Journal of Forecasting
J. Forecast. **20**, 297–314 (2001)
DOI: 10.1002/for.800

Forecasting Output Growth Rates and Median Output Growth Rates: A Hierarchical Bayesian Approach

JUSTIN L. TOBIAS*
University of California-Irvine, USA

ABSTRACT
This paper describes procedures for forecasting countries' output growth rates and medians of a set of output growth rates using Hierarchical Bayesian (HB) models. The purpose of this paper is to show how the γ-shrinkage forecast of Zellner and Hong (1989) emerges from a hierarchical Bayesian model and to describe how the Gibbs sampler can be used to fit this model to yield possibly improved output growth rate and median output growth rate forecasts. The procedures described in this paper offer two primary methodological contributions to previous work on this topic: (1) the weights associated with widely-used shrinkage forecasts are determined endogenously, and (2) the posterior predictive density of the future median output growth rate is obtained numerically from which optimal point and interval forecasts are calculated. Using IMF data, we find that the HB median output growth rate forecasts outperform forecasts obtained from variety of benchmark models.

INTRODUCTION

In this paper, I describe how hierarchical Bayesian models can be used to forecast individual countries' output (GDP) growth rates, and medians of a set of countries' output growth rates. The problem of forecasting output growth rates is of central importance, and has received considerable attention in the literature (see, for example, Garcia-Ferrer *et al*., 1987; Zellner and Hong, 1989; Zellner *et al*., 1990, 1991; Min and Zellner, 1993; Zellner and Tobias, 1999; Kazimi and Brownstone, 1999). With the exception of Kazimi and Brownstone (1999), the studies above describe and employ Bayesian forecasting procedures which appear to yield lower predictive Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE) than a variety of benchmark forecasting models. The purpose of this paper is to describe how hierarchical Bayesian models and Markov Chain Monte Carlo (MCMC) techniques (namely the Gibbs sampler) can be employed to generalize the Bayesian methodology used in these earlier studies and yield possibly improved forecasting results. Although the approach presented in this paper is applied to forecasting output growth rates, the procedures described in the following two sections are quite operational and can be applied to a variety of forecasting problems.

* Correspondence to: Justin L. Tobias, Department of Economics, University of California-Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA. E-mail: jtobias@uci.edu

Received April 2000
Accepted September 2000

Previous work on the topic of forecasting international output growth rates has noted that 'shrinkage forecasts' can perform quite well in practice. Zellner and Hong (1989) introduce one such shrinkage forecast (denoted a 'γ forecast'), in which point estimates of the regression coefficients of each country are obtained as a weighted average of the least-squares estimate for that country and a pooled least squares estimate in which coefficients are restricted to be equal across all countries.[1] This estimator formalized the notion that there may be some underlying common relationship between inputs and output growth rates across countries, while countries may also exhibit idiosyncratic differences in relating these inputs to output growth rates. Given a weighting scheme, least-squares estimates for all countries and a pooled least-squares estimate, one-step-ahead output growth rate forecasts can be obtained and RMSE and MAE of these forecasts can be determined from a hold-out sample. Zellner and Hong implemented their novel approach by specifying a variety of different values for the weight (which was denoted γ and hence the name γ-forecast) and computing the associated predictive RMSE and MAE at a set of prespecified weights. Kazimi and Brownstone (1999) determined the weights endogenously in a non-Bayesian approach using Stein's shrinkage estimator. Under certain regularity conditions, Stein's estimator is known to dominate least-squares relative to quadratic loss, and Kazimi and Brownstone documented some reduction in predictive RMSE and MAE when using Stein's estimator.

In this paper, I build upon the results in Zellner and Hong (1989) and incorporate the endogenous determination of the weights in the spirit of Kazimi and Brownstone (1999) in a finite-sample Bayesian approach. In the sections which follow, I describe how the shrinkage forecasts of these earlier studies emerge from a hierarchical Bayesian model. To express the belief that the regression coefficients for each country may be similar, I assume that these parameters are drawn from the same underlying distribution with common mean and covariance matrix. To effect a Bayesian analysis, priors are then placed on the common mean and covariance matrix as well as the remaining parameters in the model. Since the priors are chosen to be conjugate at each stage in the hierarchy, the complete conditionals are easily obtained and thus the Gibbs sampler can be easily applied to fit the model. What emerges from this hierarchical framework is that the conditional posterior mean of the regression coefficients for each country is in the form of a shrinkage estimator, where, as in Zellner and Hong (1989) and Kazimi and Brownstone (1999), the conditional mean is a weighted average of the least-squares estimator for that country and a pooled estimator. Through the use of hierarchical modelling, the weighting mechanism is determined endogenously, since the posterior distribution of the weighting matrix is simultaneously determined along with the other parameters of interest. This is a generalization of the results in Zellner and Hong (1989) where the weights were assigned and forecasts were obtained for a prespecified set of weights.

The simulation-based techniques discussed in this paper also enable us to obtain forecasts of the median of a set of countries' output growth rates, as in Zellner and Hong (1989) and Zellner and Tobias (1999).[3] In this paper, we show how the Gibbs sampler can be used to approximate the posterior predictive distribution of the future median output growth rate. In the studies mentioned

[1] Zellner and Hong (1989) also describe other shrinkage forecasts, including the η-forecast which is obtained from a loss function approach. This paper focuses attention on the γ-forecast and acknowledges that Zellner and Hong (1989) introduced several other shrinkage forecasts.

[2] Zellner and Hong (1989) acknowledged that the posterior distribution of the weights could be obtained, but deferred this topic as the subject of future research.

[3] Note that the median output growth rate may be a measure of the 'world' output growth rate that is robust to outliers. The existence of outliers may be particularly problematic if the mean is used as our criterion and the number of countries in the sample is relatively small.

above, point forecasts of the future median output growth rate were obtained by taking the median of the set of point forecasts for each country. This approach avoids calculation of the posterior predictive density of the median and thus could produce forecasts which perform poorly in practice. Using output from the Gibbs sampler, it is possible to simulate draws from the posterior predictive density of the future median output growth rate, and use these simulated values to obtain optimal point and interval forecasts, as done below. Results indicate that predictive RMSE and MAE of the median forecasts obtained from the hierarchical Bayesian model are the lowest among a variety of benchmark forecasting models.

The outline of this paper is as follows. In the next section I describe the model which has been used in the studies mentioned above and serves as the primary model employed throughout this paper. In the third section I show how the Gibbs sampler can be applied to analyse the model of the second section and derive the necessary conditional densities. A discussion of the link between the posterior mean of these conditionals and the 'shrinkage estimators' of these earlier studies is also presented. The fourth section describes the data and presents the forecasting results. The final section presents conclusions.

THE MODEL

As in Garcia-Ferrer *et al.* (1987), Zellner and Hong (1989), Zellner *et al.*, (1990, 1991), Min and Zellner (1993), Zellner and Tobias (1999), and Kazimi and Brownstone (1999), I specify the following baseline regression model[4] for the growth rate of real GDP for country i in year t (y_{it}):

$$y_{it} = x_{it}\beta_i + u_{it} \qquad i = 1, 2, \ldots M, t = 1, 2, \ldots T_i \tag{1}$$

where the vector x_{it} is specified as follows:

$$x_{it} = [1 \ \ y_{it-1} \ \ y_{it-2} \ \ y_{it-3} \ \ SR_{it-1} \ \ SR_{it-2} \ \ GM_{it-1} \ \ WSR_{t-1}]$$

It is important to note that x_{it} consists only of lagged values, and thus our one-step-ahead forecasts for time $T+1$ will depend only on covariates which are known at time T. In the above, SR_{it} and GM_{it} denote the growth rate of real stock prices and real money, respectively, for country i in year t, and WSR_t is a 'world return' variable, which is the median growth rate of real stock prices in year t. The inclusion of this last variable has been shown in previous work to reduce the contemporaneous correlation among the error terms. A third-order lag in the real annual growth rate is also included, which allows the process to have one real root associated with a trend and two complex conjugate roots associated with a cycle, as found empirically in Hong (1989). Theoretical justifications for the use of equation (1) in forecasting have appeared in Hong (1989), Min (1992) and Zellner (1999).

