non-Bayesian methods for combining models and forecasts with applications to forecasting international growth rates," *Journal of Econometrics*, 56, 89-118.

Poirier, D. (1995). *Intermediate Statistics and Econometrics: A Comparative Approach*. Cambridge: The MIT Press.

Raftery, A., Madigan, D. and Hoeting, J. (1997). "Bayesian model averaging for linear regression models," *Journal of the American Statistical Association*, 92, 179-191.

Stock, J. and Watson, M. (2002a). "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, 20, 147-162.

Stock, J. and Watson, M. (2002b). "Has the Business Cycle Changed and Why?" *National Bureau of Economic Research Working Paper* 9127.

West, M. (2002). "Bayesian factor regression models in the "Large p, Small n" paradigm," to appear in Bayesian Statistics, 7.

Zellner, A. (1986). "On assessing prior distributions and Bayesian regression analysis with g-prior distributions," in Goel, P. K. and Zellner, A. (eds.), *Bayesian Inference and Decision Techniques: Essays in Honour of Bruno de Finetti*. Amsterdam: North-Holland.

Technical Appendix

The predictive distribution of $y_{\tau+1}$ given all of the potential explanatory variables and information through time τ is given by the standard formula (see, e.g., Poirier, 1995, page 556):

$$y_{\tau+1}|Z_{\tau+1}, \gamma \sim t\left(z_{\tau+1}\overline{\alpha}_\tau, \overline{s}_\tau^2 \left[1 + z_{\tau+1}\overline{A}_\tau z'_{\tau+1}\right], \tau + \underline{\nu}\right)$$

where Z_τ is the matrix containing the first τ rows of Z (remember that Z only contains lagged dependent and explanatory variables) and z_τ is the $(\tau+1)^{th}$ row of Z. The posterior mean and scale matrix (based on data through time τ) are

$$\overline{A}_\tau = \left(\underline{A}^{-1} + Z'_\tau Z_\tau\right)^{-1}$$

and

$$\overline{\alpha}_\tau = \overline{A}_\tau \left(\underline{A}^{-1}\underline{\alpha} + Z'_\tau Y_\tau\right),$$

where $Y_\tau = (y_1, .., y_\tau)'$. Note that our orthogonalization strategy implies $\underline{A}$ is a diagonal matrix and, thus, $\overline{A}_\tau$ is a diagonal matrix of a simple form. This, and the fact we have assumed $\underline{\alpha} = 0$, simplifies computation.

The remaining feature to be evaluated is $\overline{s}_\tau^2$ which is given by:

$$\overline{s}_\tau^2 = (\tau + \underline{\nu})\left[\underline{\nu s}^2 + (y - Z_\tau\overline{\alpha}_\tau)'(y - Z_\tau\overline{\alpha}_\tau) + (\overline{\alpha}_\tau - \underline{\alpha}_\tau)' \underline{A}_\tau^{-1}(\overline{\alpha}_\tau - \underline{\alpha}_\tau)\right].$$

These formulae hold for the full model with $\gamma = \iota_K$. Results for other values of γ are obtained by deleting columns/rows/elements of all vectors/matrices as appropriate.

Data Appendix

This Appendix provides a list of all variables used in the analysis. All data are quarterly from 1959Q1-2001Q1. With a few minor exceptions, we use the same variables and transformations as Stock and Watson (2002b). The original variables were taken from the DRI Basic Economics database and we use the DRI acronyms for these. Some transformations of these original variables are taken as noted below. All variables are transformed to stationarity in the same manner as in Stock and Watson (2002b). The exact transformation required is noted below using the code:

1 = level of the series

2 = first difference

3 = second difference

4 = log of the series

5 = first difference of the log

6 = second difference of the log

Series Name	Transform. Code	Description
		NIPA Components
GDPQ	5	Gross domestic product (chained)
GOQ	5	Gross domestic product, goods
GOSQ	5	Final sales of goods
GODQ	5	Gross domestic product, durable goods
GODSQ	5	Final sales of durables
GONQF	5	Gross domestic product, nondurables
GONSQF	5	Final sales of nondurables
GOOSQ	5	Gross domestic product, services
GOCQ	5	Gross domestic product, structures
GCQ	5	Personal consumption expenditures (chained), total
GCDQ	5	Personal consumption expenditures (chained), durables
GCNQ	5	Personal consumption expenditures (chained), nondurables
GCSQ	5	Personal consumption expenditures (chained), services
GPIQ	5	Investment (chained), total
GIFQ	5	Fixed investment (chained), total
GINQ	5	Fixed investment (chained), nonresidential
GIRQ	5	Fixed investment (chained), residential
GEXQ	5	Exports of goods and services (chained)
GIMQ	5	Imports of goods and services (chained)
GGEQ	5	Govt. consumption expenditures and gross investment (chained)
GGFENQ	5	Nat. defence cons. expenditures and gross investent (chained)
GMCANQ	5	Personal cons. exp. (chained), new cars (bil 96$, saar)
GMCDQ	5	Personal cons. exp. (chained), total durables (bil 96$, saar)
GMCNQ	5	Personal cons. exp. (chained), nondurables (bil 92$, saar)
GMCQ	5	Personal cons. exp. (chained), total (bil 92$, saar)
GMCSQ	5	Personal cons. exp. (chained), services (bil 92$, saar)
GMPYQ	5	Personal income (chained) (series #52) (bil 92$, saar)
GMYXPQ	5	Personal income less transfer payments (chained) (#51) (bil 92$, saar)

Series Name	Transform. Code	Description
		Money, Credit, Interest Rates and Stock Prices
FBCUC	2	Change in Business and Consumer Credit Outstanding
FM1	6	Money stock, M1 (bil$, sa)
FM2	6	Money stock, M2 (bil$, sa)
FM2DQ	5	Money supply - M2 in 92$ (bci)
FM3	6	Money stock, M3 (bil$, sa)
FMFBA	6	Monetary base adusted for reserve requirement changes (mil$, sa)
FMRRA	6	Depository inst. reserves: adjusted for res. req. changes (mil$, sa)
FSDXP	5	S&P's composite stock index, dividend yield (% per annum)
FSNCOM	5	NYSE common stock price index, composite (12/31/65=50)
FSPCAP	5	S&Ps common stock price index, capital goods (1941-43=10)
FSPCOM	5	S&Ps common stock price index, composite (1941-43=10)
FSPIN	5	S&Ps common stock price index, industrials (1941-43=10)
FSPXE	5	S&Ps composite common stock, price-earning ratio (%, nsa)
FYAAAC	2	Bond yield, Moody's aaa corporate (% per annum)
FYBAAC	2	Bond yield, Moody's baa corporate (% per annum)
FYFF	2	Interest rate, federal funds (effective) (% per annum, nsa)
FYGM3	2	Interest rate, US t-bill, sec. market 3 mo. (% per annum, nsa)
FYGT1	2	Interest rate, US treasury const. maturities, 1-yr. (% per ann., nsa)
FYGT10	2	Interest rate, US treasury const. maturities, 10-yr. (% per ann., nsa)

Series Name	Transform. Code	Description
		Housing
HSBR	5	Housing authorized, total new priv housing units (thous., saar)
HSFR	5	Housing starts, total farm and non-farm (thous., sa)
HSMW	5	Housing starts, midwest (thous., sa)
HSNE	5	Housing starts, northeast (thous., sa)
HSSOU	5	Housing starts, south (thous., sa)
HSWST	5	Housing starts, west (thous., sa)
		Industrial Production
IP	5	Industrial production, total index (1992=100, sa)
IPC	5	Industrial production, consumer goods (1992=100, sa)
IPCD	5	Industrial production, durable consumer goods (1992=100, sa)
IPCN	5	Industrial production, nondurable consumer goods (1992=100, sa)
IPE	5	Industrial production, business equipment (1992=100, sa)
IPF	5	Industrial production, final products (1992=100, sa)
IPI	5	Industrial production, intermediate products (1992=100, sa)
IPM	5	Industrial production, materials (1992=100, sa)
IPMD	5	Industrial production, durable goods materials (1992=100, sa)
IPMFG	5	Industrial production, manufacturing (1992=100, sa)
IPMIN	5	Industrial production, mining (1992=100, sa)
IPMND	5	Industrial production, nondurable goods materials (1992=100, sa)
IPN	5	Industrial production, nondurable manufacturing (1992=100, sa)
IPP	5	Industrial production, products (1992=100, sa)
IPUT	5	Industrial production, utilities (1992=100, sa)
IPXMCA	1	Capacity utilization rate (%), manufacturing (sa, from FRB)