As in previous work on this topic, I assume the following normal sampling density for the data:

$$y_{it}|x_{it}, \beta_i, \sigma_i^2 \stackrel{ind}{\sim} N(x_{it}\beta_i, \sigma_i^2) \tag{2}$$

[4] Zellner *et al.* (1991) and Min and Zellner (1993) also analyse time-varying parameter models. In this paper, I focus on the baseline regression in equation (1) as in Zellner and Hong (1989), Kazimi and Brownstone (1999), and Zellner and Tobias (1999), and show how the Gibbs sampler can be employed to yield possibly improved forecasting results.

To express the belief that the β_i are similar, I assume that the regression coefficient vectors for each country are drawn from a common normal distribution.[5]

$$\beta_i \overset{iid}{\sim} N(\beta, \Lambda) \tag{3}$$

Third-stage priors for the common mean β and inverse covariance matrix Λ^{-1} as well as second-stage priors for the variance coefficients $\{\sigma_i^2\}_{i=1}^M$ are then specified to complete the hierarchical model. I assume that these parameters are *a priori* independently distributed, and specify the following conjugate forms for these priors:

$$p(\beta|\beta_0, \Sigma) \sim N(\beta_0, \Sigma) \tag{4}$$

$$p(\sigma_i^2|a_i, b_i) \sim IG(a_i, b_i) \qquad i = 1, 2, \ldots M \tag{5}$$

$$p(\Lambda^{-1}|\rho, R) \sim W((\rho R)^{-1}, \rho) \tag{6}$$

In the above, *IG*(*a, b*) denotes an inverted gamma density with parameters a and b and mean $1/[b(a-1)]$, $a > 1$, $b > 0$. The prior density for the inverse covariance matrix Λ^{-1} follows a Wishart (denoted *W*) density with 'degrees of freedom' parameter ρ and is parameterized so that R approximately represents the prior variability of the β_i about the common mean β. A diffuse prior for Λ^{-1} is obtained as $\rho \to 0$. The quantities $\{a_i\}_{i=1}^M$, $\{b_i\}_{i=1}^M$, ρ, R, β_0 and Σ are known and specified by the researcher.

Prior specification

For the analysis which follows, we choose the prior values above to be 'vague' or 'diffuse' and centre these priors over previous results obtained from Zellner *et al.* (1990) (henceforth ZHG). These authors analyzed the model in equation (1) for a larger set of countries, and reported key estimation results for the period 1954–1973 in Appendix Table 1 of their paper. We choose the prior mean of the pooled parameter vector β (denoted as β_0 in equation (1)) as the sample mean of the $E(\beta_i|\text{Data})$ values reported in Table 1 of ZHG. This leads us to choose

$$\beta_0 = [4.58\ \ -0.01\ \ -0.07\ \ -0.03\ \ -0.002\ \ -0.02\ 0.17\ 0.07]'$$

To express our prior uncertainty about the value of this common mean, we choose Σ to be a diagonal matrix, and select the diagonal elements to be quite large. Specifically, we set these diagonal elements equal to 25 so that the prior standard deviation of each element of β is 5. This produces priors which are quite flat over each element, and are clearly less informative than the marginal posteriors of the elements of β (see Figure 1A and 1D of in the fourth section).

The variance parameters are also taken from Appendix Table 1 of ZHG. For the 16 countries examined here, ZHG report $E(\sigma_i|\text{Data})$ as:

$$E(\sigma_i|\text{Data}) = [1.66\ 2.11\ 1.68\ 1.61\ 2.25\ 2.71\ .98\ 2.57\ 3.61\ 1.67\ 1.91\ 3.89\ 1.39\ 1.18\ 1.71\ 1.52]$$

for Australia, Austria, Belgium, Canada, Denmark, Finland, France, Ireland, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, the UK, and the USA, respectively. We set $a_i = 3\forall i$, and

[5] Note that this is similar to the 'random coefficients' model often encountered in non-Bayesian econometrics. Here, I pursue a Bayesian approach and will obtain the finite-sample posterior distribution of the country-level coefficients (β_i), variance parameters (σ_i^2), common mean (β) and common inverse covariance matrix (Λ^{-1}).

choose $b_i = [10\mathrm{E}^2(\sigma_i|\text{Data})]^{-1}$. This results in prior means and standard deviations for the variance parameters that are five times a large as the square of the results in ZHG. Again this produces priors for the variance parameters which are spread over a large region of the parameter space and are clearly less informative than the corresponding marginal posteriors [see, e.g. Figure 1(C)].

Finally, for the matrix Λ^{-1}, we set $\rho = 9$ to ensure propriety of the Wishart prior (yet keep this prior as uninformative as possible). Let $\tilde{R}$ denote the 8×8 diagonal matrix with diagonal elements equal to the sample variance of the $E(\beta_i|\text{Data})$ values reported in ZHG. Since R approximately represents the prior variability of the β_i about the common mean β, setting $R = \tilde{R}$ would seem a reasonable choice. Based on Appendix Table 1 of ZHG, we find:

$$\tilde{R} = \text{Diag}[1.2\ 0.078\ 0.073\ 0.063\ 0.004\ 0.002\ 0.023\ 0.006]$$

In Zellner and Hong (1989), Zellner *et al.* (1990), LeSage (1990), Zellner and Tobias (1999), and Baltagi *et al.* (2000), it was found that models with coefficient pooling across groups often produced improved forecasts.[6] To express this partiality for pooled models through our prior, we set $R = (1/3)\tilde{R}$, so that our prior view of coefficient variability across countries is one-third of that suggested by Appendix Table 1 of ZHG.

In general, we find that the results which follow are robust to reasonable changes in the prior, but are most sensitive to the specification of R in the Wishart prior.[7] Choosing the elements of R to be 'large', for example, expresses a prior belief for a large amount of variability in the individual countries' coefficient vectors, and as a result, our point forecasts tended to approach the country-by-country least-squares forecasts. The choice of R above seems reasonable since it is guided by previous estimation results, expresses partiality for pooled models as found in previous work and, as shown below, produces forecasts which perform well in practice.

The posterior

The data employed in the analysis of the fourth section consist of observations on $M = 16$ industrialized countries. By Bayes' Theorem, the joint posterior of the 188 parameters used in this analysis (128 country-specific parameters $(\{\beta_i\}_{i=1}^{16})$, 16 variance parameters $(\{\sigma_i^2\}_{i=1}^{16})$, 8 elements of the common mean β, and 36 parameters in the inverse covariance matrix Λ^{-1}) is given up to a normalizing constant as the product of the likelihood implied from equation (2) times equations (3)–(6). Formally, letting $\theta = [\{\beta_i\}, \beta, \{\sigma_i^2\}, \Lambda^{-1}]$ denote all the parameters in the model, we have:[8]

$$p(\theta|\text{Data}) \propto \Pi_{i=1}^{16}[p_N(y_i; X_i\beta_i, \sigma_i^2 I_{T_i})p_N(\beta_i; \beta, \Lambda)p_{IG}(\sigma_i^2)]p(\beta)p(\Lambda^{-1})$$

In the above, $p_N(x; \mu_x, \Sigma_x)$ denotes a multivariate normal density for x with mean μ_x and covariance matrix Σ_x, and y_i and X_i represent the $T_i \times 1$ output growth rate vector and $T_i \times 8$ design matrix for country i, respectively.

---

[6] Unlike the other studies which focus on output growth rate or turning point forecasts, Baltagi *et al.* (2000) reported such a result when forecasting cigarette demand.

[7] The choice of R also had an effect on the mixing properties of the chain. Specifically, the lagged autocorrelations tended to rise as the values assigned to the diagonal elements of R decreased. Since ρ is reasonably large relative to M, it is not surprising to note that the Wishart prior specification does have an effect on the fitted parameter values and the forecasting results.

[8] We treat the initial conditions as given and include them into the given data to facilitate computation with the Gibbs sampler.

Although inference in this model seems like a daunting computational task, the conjugacy of the specifications in equations (3)–(6) at each stage in the hierarchy yields simple (and easily sampled), complete conditional posterior densities for all parameters of interest. This makes implementation of the Gibbs sampler straightforward. In the following section, I show how the Gibbs sampler can be applied to fit this model to obtain point and interval estimates, marginal posterior densities, and output growth rate and median output growth rate forecasts.

IMPLEMENTATION OF THE GIBBS SAMPLER

The idea behind the Gibbs sampler and Markov Chain Monte Carlo (MCMC) techniques in general is to generate a sequence of draws which, after a suitable pre-convergence or 'burn-in' period, have converged in distribution to the joint posterior density. The post-convergence draws (draws kept after the burn-in) can then be used to estimate posterior means, interquartile ranges, smoothed posterior densities, or other quantities of interest. The Gibbs sampler exploits the powerful result of Hammersley and Clifford (1970, 'Markov fields on finite graphs and lattices,' unpublished manuscript) and Besag (1974) which states that (under weak conditions) the complete conditional distributions are sufficient to define the joint distribution.

Given this result, a natural approach for obtaining draws from the joint posterior is to cycle through samples from the complete conditionals. The Markov chain produced by successively sampling from these conditionals is stationary (once we 'arrive' at the posterior, the next draw will also be distributed according to the posterior), and under mild conditions, the posterior is also the limiting distribution of the chain. This result is extremely powerful and has made calculations possible in seemingly intractable large-dimension problems.[9] See Metropolis *et al.* (1953), Hastings (1970), Gelfand and Smith (1990), Gelfand *et al.* (1990), Casella and George (1992), and Tierney (1994) for more on the theory and application of the Gibbs sampler and Markov Chain Monte Carlo (MCMC) techniques.

Given the joint posterior derived in the previous section it is a straightforward exercise to obtain the complete conditional posterior densities of our regression parameters. For each conditional density, we can regard the remaining parameters in the conditioning set as fixed, thus simplifying our analysis considerably. For example, when deriving the complete conditional posterior of the common mean β we only need to consider the contribution of the second-stage priors: $p_N(\beta_i; \beta, \Lambda)$ and the third-stage prior: $p_N(\beta; \beta_0, \Sigma)$ since the remaining parameters are treated as fixed and are essentially absorbed into the normalizing constant of this conditional. Proceeding in this fashion for all the regression parameters, we repeatedly use the results in Lindley and Smith (1972) to obtain:[10]

$$[\beta_i | \{\sigma_i^2\}, \beta, \Lambda^{-1}, \text{Data}] \sim N(D_i d_i, D_i) \qquad i = 1, 2, \ldots M \tag{7}$$

where

$$D_i \equiv (X_i' X_i / \sigma_i^2 + \Lambda^{-1})^{-1} \qquad d_i \equiv (X_i' y_i / \sigma_i^2 + \Lambda^{-1}\beta)$$

$$[\beta | \{\beta_i\}, \Lambda^{-1}, \{\sigma_i^2\}, \text{Data}] \sim N(V(M\Lambda^{-1}\overline{\beta} + \Sigma^{-1}\beta_0), V) \tag{8}$$

[9] This is particularly true for hierarchical models with conjugate priors, making the Gibbs sampler extremely well-suited for this analysis. See also Gelfand and Smith (1990) and Gelfand *et al.* (1990) for a description of how the Gibbs sampler can be applied to analyse a variety of econometric models.

[10] See Carlin and Louis (1996) and Gelfand *et al.* (1990) for analysis of highly similar models.

where

$$V \equiv (M\Lambda^{-1} + \Sigma^{-1})^{-1} \qquad \overline{\beta} \equiv \frac{1}{M}\sum_{i=1}^{M}\beta_i$$

$$[\sigma_i^2|\beta, \{\beta_i\}, \Lambda^{-1}, \text{Data}] \sim IG(T_i/2 + a_i, [1/2(y_i - X_i\beta_i)'(y_i - X_i\beta_i) + b_i^{-1}]^{-1}) i = 1, 2, \ldots M \quad (9)$$

$$[\Lambda^{-1}|\{\sigma_i^2\}, \{\beta_i\}, \beta, \text{Data}] \sim W\left(\left[\sum_{i=1}^{M}(\beta_i - \beta)(\beta_i - \beta)' + \rho R\right]^{-1} M + \rho\right) \quad (10)$$

Some features of the above conditional densities are worth discussing. First, note from the definition of D_i and d_i in equation (7) that the conditional posterior mean of β_i is in similar in form to a weighted average of the least-squares estimate using the data from country i: $(X_i'X_i)^{-1}X_i'y_i$ and the common or 'pooled' mean for all countries, β. As discussed in the introduction, Zellner and Hong (1989) describe and employ a 'γ forecast' by first coming up with a point prediction for β_i which is a weighted average of the least-squares estimate for that country and a pooled least-squares estimate for all countries. Formally, Zellner and Hong used

$$\hat{\beta}_i^* = (\hat{\beta}_i + \gamma\hat{\beta})/(1 + \gamma)$$

and then obtained one-step-ahead predictions using $\hat{y}_{iT+1} = x_{iT+1}\hat{\beta}_i^*$. In their model, $\hat{\beta}_i$ was the least-squares coefficient vector for country i, and $\hat{\beta}$ was the pooled least-squares coefficient vector with all β_i assumed equal. Zellner and Hong implemented their approach by selecting different values of γ, using these values to obtain $\hat{\beta}_i^*$, and then using these parameters to forecast the future output growth rates.[11] In the framework above, the weighting of this shrinkage estimator is determined by Λ^{-1}, which is treated as a parameter of interest and determined endogenously. To get some intuition about how Λ^{-1} is weighting the restricted and pooled estimates, suppose that Λ is diagonal. If the diagonal entries of Λ^{-1} are near zero, or equivalently, if the diagonal elements of Λ are 'large', then the conditional posterior mean of β_i is approximately centered around the least-squares estimate for country i with a posterior covariance matrix equal to the traditional least-squares covariance matrix. From equations (3) and (10) and the discussion below we see that a 'large' Λ emerges when there is a large amount of posterior variability of the individual countries' forecasts around the common mean β. If this is the case, then equation (7) downweights the pooled β and places most weight on the least-squares forecast for that country, an intuitively sensible result. Similarly, if Λ^{-1} is large, then the conditional posterior mean of β_i is approximately the pooled coefficient, β. Thus, the conditional posterior mean of β_i in equation (7) reproduces the shrinkage estimator described in Zellner and Hong (1989) and others, and determines the weighting of the pooled and country-specific forecasts endogenously.