Series Name	Transform. Code	Description
		Inventories, Orders and Sales
IVMFDQ	5	Inventories, business durables (chained, 96$ mil, sa)
IVMFGQ	5	Inventories, business, manufacturing (chained, 96$ mil, sa)
IVMFNQ	5	Inventories, business, nondurables (chained, 96$ mil, sa)
IVMTQ	5	Mfg. and trade inventories, total (chained, 96$ mil, sa)
IVRRQ	5	Mfg. and trade inventories, retail trade (chained, 96$ mil, sa)
IVWRQ	5	Mfg. and trade inventories, merchant wholesalers (chained, 96$ mil, sa)
IVSRMQ	5	Ratio for mfg. and trade:mfg; inventory/sales (96$ sa)
IVSRQ	5	Ratio for mfg. and trade; inventory/sales (chained, 96$ sa)
IVSRRQ	5	Ratio for mfg. and trade:retail trade; inventory/sales (96$ sa)
IVSRWQ	5	Ratio for mfg. and trade:wholesaler; inventory/sales (96$ sa)
GVSQ	1	(change in inventories)/sales — goods
GVDSQ	1	(change in inventories)/sales — durable goods
MDOQ	5	New orders, durable good industries, 92$
MOCMQ	5	New orders (net), consumer goods and materials, 92
MPCONQ	5	Contracts & orders for plant and equipment, 92$
MSDQ	5	Mfg. & trade: mfg., durable goods (mil of chained 96$)
MSMTQ	5	Mfg. and trade: total (mil. of chained 96$, sa)
MSNQ	5	Mfg. and trade: mfg; nondurable goods (mil. of chained 96$, sa)
MSONDQ	5	New orders, nondefence capital goods (92$)
RTNQ	5	Retail trade, nondurable goods (mil. of 96$, sa)
WTDQ	5	Merch. wholesalers, durable goods total (mil of 96$, sa)
WTNQ	5	Merch. wholesalers, nondurable goods (mil of 96$, sa)
WTQ	5	Merch. wholesalers, total (mil of 96$, sa)

Series Name	Transform. Code	Description
		Employment
LHEL	5	Index of help wanted adv. in newspapers (1967=100, sa)
LHELX	5	Employment ratio: help wanted ads/no. unemployed clf
LHEM	5	Civilian labor force, employed, total (thous., sa)
LHNAG	5	Civilian labor force, employed, nonagric. industries (thous., sa)
LHU14	5	Unempl. by duration, persons unemp. 5-14 wks (thous., sa)
LHU15	5	Unempl. by duration, persons unemp. 15+ wks (thous., sa)
LHU26	5	Unempl. by duration, persons unemp. 15-26 wks (thous., sa)
LHU5	5	Unempl. by duration, persons unemp. <5 wks (thous., sa)
LHU680	5	Unempl. by duration, average duration in weeks (thous., sa)
LHUR	2	Unemployment rate, workers 16 yrs. and over (%, sa)
LP	5	Employees on nonagric. payrolls, total private (thous., sa)
LPCC	5	Employees on nonagric. payrolls, contract construction (thous., sa)
LPED	5	Employees on nonagric. payrolls, durable goods (thous., sa)
LPEM	5	Employees on nonagric. payrolls, manufacturing (thous., sa)
LPEN	5	Employees on nonagric. payrolls, nondurable goods (thous., sa)
LPFR	5	Employees on nonagric. payrolls, finance, ins. and real est. (thous., sa)
LPGD	5	Employees on nonagric. payrolls, goods-producing (thous., sa)
LPGOV	5	Employees on nonagric. payrolls, government (thous., sa)
LPHRM	5	Avg. wkly. hours of production workers, manufacturing (sa)
LPMOSA	5	Avg. wkly. hours of production workers, mfg, overtime (sa)
LPNAG	5	Employees on nonagric. payrolls, total (thous., sa)
LPS	5	Employees on nonagric. payrolls, services (thous., sa)
LPSP	5	Employees on nonagric. payrolls, service-producing (thous., sa)
LPT	5	Employees on nonagric. payrolls, wholesale & retail trade (thous., sa)