Thus far, I have not discussed how the simulated parameter values obtained from the Gibbs sampler can be employed to predict countries' output growth rates and the median of a set of countries' output growth rates. To this end, let y_{if} denote the future, as yet unobserved growth rate for country i and further assume:

$$y_{if} = x_{if}\beta_i + \varepsilon_{if} \qquad \varepsilon_{if} \sim N(0, \sigma_i^2) \quad (11)$$

[11] Zellner and Hong (1989) found that the weight γ was equal to the ratio of two variance parameters, but did not compute the posterior distribution of γ.

so that our model in equation (1) holds for the future growth rates of each country. The posterior predictive density of y_{if} can be obtained by marginalizing the conditional predictive density over the posterior $p(\theta|\text{Data})$:[12]

$$p(y_{if}|x_{if}, \text{Data}) = \int_{\Theta} p_N(y_{if}; x_{if}\beta_i, \sigma_i^2|\text{Data})p(\theta|\text{Data})\text{d}\theta \tag{12}$$

Draws from the posterior predictive density for y_{if} can be obtained using the post-convergence draws from the Gibbs output. This is like performing direct Monte Carlo integration using equation (12) where draws from $p(\theta|D)$ are taken from the Gibbs sequence after it has been determined to converge. Draws from the marginal posterior predictive density of y_{if}, say $\{y_{if}^{(j)}\}_{j=1}^{J}$ can thus be obtained by taking draws from the following normal density:

$$y_{if}^{(j)} \sim N(x_{if}\beta_i^{(j)}, \sigma_i^{2(j)})$$

where $\beta_i^{(j)}$ and $\sigma_i^{2(j)}$ denote the *jth* post-convergence draw obtained from the Gibbs sampler. Repeating this process $\forall i = 1, 2, \ldots, M$ gives a set of draws from the marginal posterior predictive densities of all countries. Repeating this process for many j allows computation of posterior predictive means, interquartile ranges, and smoothed predictive densities for each country. For example, the posterior predictive mean of y_{if} can be consistently estimated by the sample average of the simulated values described above: $\hat{E}(y_{if}|x_{if}, \text{Data}) = 1/J\sum_{j=1}^{J} y_{if}^{(j)}$. Intervals, smoothed posterior predictive densities, and other statistics can be obtained similarly, as shown in the next section. Given draws from the marginal posterior predictive density of each country, we can use these draws to simulate the posterior predictive distribution of the future median growth rate, denoted $w_f \equiv \text{Median}\{y_{1f}, y_{2f}, \ldots, y_{Mf}\}$, which was the object of interest in Zellner and Hong (1989), and Zellner and Tobias (1999). Given the j^{th} round of draws for each country, $(\{y_{if}^{(j)}\}_{i=1}^{M})$, the sample median of these draws can be used as a draw from the posterior predictive density of w_f:[13]

$$w_f^{(j)} = \text{Median } \{y_{1f}^{(j)}, y_{2f}^{(j)}, \ldots, y_{Mf}^{(j)}\}. \tag{13}$$

Repeating this process for many j enables calculation of the posterior predictive density of w_f from which point forecasts can be obtained, as done below. In previous work on this topic, point forecasts of the future median output growth rate were obtained as the median of a set of point forecasts for each country. The approach listed above offers improvement over this method, since the distribution of the future median output growth rate can be obtained, and point and interval forecasts can be based on this simulated posterior predictive density.

[12] Again, it is important to note from equation (1) that x_{if} is known at time *T*, and thus we do not require any knowledge of future covariates to obtain our one-step-ahead forecasts.

[13] Using the post-convergence draws of the Gibbs sampler enables us to regard $y_f^{(j)} \equiv [y_{1f}^{(j)}, \ldots y_{Mf}^{(j)}]$ as a draw from the joint posterior predictive density of the countries' output growth rates. This holds for each j so that $\{y_f^{(j)}\}_{j=1}^{J}$ is stationary. Provided $E(w_f) < \infty$, $\text{Var}(w_f) < \infty$, and the covariance of the terms w_f^j, w_f^{j+k} converges to zero as k grows, we can apply Chebyshev's inequality to show that $\frac{1}{J}\sum_{j=1}^{J} w_f^{(j)} \xrightarrow{P} E(w_f)$. We can also express $\Pr(a \leq w_f \leq b)$ as an expectation of an indicator function, whence we can again obtain a consistent estimate of this probability by taking the fraction of the times that w_f^j falls in the interval [*a*, *b*]. This suggests that a histogram of the w_f^j's will recover the density of the future median output growth rate. Alternatively, this histogram can be smoothed by applying a kernel density estimator to the values of w_f^j's, as stated below (13). See Thisted (1988) for a similar discussion in which Monte Carlo integration is used to simulate the distribution of the α-trimmed mean.

The procedure above can be extended to obtain multiple-step-ahead forecasts. For simplicity (and without loss of generality), take the restricted case of an AR(1) model, and suppose we wish to forecast two periods into the future. To do this, we note that

$$\begin{aligned} p(y_{iT+2}|\text{Data}) &= \int_{y_{iT+1}} \int_{\Theta} p(y_{iT+2}, y_{iT+1}, \theta|\text{Data})\mathrm{d}\theta \mathrm{d}y_{iT+1} \\ &= \int_{y_{iT+1}} \int_{\Theta} pN(y_{iT+2}|y_{iT+1}, \theta, \text{Data})pN(y_{iT+1}|\theta, \text{Data})p(\theta|D)\mathrm{d}\theta \mathrm{d}y_{iT+1} \end{aligned}$$

The first conditional in the last line is known (given y_{iT+1}) from equation (11) since we assume that our model continues to hold two steps into the future. To obtain a draw from the posterior predictive of y_{iT+2}, we simply add one more step to our simulation algorithm. That is, we draw θ from its marginal posterior, use this draw to draw from the predictive for y_{iT+1} (as stated below equation (12)), and then use both draws to draw y_{iT+2} from its conditional normal given the drawn value of y_{iT+1} and θ. Higher-order AR relationships can be handled similarly. If the model contains leading indicator variables, then values for these variables must be forecasted in order to forecast y_{iT+2}. Such an approach has been used in Zellner and Hong (1989), and is also implemented in the following section.

The forecasting algorithm

Since the procedure just described is rather computationally involved, it is useful to clearly articulate the steps needed to obtain our one-step-ahead output growth rate and median output growth rate forecasts. The following provides an outline of these steps as well as further details regarding the Gibbs sampling algorithm.

(1) Obtain the past data to be used in estimation and the true one-step-ahead output growth rates from the hold-out sample. For the base year, for example, we get data for years $\leq$ 1984 to estimate the model and keep the actual outcomes in 1985.
(2) Run the Gibbs sampler to fit the model using the past data. In practice, we implement the Gibbs sampler as follows:
   (a) Choose initial parameter values, $\beta(0)$, $\{\beta_i(0)\}$, $\{\sigma_i^2(0)\}$, and $\Lambda^{-1}(0)$.
   (b) Draw $\{\beta_i(1)\}$ from equation (7) after setting $\beta = \beta(0)$, $\{\sigma_i^2\} = \{\sigma_i^2(0)\}$, $\Lambda = \Lambda(0)$.
   (c) Draw $\beta(1)$ from equation (8) after setting $\{\beta_i\} = \{\beta_i(1)\}$, $\Lambda = \Lambda(0)$.
   (d) Draw $\{\sigma_i^2(1)\}$ from equation (9) after setting $\{\beta_i\} = \{\beta_i(1)\}$.
   (e) Draw Λ^{-1} from equation (10) after setting $\{\beta_i\} = \{\beta_i(1)\}$, and $\beta = \beta(1)$.
   (f) Repeat until the desired number of draws are obtained. At each iteration, condition on the most recent values of the simulated parameters.
(3) Discard the draws from the burn-in period, and keep the post-convergence draws. Use each post-convergence draw to obtain a draw from the posterior predictive density for y_{if}, as done numerically below equation (12). For each set of M simulated future growth rates, obtain a draw from the posterior predictive density of the future median output growth rate as in equation (13).
(4) Compute one-step-ahead forecasts of the output growth rates for each country and the median output growth rate. (Here we use the mean as our point estimate.) Compare these predictions to the true values from the hold-out sample to calculate RMSE and MAE.

(5) Repeat. Add the data for the year just forecasted to the past data, and fit the model again for this expanded data set. Compute output growth rate and median output growth rate forecasts numerically from the new post-convergence Gibbs output.

THE DATA AND RESULTS

The data used in this analysis come from the IMF International Financial Statistics database which is maintained and updated online at the University of Chicago's Graduate School of Business. I use data on the following sixteen industrialized countries: Australia, Austria, Belgium, Canada, Denmark, Finland, France, Ireland, Japan, the Netherlands, Norway, Spain, Sweden, Switzerland, the UK, and the USA.[14] The longest series contain annual data from 1948 to 1995, but data points are missing for some countries at the beginning of the series, making the panel unbalanced. Real growth rates for all variables are created by first dividing each nominal variable from the IMF database by a country-specific price index. The real variables are then logged and first-differenced to create growth rates.

To fit the model, a burn-in period of 1000 iterations is discarded each time the Gibbs sampler is run, and every third value of the following 20,000 draws is used to compute the output growth rate and median output growth rate forecasts. The latter is done to reduce the autocorrelation in the chain, although results are very similar when the full chain is used. Following Gelman and Rubin (1992), several chains were run with different and overdispersed starting values to check for convergence. After the initial 1000 values, all chains appeared to settle down to explore the same region. Further, kernel-smoothed posterior predictive densities of the future growth rates of several countries were nearly identical across the chains, providing evidence that this Monte Carlo sample size is adequate for this application.

Figure 1 plots the priors and posteriors for a representative set of regression parameters. These results are obtained for the last run of the Gibbs sampler, and thus all data up to and including 1994 are used in the calculations. Figure 1A presents results for the pooled AR(1) coefficient (the (2,1) element of β), Figure 1B for the AR(1) parameter for the USA,[15] Figure 1C for the US variance parameter, and Figure 1C for the pooled growth rate of real money coefficient (the (7,1) element of β). These density estimates are computed non-parametrically by kernel-smoothing simulated values from the priors and posteriors. We first see that a significant amount of learning occurs from the data, since the priors are much flatter than their corresponding posteriors for all parameters. Further, the pooled AR1 and the lagged money posteriors generally place small mass to the left of zero, indicating that these variables tend to have a positive effect on future output growth rates. Finally, we see some departure between the USA and pooled AR1 coefficient, indicating that the data always do not push us toward the pooled forecast, at least for the USA.

---

[14] In earlier work on this topic, Garcia-Ferrer *et al.* (1987), Zellner and Hong (1989), Zellner *et al.* (1991), Min and Zellner (1993), and Zellner and Tobias (1999) combine similar data with data on Germany and Italy and use these data in their analysis. In this paper, I have focused on the above sixteen countries and omitted Germany and Italy due to several missing observations.

[15] The plot of the prior is obtained by drawing β and Λ^{-1} from their marginal prior distributions, plugging these drawn values into the conditional normal prior for β_{US} (equation (3)), and then taking a draw from this distribution. The (2,1) element of this vector is stored, and the kernel-smoothed density of these values is reported.

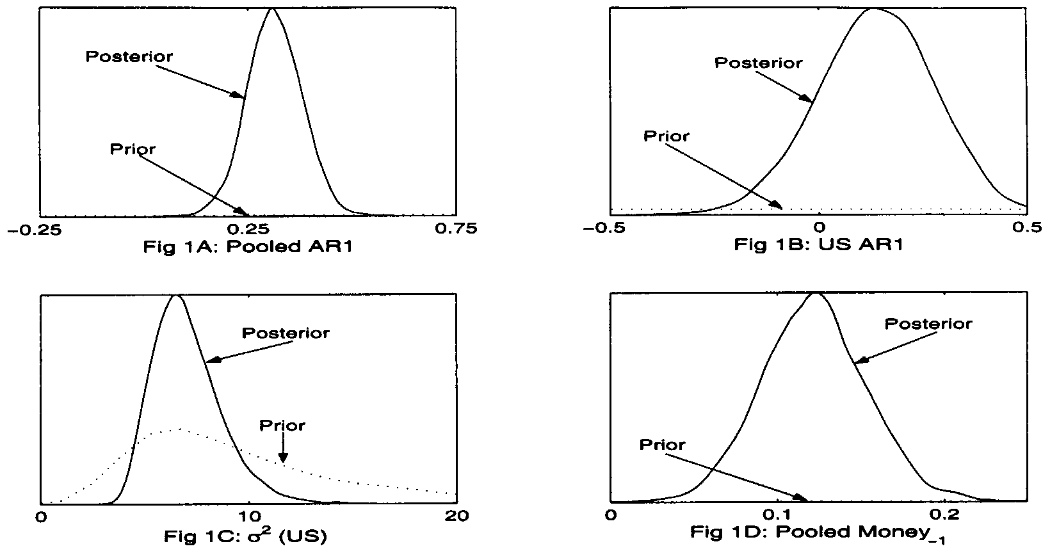

Figure 1. Priors versus posteriors for a selection of parameters

Forecasting results

We begin by discussing the performance of our median output growth rate forecasts. To provide a benchmark for the performance of forecasts obtained using the methods of the second and third sections predictive RMSEs and MAEs[16] associated with a variety of forecasting models were obtained for the forecasting period 1985–1995. One of these benchmark forecasting models has been employed in Zellner and Hong (1989), and Zellner and Tobias (1999). This model specifies an AutoRegressive Leading Indicator (AR(3)LI) model for the median output growth rate w_t:

$$w_t = \alpha_0 + \alpha_1 w_{t-1} + \alpha_2 w_{t-2} + \alpha_3 w_{t-3} + \alpha_4 MSR_{t-1} + \alpha_5 MedMon_{t-1} + u_t \tag{14}$$

where MSR_t denotes the median of the sixteen countries' growth rates of real stock prices in year t, and $MedMon_t$ is the median of the sixteen countries' growth rates of real money in year t. Least-squares forecasts using the model above are obtained and predictive RMSE and MAE of the one-step-ahead least-squares forecasts are reported in Table AI of the Appendix.