Series Name	Transform. Code	Description
		NAPM Indexes
PMCP	1	NAPM commodity prices index (%)
PMDEL	1	NAPM vendor deliveries index (%)
PMEMP	1	NAPM employment index (%)
PMI	1	Purchasing managers' index (%)
PMNO	1	NAPM new orders index (%)
PMNV	1	NAPM inventories index (%)
PMP	1	NAPM production index (%)
		Industrial Production in Other Countries
IPCAN	5	Industrial production, Canada (1990=100, sa)
IPFR	5	Industrial production, France (1990=100, sa)
IPIT	5	Industrial production, Italy (1990=100, sa)
IPJP	5	Industrial production, Japan (1990=100, sa)
IPUK	5	Industrial production, UK (1990=100, sa)
IPWG	5	Industrial production, West Germany (1990=100, sa)

Series Name	Transform. Code	Description
		Wages and Prices
R_LEHCC	2	ln(LEHCC/GDPD)
LEHCC	6	Avg. hourly earnings of workers, construction ($, sa)
R_LEHM	2	ln(LEHM/GDPD)
LEHM	6	Avg. hourly earnings of prod. workers, manufacturing ($, sa)
GDPD	6	GDP implicit price deflator (1992=100)
GDC	6	Implicit price deflator, personal consumption expenditure
PUNEW	6	CPI-u, all items (82-84=100, sa)
PUXF	6	CPI-u, all items less food (82-84=100, sa)
PUXHS	6	CPI-u, all items less shelter (82-84=100, sa)
PUXM	6	CPI-u, all items less medical care (82-84=100, sa)
PW	6	Producer price index, all commodities (82=100, nsa)
PSCCOM	6	Spot market price index, bls & crb, all commodities (67=100, nsa)
R_PSCCOM	2	ln(PSSCOM/GDPD)
PSM99Q	6	Index of sensitive materials prices (1990=100)
R_PSM99Q	2	ln(PSM99Q/GDPD)
PU83	6	CPI-u, apparel and upkeep (82-84=100, sa)
R_PU83	2	ln(PU83/GDPD)
PU84	6	CPI-u, transportation (82-84=100, sa)
R_PU84	2	ln(PU84/GDPD)
PU85	6	CPI-u, medical care (82-84=100, sa)
R_PU85	2	ln(PU85/GDPD)
PUC	6	CPI-u, commodities (82-84=100, sa)
R_PUC	2	ln(PUC/GDPD)
PUCD	6	CPI-u, durables (82-84=100, sa)
R_PUCD	2	ln(PUCD/GDPD)
PUS	6	CPI-u, services (82-84=100, sa)
R_PUS	2	ln(PUS/GDPD)
PW561	6	Producer price index, crude petroleum (82=100, nsa)
R_PW561	2	ln(PW561/GDPD)
PWFCSA	6	Producer price index, finished consumer goods, (82=100, sa)
R_PWFCSA	2	ln(PWFCSA/GDPD)
PWFSA	6	Producer price index, finished goods (82=100, sa)
R_PWFSA	2	ln(PWFSA/GDPD)