Four other benchmark forecasting models for the median output growth rate are also analysed, and the performances of these models are compared to the performance of forecasts obtained from our hierarchical Bayesian analysis. Two of these four models are based on least-squares forecasts of the relationships in equation (1). As in previous work, given a set of least-squares forecasts for each country ($\{\hat{y}_{if}\}_{i=1}^{M}$), we can obtain an approximate point forecast of the future median output growth rate as follows:

$$\hat{w}_f = \text{median } \{\hat{y}_{1f}, \hat{y}_{2f}, \ldots, \hat{y}_{Mf}\} \tag{15}$$

One-step-ahead least-squares forecasts are obtained as above using both the unrestricted relationships in equation (1), and restricted least-squares forecasts in which the regression coefficients and

[16] We define RMSE of the Median forecast as $\sqrt{1/T \sum_{t=1}^{T} (\hat{w}_t - w_t)^2}$ and MAE as $1/T \sum_{t=1}^{T} |\hat{w}_t - w_t|$. For this analysis $T = 11$.

variance parameters (β_i and σ_i^2) are assumed equal across countries. Note that the procedure in equation (15) offers a computationally convenient, but perhaps inappropriate, method for forecasting the future median output growth rate. The final two forecasting models carried along in this analysis are the simple random walk model: $\hat{w}_{T+1} = w_T$, and a model which forecasts the future median as the sample average of countries' current growth rates: $\hat{w}_{T+1} = 1/M \sum_{i=1}^{M} y_{iT}$.[17]

The five benchmark forecasting models above are combined with the hierarchical Bayes (HB) model of the second and third sections and predictive RMSE and MAE for all six models are presented in Table AI. One-step-ahead forecasts using the post-convergence draws of the Gibbs sampler were computed as the sample mean of the simulated values from the posterior predictive density of w_f. This point estimate is an approximation to the posterior predictive mean of w_f, which is known to be optimal relative to quadratic loss.

The first item to note from Table AI is that the median forecasts obtained from the hierarchical Bayesian model described in the second and third sections produced the lowest predictive RMSE and MAE, which are 1.42 and 1.11, respectively. Second-best is the random walk model, which produces a RMSE and MAE of 1.50 and 1.22, respectively. The results in the table are also in agreement with the conclusions of Garcia-Ferrer *et al.* (1987), Zellner and Hong (1989), Zellner *et al.* (1990), LeSage (1990), and Zellner and Tobias (1999). These authors found that a reduction in parameterization by pooling coefficients across countries lowers predictive RMSE and MAE of the least-squares forecasts. From the table, we see that this is indeed the case when forecasting the median output growth rate. Unrestricted least-squares forecasts based on equation (1) yield a RMSE and MAE of 1.72 and 1.37, respectively, while pooled least squares forecasts which impose $\beta_i = \beta \forall_i$ produce a RMSE and MAE of 1.62 and 1.19, respectively.[18] Least-squares forecasting results based on the AR(3)LI for the median growth rate in equation (14) are also higher than the hierarchical Bayes results, and are quite similar to the pooled least-squares results based on equation (1).

In this paper I have argued that the median forecasts produced using the hierarchical Bayes model offer a methodological improvement over the estimation approach based on equation (9). For the hierarchical Bayesian calculations, we regarded the future median output growth rate as the random variable of interest, and described a way to simulate numerically the posterior predictive density of that random variable. We then used the mean of this posterior predictive density as our point forecast. The competing approach, which has been used in past work, obtains a set of point predictions for each country and then takes the median of this set of forecasts as the forecast of the future median growth rate. This approach entirely avoids calculation of the predictive density for the median growth rate, and thus we might expect that it will produce relatively poor forecasts in practice. As shown in Table AI, the hierarchical Bayes approach does produce the lowest RMSE and MAE of all the competing forecasting models.

Simulating the posterior predictive density of the future median output growth rate w_f also enables calculation of intervals associated with the median forecasts and thus provides a means to further assess the accuracy of our forecasting procedure. Figure 2 contains a graph of the true and predicted median growth rates as well as a vertical band giving ± 2 standard deviations from the estimated posterior mean (the point forecast) over the forecast period: 1985–1995. From this

[17] I thank a referee for making this suggestion as a benchmark forecasting model.

[18] Similar conclusions are drawn from Table AII which breaks the forecasts down by country. Pooled least-squares forecasts produce a lower RMSE than the unrestricted least-squares forecasts for all countries except Finland, France and the UK. There was a also a tie for Austria.

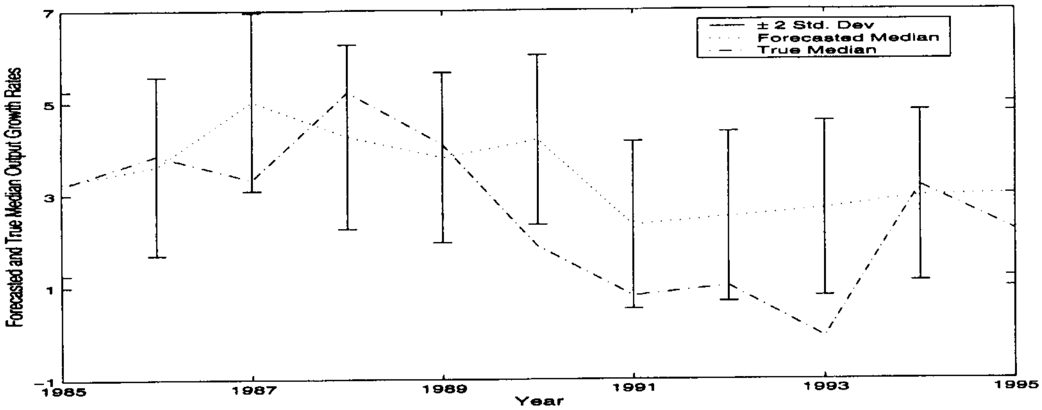

Figure 2. True and predicted medians: 1985–1995

graph we note that the true median output growth rate falls within these standard error bands in 9 of 11 of our forecasts, or in all years except 1990 and 1993. From the figure we also see that our forecasting results fail to predict the extent of the 'world downturn' of the early 1990s, suggesting that there remains room for improvement in our forecasting performance. It is important to note that the studies previously mentioned have not presented a graph such as Figure 2 because of the difficulty in determining standard errors when using equation (15) as a forecasting rule. The results presented here suggest that median forecasts obtained using the method in the second and third sections offer a methodological improvement over previous work on this topic, and given the results in Table AI, also seem to offer an improvement in actual forecasting performance. This improvement in forecasting performance also extends when conditioning on the forecasting period used in Zellner and Hong (1989). Use of the hierarchical Bayes model produced a RMSE and MAE of the median output growth rate equal to 1.78 and 1.54, respectively over the period 1974–1984. Pooled least-squares (the 'complete shrinkage' case) yields a RMSE and MAE of 2.59 and 2.20, respectively, while unrestricted least-squares yields 2.48 and 2.08 as the RMSE and MAE for the same forecasting period.[19]

In addition to forecasting the median output growth rate, output growth rate forecasts can also be obtained and evaluated at the country level. Shown in Table AII are the RMSEs and MAEs broken down by country for the same forecasting period 1985–1995. Again we carry along forecasts obtained from the hierarchical Bayes model, the random walk and current mean forecasts, and the disaggregated and pooled AR(3)LI models fitted by least-squares. As shown in the table, forecasts obtained from the Hierarchical Bayes model produce a median RMSE which is lower than both the pooled least-squares and unrestricted least squares forecasts, although the improvement relative to pooled least-squares is very modest. In fact, forecasts obtained using the hierarchical Bayes model only beat the pooled least-squares forecasts in 7 of 16 of the countries' forecasts (Australia, Austria, Denmark, France, Switzerland, the UK, and the USA).[20] This result is consistent with

[19] Zellner and Hong (1989) report smaller values for RMSE and MAE for the period 1974–1984. It is important to note that this paper focuses on a smaller set of sixteen countries and uses revised IMF data.

[20] The reader may note the poor performance of forecasting output growth rates for Finland across all competing models. The data used in this analysis revealed a large negative growth rate for Finland in 1991 and 1992 which was not predicted by any of our forecasting models.

the conclusions of Zellner and Hong (1989), and Zellner and Tobias (1999), where pooled least-squares forecasts (or the case of 'complete shrinkage') generally produced the lowest RMSE and MAE, and were clearly better than unrestricted least-squares forecasts. The competitiveness of the pooled forecasts relative to the hierarchical Bayes model is somewhat surprising, given that conditional posterior mean of each β_i when using HB can be thought of as an endogenously weighted average of the pooled and unrestricted least-squares estimate, as discussed in the previous section. However, there is no guarantee that this endogenous determination of the weights will improve the performance of each countries' forecasts.

Further, we note from Table AII that the simple random walk model produces the lowest median RMSE and MAE of the competing models and beats the hierarchical Bayes forecasts for 8 of 16 countries (Belgium, Canada, Denmark, Finland, Ireland, Japan, the Netherlands, and Sweden). Although the hierarchical Bayes forecasts generalize the pooled and unrestricted forecasts and seem to outperform these least-squares forecasts given the same model specification, there is no reason to believe that these forecasts will necessarily beat all simple forecasting rules. Indeed, these results show that the random walk model produces a median RMSE slightly lower than the hierarchical Bayes results, and beats HB for one-half of the 16 countries.

One possible direction to take, which is motivated by the solid performance of these simple forecasting models, is to simplify our AR(3)LI specification in equation (1). To this end, I analysed a simplified AR(1) in our hierarchical Bayesian framework. For these calculations, I kept the same prior specifications for the countries' variance parameters, and centred the pooled β vector at [3 0] (corresponding to the constant and AR(1) parameters, respectively). To complete the vague prior specifications, I set $\rho = 3, R = \text{Diag}[4\ 1]$, and $\Sigma = \text{Diag}[100\ 25]$. The country-level forecasts obtained from this HB AR(1) are provided in the last column of Table AII. We see that this forecasting model appears to offer a significant reduction in the RMSE of the individual countries' forecasts relative to the other competing models in the table. The hierarchical Bayes AR(1) clearly produced the lowest median RMSE of the individual countries' forecasts, which was 2.02. Though not shown in Table AII, least-squares forecasts for the pooled AR(1) produced a median RMSE equal to 2.09 while the disaggregated AR(1) fitted by least squares also yielded a median RMSE of 2.02. Although the disaggregated AR(1) forecasts are clearly competitive with the hierarchical Bayes results, the HB AR(1) forecasts beat the disaggregated least-squares AR(1) forecasts for 11 of the 16 countries (Australia, Austria, Belgium, Canada, Finland, France, Japan, Norway, Sweden, Switzerland and the USA) and tied for the UK. This again shows that given the same model specification, the endogenously-weighted HB forecasts perform better than (or at least as well as) both pooled and restricted least-squares.[21] Further, we see that simple forecasting models perform quite well when evaluated at the country level.

A different direction to take, which has been suggested by Zellner and Hong (1989) and Zellner and Tobias (1999), is to complicate the AR(3)LI model in equation (1) and determine if this elaborated model can produced improved forecasts. These authors augment the specification in equation (1) by including the current median output growth rate in the growth rate equations for each country:

$$y_{it} = w_t\phi_i + x_{it}\gamma_i + u_{it} \tag{16}$$

[21] None of these AR(1) forecasting models produced lower. RMSE or MAE for the median output growth rate than the HB AR(3)LI. The RMSE and (MAE) of the median forecasts were 1.59 (1.30), 1.48 (1.26), and 1.45 (1.18) for the disaggregated AR(1), pooled AR(1) and HB AR(1), respectively.

Note that the specification above also induces another type of shrinkage forecast. If $\phi_i = 1$ and $\gamma_i = 0$ then forecasts for each country will just reduce to the median forecast. For $\phi_i = 0$, forecasts will reduce to the forecasts obtained from equation (1). Other parameter values will reflect a compromise between the median forecast and information that is specific to each country. Since w_{T+1} is not known during the forecasting period, it must be forecasted in order to obtain output growth rate forecasts from equation (16). Zellner and Hong (1989) and Zellner and Tobias (1999) use equation (14) to forecast w_{T+1} and then use this point forecast to forecast the future growth rates for each country. I implement this approach using the Gibbs sampler, after first augmenting x_{it} in equation (1) by also including w_t. To obtain forecasts in the spirit of Zellner and Hong (1989) and Zellner and Tobias (1999), I fix $w_{T+1} = \hat{w}_{T+1}$ and obtain the conditional posterior predictive density for each country: $p(y_{iT+1}|w_{T+1} = \hat{w}_{T+1}, x_{if}, D)$.[22] Draws from this predictive density can then be obtained numerically as in equation (12) using the post-convergence draws of the Gibbs sampler.

Forecasts using this expanded model did not lower the RMSE and MAE of the one-step-ahead output growth rate forecasts for each country. The median RMSE obtained from our hierarchical Bayesian analysis of the model in (16) was 2.42, which was higher than the results obtained from HB model without including w_t in equation (1) and also higher than the pooled least-squares forecasts presented in Table AII. Further, the elaborated results from the Hierarchical Bayes model in equation (16) produced lower RMSE and MAE than HB AR(3)LI model in equation (1) in only six of the sixteen countries' forecasts (Belgium, Canada, Finland, Ireland, Norway, and the USA). This indicates that country-level forecasts did not improve through the inclusion of the current median output growth rate, w_t, unlike earlier work on this topic.[23]

CONCLUSIONS

This paper has described Hierarchical Bayesian procedures for forecasting countries' output growth rates, and medians of a set of output countries' growth rates. This model generalized earlier work on this topic in Garcia-Ferrer *et al.* (1987), Zellner and Hong (1989), Zellner *et al.* (1991), Min and Zellner (1993), and Zellner and Tobias (1999). These past studies argued that shrinkage forecasts, in which coefficient estimates for each country are obtained as a weighted average of least squares estimates for that country and a pooled least-squares estimate for all countries, can improve predictive RMSE and MAE of countries' output growth rates and median output growth rate forecasts. In this paper, I showed how such a shrinkage estimator emerges from a Hierarchical Bayesian model. In this framework, it was shown that the conditional posterior means of the regression coefficients for each country were in the form of the shrinkage estimator described in these earlier studies, where the weights associated with the shrinkage estimator were determined endogenously. The approach described in this paper also allowed us to simulate the posterior predictive density of the median output growth rate and obtain point and interval forecasts based on this predictive density. This approach offered a significant methodological improvement over previous work in which point forecasts of the future median output growth rate were obtained using a potentially inaccurate approximation. Finally, it is important to note that the techniques described in this paper are quite operational and can be applied to a variety of forecasting problems.

[22] Of course, we could also obtain the marginal posterior predictive density by integrating out w_{T+1} numerically.

[23] It is important to note that earlier studies used an expanded set of eighteen countries and generally used data prior to 1985 for estimation and prediction.

APPENDIX

Table AI. RMSEs and MAEs of one-step-ahead forecasts of the median output growth Rate, w_t, using alternative forecasting models: forecasting period: 1985–1995

Model rule	RMSE	MAE
Random walk: $\hat{w}_{T+1} = w_T$	1.50	1.22
Current mean: $\hat{w}_{T+1} = 1/M\,\Sigma_{i=1}^{M} y_{iT}$	1.65	1.36
AR(3)LI for w_t (equation (14))	1.61	1.19
Pooled AR(3)LI (equation (1) with $\beta_i = \beta$)	1.62	1.19
AR(3)LI (equation (1))	1.72	1.37
Hierarchical Bayes AR(3)LI	1.42	1.11

Table AII. RMSEs of one-step-ahead forecasts broken down by country: forecasting period: 1985 –1995

Country	Hierarchical Bayes (HB)	Pooled least-squares ($\beta_i = \beta$)	Unrestricted least-squares	Random walk	$\hat{y}_{iT+1} = 1/M\,\Sigma_{i=1}^{M} y_{iT}$	HB AR(1)
Australia	2.85	3.07	3.47	3.57	2.86	2.78
Austria	1.52	1.91	1.91	1.54	1.44	1.41
Belgium	2.05	1.82	2.61	1.85	1.50	1.53
Canada	3.22	3.19	3.46	2.87	2.98	2.88
Denmark	2.27	2.28	2.59	1.59	2.31	1.73
Finland	6.31	6.17	5.87	4.75	4.51	4.63
France	1.45	1.98	1.22	2.11	1.48	1.92
Ireland	2.64	2.34	2.77	2.27	2.96	2.13
Japan	2.11	1.65	3.15	1.34	1.14	1.78
Netherlands	2.14	1.81	2.46	1.59	1.02	1.56
Norway	3.24	3.22	3.73	3.89	3.01	3.00
Spain	1.96	1.93	2.58	2.19	1.87	2.03
Sweden	3.06	2.94	3.47	2.98	2.72	3.06
Switzerland	2.10	2.12	2.14	2.15	1.73	2.02
UK	2.31	3.31	2.68	2.33	2.81	2.41
USA	2.26	2.29	3.17	2.66	2.56	1.81
Median	2.26	2.29	2.73	2.23	2.43	2.02

Table AIII. Selected coefficients' posterior means and (standard deviations) using full sample ($\leq$1994)

	Constant	y_{t-1}	y_{t-2}	y_{t-3}	SR_{t-1}	SR_{t-2}	GM_{t-1}	WSR_{t-1}
Common mean β	1.65	0.315	−0.040	0.122	0.018	−0.019	0.122	0.017
	(0.303)	(0.066)	(0.068)	(0.061)	(0.016)	(0.011)	(0.032)	(0.022)
β_i (USA)	2.18	0.136	0.026	−0.011	0.066	−0.032	0.186	−0.047
	(0.571)	(0.142)	(0.136)	(0.121)	(0.033)	(0.022)	(0.049)	(0.044)
β_i (UK)	1.78	0.277	−0.140	0.026	0.045	−0.010	0.051	−0.048
	(0.536)	(0.141)	(0.143)	(0.127)	(0.030)	(0.022)	(0.037)	(0.044)
Diagonal elements of	0.645	0.035	0.036	0.028	0.003	0.001	0.010	0.005
covariance matrix Λ	(0.336)	(0.017)	(0.018)	(0.014)	(0.001)	(0.001)	(0.005)	(0.002)

ACKNOWLEDGEMENTS

This paper evolved from previous work with Arnold Zellner during my graduate studies in the Department of Economics at the University of Chicago. I would like to thank Arnold for his collaboration on a variety of projects, and also for introducing me to his vast and important work on forecasting international output growth rates. I would also like to thank David Brownstone, John DiNardo, Kaku Furuya, Garance Genicot, Mark Moore, Dale Poirier, Priya Ranjan, Gary Richardson, Alex Robson, two anonymous referees and the Editor for helpful comments and suggestions. All errors are, of course, my own.

REFERENCES

Baltagi BH, Griffin JM, Xiong W. 2000. To pool or not to pool: Homogeneous versus heterogeneous estimators applied to cigarette demand. *Review of Economics and Statistics* **82**: 117–126.

Besag J. 1974. Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Ser. B* **36**: 192–196.

Carlin B, Louis T. 1996. *Bayes and Empirical Bayes Methods for Data Analysis: Monographs on Statistics and Applied Probability 69*. Chapman & Hall: London.

Carlin B, Polson N. 1991. Inference for nonconjugate Bayesian models using the Gibbs sampler, *Canadian Journal of Statistics* **19**: 399–405.

Casella G, George EI. 1992. Explaining the Gibbs sampler. *The American Statistician* **46**: 167–174.

Garcia-Ferrer A, Highfield RA, Palm F, Zellner A. 1987. Macroeconomic forecasting using pooled international data. *Journal of Business and Economic Statistics* **5**: 53–67.

Gelfand AE, Hills SE, Racine-Poon A, Smith AFM. 1990. Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association* **85**: 972–985.

Gelfand AE, Smith AFM. 1990. sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**: 398–409.

Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* **87**: 97–109.

Hong C. 1989. Forecasting Real Output Growth Rates and Cyclical Properties of Models: A Bayesian Approach. PhD dissertation, Department of Economics, University of Chicago.

Kazimi C, Brownstone D. 1999. Bootstrap confidence bands for shrinkage estimators. *Journal of Econometrics* **90**: 99–127.

LeSage JP. 1990. Forecasting turning points in metropolitan employment growth rates using Bayesian exponentially weighted regression, time-varying parameter and pooling techniques. *Journal of Regional Science* **30**: 533–548.

Lindley DV, Smith AFM. 1972. Bayes estimates for the linear model. *Journal of the Royal Statistical Society, Ser. B* **34**: 1–41.

Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equations of state calculations by fast computing machines. *Journal of Chemical Physics* **21**: 1087–1091.

Min C-K. 1992. Economic Analysis and Forecasting of International Growth Rates Using Bayesian Techniques. PhD dissertation, Department of Economics, University of Chicago.

Min C-K, 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.

Robert CP, Casella G. 1999. *Monte Carlo Statistical Methods*. Springer-Verlag: New York.

Smith AFM, Gelfand AE. 1992. Bayesian statistical without tears: a sampling–resampling perspective. *The American Statistician* **46**: 84–88.

Thisted RA. 1988. *Elements of Statistical Computing: Numerical Computation*. Chapman & Hall: New York.

Tierney L. 1994. Markov chains for exploring posterior distributions. *The Annals of Statistics* **22**: 1701–1762.

Zellner A. 1999. Bayesian and non-Bayesian approaches to scientific modeling and inference in economics and econometrics. H.G.B. Alexander Research Foundation, Graduate School of Business, University of Chicago.

Zellner A, Hong C. 1989. Forecasting international growth rates using Bayesian shrinkage and other procedures. *Journal of Econometrics* **40**: 183–202.

Zellner A, Hong C, Gulati G. 1990. Turning points in economic time series, loss structures and Bayesian forecasting. In *Bayesian and Likelihood Methods in Statistics and Econometrics*, Geisser S, Hodges JS, Press SJ, Zellner A (eds). Elsevier Science Publishers: New York 371–393.

Zellner A, Hong C, Min C-K. 1991. Forecasting turning points in international output growth rates using Bayesian exponentially weighted autoregression, time-varying parameter and pooling techniques. *Journal of Econometrics* **49**: 275–304.

Zellner A, Min C-K. 1995. Gibbs sampler covergence criteria. *Journal of the American Statistical Association* **90**: 921–927.

Zellner A, Tobias J. 1999. A note on aggregation, disaggregation and forecasting performance. *Journal of Forecasting*. (forth coming).

Author's biography:

Justin Tobias is Assistant Professor of Economics at the University of California-Irvine. He received his PhD in Economics from the University of Chicago in 1999. His research interests focus on simulation-based methods for inference, partially linear modelling, and the application of such techniques to issues in the economics of education.

Author's addresses:

Justin Tobias, Department of Economics, University of California-Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA.