CONSTRAINED FORECASTS IN ARMA MODELS: A BAYESIAN APPROACH

Enrique de Alba
I.T.A.M.
Rio Hondo No. 1
Mexico, D.F.

Omar Aguilar
Institute of Statistics and Decision Sciences
Duke University
Durham, N.C.

September, 1995

Abstract

A Bayesian approach is developed to generate constrained and unconstrained forecasts in autoregressive-moving average time series models. Both are calculated by formulating the ARMA(p,q) model in such a way that it is possible to numerically compute the predictive distribution for any number of forecasts as in de Alba (1993). We obtain the posterior distribution of the parameters via Gibbs sampling and the predictive distribution using Monte Carlo integration. The kind of constraints used are that a linear combination of the forecasts equals a given value. The constrained forecasts are obtained by conditioning in the predictive distribution of the unconstrained forecasts. The results are applied to a series of quarterly inflation in Mexico.

1 INTRODUCTION

The literature on Bayesian methods applied to ARMA time series models is rather limited. It refers mostly for AR models. Early references include Zellner (1971), Harrison and Stevens (1976) and Newbold (1973). In most cases the applications are restricted to very simple models, or they only forecast one or two future periods. In part this has been due to the impossibility or the cost of dealing with the computational problems associated with the theoretical results.Thompson and Miller (1986) improve on previous Bayesian results by simulating joint predictive distributions. They first randomly generate parameter sets from the posterior distribution. These are then used to simulate future paths of the time series for AR models. Further, Naylor and Marriott (1994) carry out exact Bayesian analysis of non-stationary AR series relaxing the usual stationarity assumptions. McCulloch and Tsay (1994) use the Gibbs sampler for Bayesian analysis also of AR models.

Specifically in relation to ARMA models, Monahan (1983) presents a procedure that combines selection and estimation. Selection of the model is based on the posterior probability that a given model is correct. Broemeling and Shaarawy (1986,1988) and Shaarawy and Broemeling (1985) proposed approximate analysis for MA and ARMA models. An extensive review of the literature can be found in Broemeling and Shaarawy (1988). With respect to finding the joint predictive density of (several) future observations with ARMA models, Zellner (1971) derived the predictive distributions for AR(p) models, p=1,2. Monahan (1983) also considers the simplest ARMA(p,q) models with p,q = 0,1,2. He computes the means and standard deviations of predictive densities up to four periods ahead. Shaarawy and Broemeling (1984,1985) approximate the predictive distribution for the one step ahead forecast with ARMA and MA processes. Marriott et.al. (1994) show how to obtain the predictive distribution for a vector of future values via the Gibbs sampler adn Monte Carlo Integration, assuming a prior distribution on the initial values of the series as well as the initial errors. West and Harrison (1989) developed a complete theory in Bayesian forecasting and dynamic models.

The problem of obtaining forecasts for multiple future values of a time series may be further complicated if these are expected to satisfy certain given constraints. The analyst might be in the situation where he has information for some future value of the variable, or for some function of future values. Suppose, for example, that the model was specified and estimated using quarterly data, but the analyst has information on the annual figure following the last year used in estimation. There are many possible sources for this kind of information, but mostly it will come from expert opinion. In any event, an informed user of the forecasts derived from the model will expect them to be consistent with whatever additional information may be available to him. For instance, it may be required that quarterly forecasts add up or average out to a given annual figure. This naturally imposes constraints on the forecasts. The problem has been dealt with using a Bayesian approach in a regression framework by de Alba (1988, 1992) and for AR(p) models by de Alba (1993). There are some related results on constrained forecasts in time series models derived from a frequentist point of view. Cholette (1982) considers a procedure to modify forecasts that takes into account the historical information of a series, via an ARIMA model, as well as partial prior information. The forecasts from the ARIMA model are combined with the given prior information assuming both are deterministic. Guerrero (1989, 1990, 1991a, 1993) extends the previous results to more general models than those entertained by Cholette.

This paper presents a Bayesian procedure to obtain forecasts of a quarterly time series that are consistent with given future annual figures. It uses a conditional likelihood. The time series follows an ARMA(p,q) process and there are no restrictions on p nor q. The constrained forecasts are derived from the predictive distribution by conditioning on the given constraint, as was done in de Alba (1993) for autoregressive models. In addition the predictive distribution of the annual figure is used to determine if a given annual value is consistent with the model estimated from quarterly historical data.

The approach presented here is similar to that in de Alba (1993). In this paper the ARMA(p,q) model is expressed in such a way that the mean and variance of the predictive distribution can be easily computed numerically by Monte Carlo integration. But more important, with our approach it is also easy to compute constrained (conditional) forecasts. For ease of exposition we assume quarterly data are used in the model and that the known future information is available in yearly terms. However the results presented here are valid in more general situations. They can easily be adapted for monthly data and quarterly future information or any other combination of frequencies for the time series and the constraint. Throughout the paper we assume the order of the process is known. A similar approach was used in de Alba and Aguilar-Chavez (1995), but using an approximation to the conditional Likelihood whereas here we use the exact conditional Likelihood.

The paper is structured as follows. Section 2 presents some analytic results about the likelihood function. A Bayesian analysis of an ARMA(p,q) model is developed in Section 3 along with procedures to compute both constrained and unconstrained forecasts. Section 4 describes how to use Gibbs sampling to produce Bayesian forecasts in ARMA(p,q) models . An example is given in Section 5 where the results are applied to a quarterly Mexican Inflation series. The Appendix describes the way to implement numerically the Bayesian forecasts with and without restrictions.

2 LIKELIHOOD FUNCTION IN ARMA(p,q) MODELS

Consider the following ARMA(p,q) model

$$\phi(B)W_t = \theta(B)\epsilon_t \quad where \quad \epsilon_t \sim N(0,\sigma^2) \;\; \forall t \in \mathbf{Z} \tag{1}$$

where $\phi(B)$ and $\theta(B)$ are lag polynomials of order p and q respectively, that is,

$$\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \phi_3 B^3 - ... - \phi_p B^p,$$

$$\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_3 B^3 - ... - \theta_q B^q,$$

and $\{W_t\}$ is a discrete stochastic process. We assume stationarity and invertibility, that is, the solutions of $\phi(x) = 0$ y $\theta(x) = 0$ are outside the unit circle.

Assume we have N observations of W_t corresponding to the process (1), let $S_N = \begin{pmatrix} W_1, & W_2, & ..., & W_N \end{pmatrix}'$ be the vector of observations and assume $\epsilon_p = \epsilon_{p-1} = \epsilon_{p-2} = ... = \epsilon_{1-q} = 0$. We can write the model (1) conditioning on the first p observations, assume they are known, and obtain the last $N-p$ errors, i.e.,

$$\epsilon_t = W_t - \sum_{i=1}^{p} \phi_i W_{t-i} + \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \qquad with \qquad t = p+1, p+2, ..., N.$$

We thus have a random sample of size $T = N - p$ form a Normal distribution $N(0, \sigma^2)$. The (conditional) likelihood function is

$$L(\underline{\phi}, \underline{\theta}, \sigma^2 \mid S_N) \propto (\sigma^2)^{-\frac{N-p}{2}} \exp\left\{-\frac{1}{2\sigma^2} \sum_{t=p+1}^{N} \epsilon_t^2\right\}, \tag{2}$$

where $\underline{\phi} \in \mathbf{R}^p$ and $\underline{\theta} \in \mathbf{R}^q$, that is,

$$\underline{\phi} = \begin{pmatrix} \phi_1, & \phi_2, & ..., & \phi_p \end{pmatrix}' \quad and \quad \underline{\theta} = \begin{pmatrix} \theta_1, & \theta_2, & ..., & \theta_q \end{pmatrix}'.$$

We can write the model as:

$$W_t = \underline{z}_t' \underline{\phi} + \epsilon_t + \underline{\epsilon}_t' \underline{\theta}$$

with $\underline{z}_t' = (W_{t-1}, W_{t-2}, \ldots, W_{t-p})$ and $\underline{\epsilon}_t' = (\epsilon_{t-1}, \epsilon_{t-2}, \ldots \epsilon_{t-q})$. Furthermore, with matrix notation we can rewrite it as:

$$\mathbf{W} = \mathbf{X}\underline{\phi} + \mathbf{G}\underline{\epsilon},$$

with the following matrices:
$\mathbf{W} = (W_{p+1}, W_{p+2}, \ldots, W_N)$, $\mathbf{X}' = \left(\underline{z}_{p+1}, \underline{z}_{p+2}, ..., \underline{z}_N\right)$, $\underline{\epsilon} = (\epsilon_{p+1}, \epsilon_{p+2}, \ldots, \epsilon_N)$, and

$$\mathbf{G} = \begin{bmatrix} 1 & 0 & 0 & 0 & ... & 0 \\ \theta_1 & 1 & 0 & 0 & ... & 0 \\ \theta_2 & \theta_1 & 1 & 0 & ... & 0 \\ ... & ... & ... & ... & ... & ... \\ 0 & 0 & 0 & 0 & ... & 1 \end{bmatrix}_{(T \times T)},$$

where $\underline{\epsilon} \sim N(0, \sigma^2 I_T)$. Thus, we can write the likelihood functions as:

$$L(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W}) \propto (\sigma)^{-T} |\mathbf{V}|^{1/2}\, exp\left\{-\frac{1}{2\sigma^2}(\mathbf{W} - \mathbf{X}\underline{\phi})'\mathbf{V}(\mathbf{W} - \mathbf{X}\underline{\phi})\right\} \tag{3}$$

with $\mathbf{V} = (\mathbf{G}\mathbf{G}')^{-1}$ and where $\underline{W}$ represents all the data.

Notice that in (3) $\underline{\theta}$ enters only through the inverse of the covariance matrix $\mathbf{V}$.

3 BAYESIAN ANALYSIS

We need a prior distribution for the parameters $\underline{\phi}, \mathbf{V}$ and σ. We could try to use infromative, conjugate prior distributions. Instead, we use non-informative priors, assuming a-priori that $\underline{\phi}, \mathbf{V}$ and σ are independent so that $f(\underline{\phi}) \propto$ *const*, $f(\sigma) \propto \sigma^{-1}$, Jeffrey's priors, and $f(\underline{\theta}) \propto$ *const*. Hence $f(\underline{\phi}, \underline{\theta}, \sigma) \propto (\sigma)^{-1}$, and the posterior distribution will be:

$$f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W}) \propto (\sigma)^{-(T+1)} |\mathbf{V}|^{1/2}\, exp\left\{-\frac{1}{2\sigma^2}(\underline{\phi} - \hat{\underline{\phi}})'\Sigma_G^{-1}(\underline{\phi} - \hat{\underline{\phi}}) + \nu_G S_G^2\right\}, \tag{4}$$

where:

$$\begin{aligned} \hat{\underline{\phi}} &= (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1} (\mathbf{X}'\mathbf{V}\mathbf{W}), \\ \Sigma_G^{-1} &= (\mathbf{X}'\mathbf{V}\mathbf{X}) \end{aligned}$$

and

$$\nu_G S_G^2 = (\mathbf{W} - \mathbf{X}\hat{\underline{\phi}})'\mathbf{V}(\mathbf{W} - \mathbf{X}\hat{\underline{\phi}}).$$

With this form of the posterior distribution a Gibbs sampler can be used to obtain samples from the overall posterior distribution. The conditional distributions can be obtained easily and thus we will obtain a random sample from the joint posterior of $\underline{\phi}, \sigma$ and $\mathbf{V}$. We can find the values of $\underline{\theta}$ easily from the matrix $\mathbf{V}$. As we know $\mathbf{V} = (\mathbf{G}\mathbf{G}')^{-1}$, thus, calculating the Cholesky decomposition of $\mathbf{V}^{-1}$ we find $\mathbf{G}$ and immediately the values of the vector $\underline{\theta}$.

3.1 UNCONSTRAINED FORECASTS

In this subsection we obtain Bayesian forecasts for k periods ahead. Let $\underline{W}_f$ be the forecasts vector

$$\underline{W}_f = \left(\begin{array}{ccccc} W_{T+1}, & W_{T+2}, & W_{T+3}, & ...., & W_{T+k} \end{array} \right)',$$

thus, the predictive distribution for $\underline{W}_f$ will be

$$f(\ \underline{W}_f \mid \underline{W}\ ) = \int \cdots \int\ f(\ \underline{W}_f \mid \underline{\phi}, \underline{\theta}, \sigma, \underline{W}\ \ ) f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W})\ d\underline{\phi} d\underline{\theta} d\sigma. \tag{5}$$

This predictive distribution can be computed with samples generated from the overall posterior distribution via Gibbs sampling. That is, we will generate samples of $f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W})$ in order to calculate (5) using Monte Carlo integration.

Furthermore, writing the forecasts in a special way we can find $f(\ \underline{W}_f \mid \underline{\phi}, \underline{\theta}, \sigma, \underline{W})$. We can write the future values as

$$W_{T+k} - \phi_1 W_{T+k-1} - \phi_2 W_{T+k-2} - ... - \phi_p W_{T-(p-k)} = \epsilon_{T+k} - \theta_1 \epsilon_{T+k-1} - \theta_2 \epsilon_{T+k-2} - ... - \theta_q \epsilon_{T-(q-k)}.$$

Let

$$\begin{aligned} \underline{W}_p &= \left(\begin{array}{cccc} W_{T-(p-1)}, & W_{T-(p-2)}, & \ldots, & W_T \end{array} \right)' \text{ the last } p \text{ observations,} \\ \hat{\underline{\epsilon}}_q &= \left(\begin{array}{cccc} \hat{\epsilon}_{T-(q-1)}, & \hat{\epsilon}_{T-(q-2)}, & ..., & \hat{\epsilon}_T \end{array} \right)' \text{ the last } q \text{ estimated errors,} \\ \underline{\epsilon}_f &= \left(\begin{array}{cccc} \epsilon_{N+1}, & \epsilon_{N+2}, & ..., & \epsilon_{N+k} \end{array} \right)' \text{ the forecast errors,} \end{aligned}$$

where $\underline{\epsilon}_f \sim \mathbf{N}_k(\underline{0}, \sigma^2 I_k)$, has a k-variate Normal distribution and $\underline{\epsilon}_q \sim \mathbf{N}_k(\underline{0}, \sigma^2 I_q)$. If we define the following matrices

$$A_{k\times k} = \left[\begin{array}{ccccc} 1 & 0 & 0 & ... & 0 \\ -\phi_1 & 1 & 0 & ... & 0 \\ -\phi_2 & -\phi_1 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ -\phi_{k-1} & -\phi_{k-2} & -\phi_{k-3} & ... & 1 \end{array} \right], \qquad C_{k\times k} = \left[\begin{array}{ccccc} 1 & 0 & 0 & ... & 0 \\ -\theta_1 & 1 & 0 & ... & 0 \\ -\theta_2 & -\theta_1 & 1 & ... & 0 \\ ... & ... & ... & ... & ... \\ -\theta_{k-1} & -\theta_{k-2} & -\theta_{k-3} & ... & 1 \end{array} \right],$$

$$B_{k\times p} = \begin{bmatrix} -\phi_p & -\phi_{p-1} & -\phi_{p-2} & \dots & -\phi_1 \\ 0 & -\phi_p & -\phi_{p-1} & \dots & -\phi_2 \\ 0 & 0 & -\phi_p & \dots & -\phi_3 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\phi_k \end{bmatrix} \qquad D_{k\times q} = \begin{bmatrix} -\theta_q & -\theta_{q-1} & -\theta_{q-2} & \dots & -\theta_1 \\ 0 & -\theta_q & -\theta_{q-1} & \dots & -\theta_2 \\ 0 & 0 & -\theta_q & \dots & -\theta_3 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\theta_k \end{bmatrix}$$

we can write the forecasts as:

$$A\underline{W}_f + B\underline{W}_p = C\underline{\epsilon}_f + D\underline{\hat{\epsilon}}_q,$$

or

$$\underline{W}_f = -A^{-1}B\underline{W}_p + A^{-1}C\underline{\epsilon}_f + A^{-1}D\underline{\hat{\epsilon}}_q.$$

It is important to point out that writing the future values or the ARMA(p,q) model in this way has several advantages. First, it is relatively easy to compute any number of future realizations from an ARMA(p,q) model using Monte Carlo integration combined with Gibbs sampling. Second, it is easy to write the predictive distribution for any number of future values conditional on the model parameters. Third, assuming a quadratic loss function, the mean of a specific conditional predictive distribution yields the 'constrained' forecasts. Thus, given past observations up to time T, $\underline{W}$, and all the parameters $\underline{\Theta} = \left\{\underline{\phi}, \underline{\theta}, \sigma\right\}$, we have

$$\underline{W}_f \mid \underline{\phi}, \underline{\theta}, \sigma, \underline{W} \sim \mathbf{N}_k\left(-A^{-1}B\underline{W}_p + A^{-1}D\underline{\hat{\epsilon}}_q, \sigma^2(A^{-1}C)(A^{-1}C)'\right), \tag{6}$$

with $\underline{\mu}_f = -A^{-1}B\underline{W}_p + A^{-1}D\underline{\hat{\epsilon}}_q$ and $\Sigma_f = \sigma^2(A^{-1}C)(A^{-1}C)'$.

Finally, the predictive function (5) is obtained numerically from this distribution and the samples generated via Gibbs sampling. If we assume a quadratic loss function, the Bayes estimator will be the mean of the predictive distribution . That is,

$$E(\underline{W}_f \mid \underline{W}) = \int \cdots \int \underline{\mu}_f f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W})\, d\underline{\phi} d\underline{\theta} d\sigma,$$

and the variance is computed as:

$$Var(\underline{W}_f \mid \underline{W}) = E_{\underline{\phi},\underline{\theta},\sigma}\left(Var(\underline{W}_f \mid \underline{W}, \underline{\phi}, \underline{\theta}, \sigma)\right) + Var_{\underline{\phi},\underline{\theta},\sigma}\left(E(\underline{W}_f \mid \underline{W}, \underline{\phi}, \underline{\theta}, \sigma)\right),$$

where

$$E_{\underline{\phi},\underline{\theta},\sigma}\left(Var(\underline{W}_f \mid \underline{W}, \underline{\phi}, \underline{\theta}, \sigma)\right) = \int \cdots \int Var(\underline{W}_f \mid \underline{W}, \underline{\phi}, \underline{\theta}, \sigma) f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W}) d\underline{\phi} d\underline{\theta} d\sigma$$

$$= \int \cdots \int \Sigma_f f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W}) d\underline{\phi} d\underline{\theta} d\sigma.$$

This expression can be computed numerically with Monte Carlo methods, de Alba (1993).

If we are interested in a linear combination of the forecasts, $W_c = \underline{i}'\underline{W}_f$, where $\underline{i}$ is a constant vector. We know from the properties of the multivariate Normal distribution that W_c has a univariate Normal with $E(W_c \mid \underline{W}, \underline{\phi}, \underline{\theta}, \sigma) = \underline{i}'\underline{\mu}_f$ and $Var(W_c \mid \underline{W}, \underline{\phi}, \underline{\theta}, \sigma) = \underline{i}'\Sigma_f\underline{i}$. For instance, if we have quarterly data and we want to estimate the future value of the sum of the forecasts for the next year, say $z = \underline{i}'\underline{W}_f$, i.e. $k = 4$, we have to select the constant vector as $\underline{i}' = (1, 1, 1, 1)$. The adequate selection of $\underline{i}$ allows to forecast several linear combinations such as weighted average of the forecasts. Using $\underline{i}' = (0, 0, 0, 1)$ then z would be the last value of the series in the forecasts, which would be useful in forecasting stock variables.

Define the yearly total in the (n+1)-st year as z above. Then, from (6) and properties of the Normal distribution we have $f(z \mid \underline{W}, \underline{\phi}, \underline{\theta}, \sigma) = N(\underline{i}'\underline{\mu}_f, \underline{i}'\Sigma_f\underline{i})$. Hence the predictive density for z is

$$f(z \mid \underline{W}) = \int \cdots \int f(z \mid \underline{W}, \underline{\phi}, \underline{\theta}, \sigma) f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W}) d\underline{\phi} d\underline{\theta} d\sigma. \tag{7}$$

This density can be evaluated numerically for each value of z. We use it to analyze how much probability mass is concentrated in the neighborhood of certain given values. Analysis of the density allows an assesment of how consistent are specific values of z with the ARMA(p,q) model being used to forecast, de Alba(1993).

3.2 CONSTRAINED BAYESIAN FORECASTS

From the results in the previous section it is possible to obtain constrained forecasts. The kind of restrictions which can be imposed on the forecasts are that a linear combination of the forecasts will be equal to a fixed value. This can be done with a good selection of the vector $\underline{i}$. The vector $\underline{i}$ will determine the kind of linear restriction that the forecasts have to satisfy. This is done by conditioning the forecasts $\underline{W}_f$ on the value of W_c , i.e. we obtain the predictive distribution for $\underline{W}_f$ given W_c and the data $\mathbf{W}$:

$$f(\underline{W}_f \mid W_c, \underline{W}) = \int \cdots \int f(\underline{W}_f \mid W_c, \underline{\phi}, \underline{\theta}, \sigma, \underline{W}) f(\underline{\phi}, \underline{\theta}, \sigma \mid W_{\mathbf{c}}, \underline{W}) \, d\underline{\phi} d\underline{\theta} d\sigma.$$

If we assume independence between the parameters and the constraint we will have, $f(\underline{\phi}, \underline{\theta}, \sigma \mid W_c, \underline{W}) = f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W})$, and we can obtain samples from the posterior distribution via Gibbs sampling. On the other hand, in order to find $f(\underline{W}_f \mid W_c, \underline{\phi}, \underline{\theta}, \sigma, \underline{W})$, let

$$\underline{W}^* = \begin{pmatrix} \underline{W}_f \\ W_c \end{pmatrix} = \begin{pmatrix} I_k \\ \underline{i} \end{pmatrix} \underline{W}_f,$$

so that $\underline{W}^* \mid \underline{\phi}, \underline{\theta}, \sigma, \mathbf{W} \sim N_{k+1}(\underline{\mu}^*, \not\Sigma^*)$, a **singular multivariate normal distribution**. It is useful to make a decomposition of the covariance matrix of $\underline{W}^*$

$$\Sigma^*_{(k+1)\times(k+1)} = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$$

where $\Sigma_{11_{(1\times1)}} = \Sigma_c = \underline{i}'\Sigma_f\underline{i}$, $\Sigma_{12_{(1\times k)}} = \Sigma_{21}'$, $\Sigma_{22_{(k\times k)}} = \Sigma_f$ and

$$\Sigma_{21_{(k\times1)}} \;=\; Cov(\underline{W}_f\ , W_c) = Var(\underline{W}_f)\ \underline{i} = \Sigma_f\ \underline{i} \quad ;$$

thus,

$$\underline{W}_f \ \mid W_c, \underline{\phi}, \underline{\theta}, \sigma, \underline{W} \ \sim N_k(\underline{\mu}_R, \Sigma_R),$$

where the mean is

$$\underline{\mu}_R = \underline{\mu}_f \ + \Sigma_f\ \underline{i}\ (\ \underline{i}'\Sigma_f\ \underline{i})^{-1}(W_c - \mu_c)$$

and the covariance matrix

$$\Sigma_R = \Sigma_f - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12}.$$

Finally, assuming a quadratic loss function, the optimum constrained forecast vector is given by the conditional predictive mean, that is,

$$E(\underline{W}_f \mid W_c, \underline{W}\ ) = \int\cdots\int \underline{\mu}_R\ \ f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W})\ d\underline{\phi}d\underline{\theta}d\sigma. \tag{8}$$

This mean and variance can be obtained numerically with Monte Carlo methods. The variance can be computed using the following expression,

$$Var(\underline{W}_f \mid W_c, \underline{W}\ ) = E_{\underline{\phi},\underline{\theta},\sigma}\left(Var(\underline{W}_f \mid W_c, \underline{\phi}, \underline{\theta}, \sigma, \underline{W}\ )\right) + Var_{\underline{\phi},\underline{\theta},\sigma}\left(E(\underline{W}_f \mid W_c, \underline{\phi}, \underline{\theta}, \sigma, \underline{W}\ )\right), \tag{9}$$

where $E_{\underline{\phi},\underline{\theta},\sigma}$ and $Var_{\underline{\phi},\underline{\theta},\sigma}$ represent expectation and variance unde the posterior distribution of $\underline{\phi}, \underline{\theta}, \sigma$.

Whatever the source of information on the value of z, whether it is expert opinion or any other, it may yield a value that seem 'unreasonable'. This would raise doubts about the expert's ability to produce good forecasts, or about the adequacy of the model. It may also indicate that there has been a structural change in the process under study. We thus need a criterion for evaluating the consistency of given values of z with the model.

The predictive distribution of the aggregate values can be used to carry out such an analysis from the Bayesian point of view. In general the source fo external information about z will not be the model in (1). Hence assuming it is the correct model some values of the future yearly figure z are very unlikely and will appear to be completely inconsistent with it. We consider a given value z_0 of z to be 'unlikely' if it has a small P-value, i.e. if $Pr\left\{z - E(z \mid \underline{W}) > |z_0 - E(z \mid \underline{W})\ | |\underline{W}\right\} \longrightarrow 0$ under the predictive density in (7), which must be evaluated numerically for each value of z. See the Appendix for more detail.

4 FORECASTS USING GIBBS SAMPLER

As mentioned, the purpose of this paper is to derive methods to compute constrained forecasts, using Bayesian techniques. In order to do that, we need to generate samples from either the posterior distribution or the predictive distribution. By using Markov Chain Monte Carlo methods we can avoid any analytical integration by generating samples from the full conditionals and obtain samples from the posterior distributions. In turn, these can be used to compute moments of the predictive distribution. In this case it is not possible to sample from the exact posterior distribution directly. Over the past decade most work on sampling methods for exploring posterior distributions has centered on importance sampling, Geweke (1989), Stewart (1979) and Zellner and Rossi (1984), among others. More recently, results of Gelfand and Smith (1990) on the Gibbs sampler have rekindled interest in the use of dependent samples generated using Markov Chains with equilibrium distribution π. Gelfand and Smith extend the Gibbs sampling algorithm of Geman and Geman (1984), originally developed for Bayesian image reconstruction, to continuous distributions and show how the algorithm can be used in a wide variety of problems. We apply Gibbs sampling to obtain constrained and unconstrained forecasts in ARMA(p,q) models.

As we have seen before in (4), the overall posterior distribution is

$$f(\underline{\phi}, \underline{\theta}, \sigma \mid \underline{W}) \propto (\sigma)^{-(T+1)} |\mathbf{V}|^{1/2} \, exp\left\{-\frac{1}{2\sigma^2}(\underline{\phi} - \underline{\hat{\phi}})'\Sigma_G^{-1}(\underline{\phi} - \underline{\hat{\phi}}) + \nu_G S_G^2\right\}$$

where

$$\begin{aligned} \underline{\hat{\phi}} &= (\mathbf{X}'\mathbf{V}\mathbf{X})^{-1}(\mathbf{X}'\mathbf{V}\mathbf{W}), \\ \Sigma_G^{-1} &= (\mathbf{X}'\mathbf{V}\mathbf{X}) \end{aligned}$$

and

$$\nu_G S_G^2 = (\mathbf{W} - \mathbf{X}\underline{\hat{\phi}})'\mathbf{V}(\mathbf{W} - \mathbf{X}\underline{\hat{\phi}}).$$

The full conditionals for the parameters follow immediately from the overall posterior distribution as follows. They are the following: $\underline{\phi} \mid \underline{\theta}, \sigma, \underline{W} \sim \mathbf{N}_p(\underline{\hat{\phi}}, \sigma^2\Sigma_G)$; $\sigma \mid \underline{\phi}, \underline{\theta}, \underline{W} \sim$ Inverse-Gamma(ν_1, S_1^2) with $\nu_1 = T$ and $S_1^2 = (\mathbf{W} - \mathbf{X}\underline{\phi})'\mathbf{V}(\mathbf{W} - \mathbf{X}\underline{\phi})$; and, using the fact that $\underline{\theta}$ enters only through the matrix $\mathbf{V}$, $\mathbf{V} \mid \underline{\phi}, \sigma, \underline{W} \sim$Wishart (Σ, ν_2, q), Zellner (1971), where $\Sigma = \sigma^2\left[(\mathbf{W} - \mathbf{X}\underline{\phi})(\mathbf{W} - \mathbf{X}\underline{\phi})'\right]^{-1}$, $\nu_2 = q+2$ and q is the order of the MA component of the model.

Thus, a Gibbs sampler will be used to obtain samples from the overall posterior distribution. Note that using these conditional distributions we will obtain a random sample from the joint posterior of $\mathbf{V}$,$\underline{\phi}$ and σ . However, we can find the values of $\underline{\theta}$ easily from the matrix $\mathbf{V}$. As we know $\mathbf{V} = (\mathbf{G}\mathbf{G}')^{-1}$, thus, calculating the Cholesky decomposition of $\mathbf{V}^{-1}$ we find $\mathbf{G}$ and immediately the values of the vector $\underline{\theta}$. Generation of random samples is now straightforward.

5 AN APPLICATION

In this section we applied the theoretical results to real data. The series used was y_t,the Mexican inflation during 1969-1976, Guerrero (1991b). The series was generated as the first difference of the natural logarithms of the National Consumer Price Index, (Indice Nacional

	BOX & JENKINS		BAYESIAN	
	PARAMETER	ST. DEV.	PARAMETER	ST. DEV.
CONST	0.0082		0.0029	
AR(1)	0.8380	0.0887	0.7982	0.1264
MA(1)	0.3985	0.1484	0.4338	0.1739

Table 1: Parameters

de Precios al Consumidor), IPC_t . This series, $y_t = \log(IPC_t) - \log(IPC_{t-1})$, was identified as an ARMA(1,1) model. There are N=91 observations, from February 1969 to august 1976. We were interested in forecasts the last four months of 1976, because we had some information about the annual value. The experts had indicated that the total value of Inflation this last year would be 24.1%. Since the time series was available uo to August, then the restriction we imposed was 15.7% as the inflation from august to December 1976. Note that,

$$\log(IPC_{dic76}) - \log(IPC_{dic75}) =$$
$$[(\log(IPC_{dic76}) - \log(IPC_{ago76})) + (\log(IPC_{ago76}) - \log(IPC_{dic75}))] \,,$$

but $\log(IPC_{dic76}) - \log(IPC_{ago76}) = y_{92} + y_{93} + y_{94} + y_{95}$ and $\log(IPC_{ago76}) - \log(IPC_{dic75}) = \sum_{t=84}^{91} y_t$. Furthermore,$(\log(y_{dic76}) - \log(y_{ago76})) = 0.241 - 0.084 = 0.157$. Thus, the constraint vector to use in the restriction was $\underline{\iota}' = (1,1,1,1)$.

In Table 1 we present the parameters estimations and standard deviations using Box-Jenkins methodology and Bayesian approach. Note that the estimated variances obtained with the Bayesian method are bigger than the others. This is due to the fact that the Bayesian methodology incorporates the posterior variation of the parameters. Table 2 presents the forecasts obtained with both methods. Bayesian forecasts include the values calculated without restrictions and with the constraint mentioned before. Note that during August of this year a peso devaluation occurred, thus inflation was affected by this structural change. If we had estimated the model with the data until August of 1976, we would have obtained the forecasts without constraints shown in the table. On the other hand, if we had included the information about the beliefs on the inflation between august and December, we would have obtained the constrained forecasts presented. As we can see in figure 1, the constrained forecasts are very close to the real values.

The constrained forecasts are computed for the following annual values of inflation: z = 0.157 and z = 0.028. The first value of z is the one actually observed for 1976. Since there was a dramatic change in the series, a structural change, this value was very 'unlikely' to observe such a large value with the model that was true up to then. It is in one of the tails of the predictive distribution of z based on observations up to 1975-IV and its P-value computed according to the formulas in the Appendix is 2.760185e-05. The second value of z would be obtained if prediction was done without constraints . It is near the mode of the predictive distribution. Its P-value is 0.504513. Figure 2 shows the corresponding predictive distribution.

This method is very useful to generate Bayesian constrained forecasts. The principal advantage is that the forecasts are derived easily from the predictive distribution. Moreover,

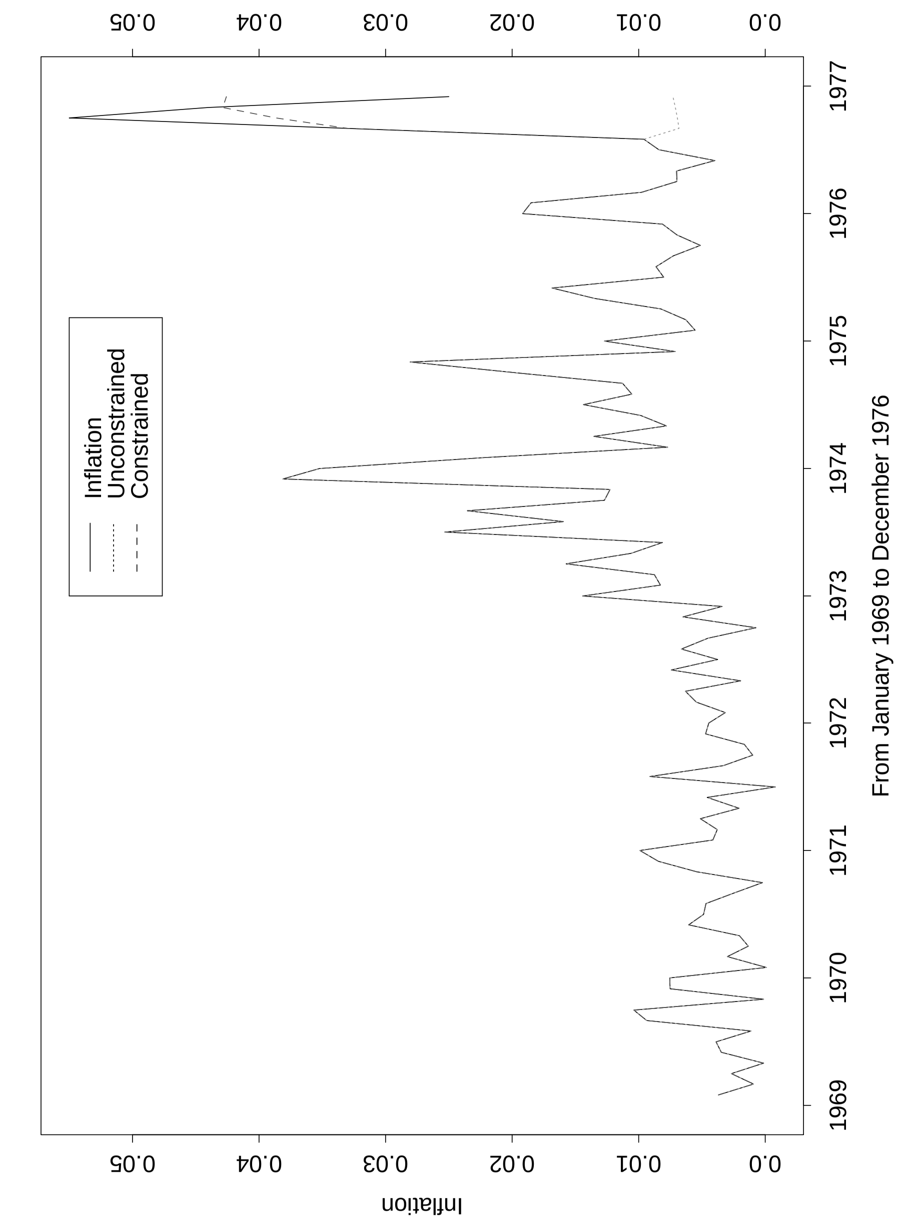

Figure 1: Bayesian Forecasts in an ARMA(1,1) model

		BOX & JENKINS		BAYESIAN			
				UNCONST		CONST	
Period	Observed Value	Puntual Forecasts	Standard Deviation	Puntual Forecasts	Standard Deviation	Puntual Forecasts	Standard Deviation
1976-9	0.033	0.008517	0.005700	0.006791	0.006918	0.032976	0.005178
1976-10	0.055	0.008459	0.006300	0.006915	0.007245	0.038647	0.004793
1976-11	0.044	0.008410	0.006600	0.007087	0.007398	0.042859	0.004895
1976-12	0.025	0.008370	0.006800	0.007296	0.007531	0.042580	0.005128
SUM	0.157	0.033756		0.028089		0.1570620	

Table 2: Forecasts

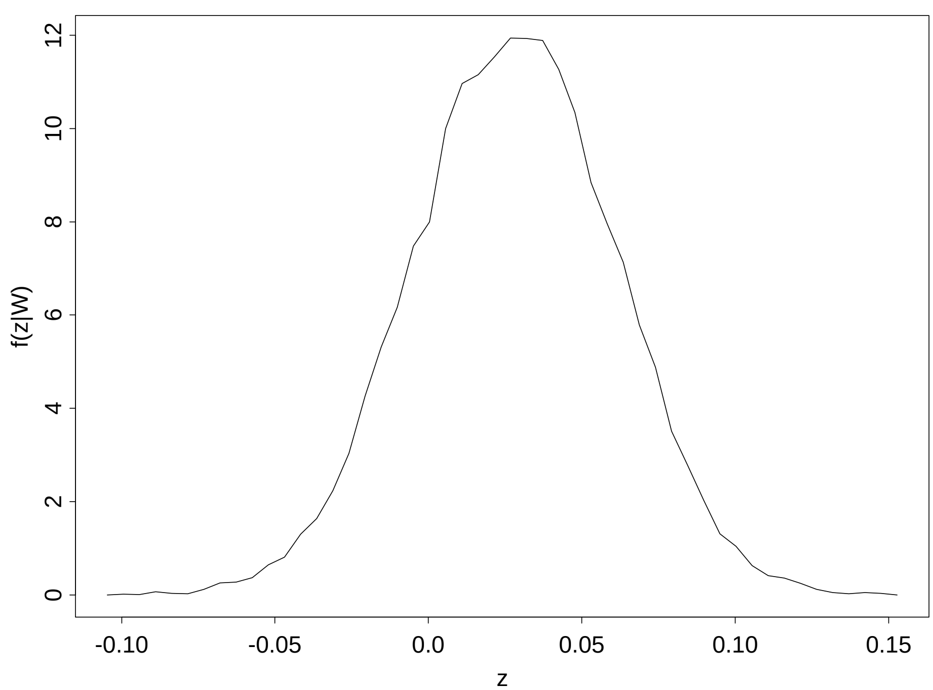

Figure 2: Predictive distribution $f(z|W)$

it is not necessary to generate too many replications for the Monte Carlo integration. Using about 1,000 iterations the results are stable, that is, the results do not change when the iterations are increased to 10,000. To generate from the posterior distribtuion in every case a Gibbs sampler is run for 1,000 iterations using the first 150 as a burn-in period and taking the average of such 150 samples to initialize the parameters. All the calculations were made in S+.

6 APPENDIX. NUMERICAL CALCULATIONS

In this appendix we are going to show the way to calculate all the main results from the previous sections. Let $\Theta = (\underline{\phi}, \sigma, \mathbf{V})$, then the numerical calculations can be carried out as follows.

6.1 MEANS

For the forecasts without restrictions we need,

$$E(\ \underline{W}_f \mid \underline{W}\ )\ = \int \cdots \int \underline{\mu}_f\ \ f(\ \underline{\Theta} \mid \underline{W}\ )\ \ d\underline{\Theta}.$$

where samples from $f(\ \underline{\Theta} \mid \underline{W})$ are obtained via Gibbs sampling. We can do Monte Carlo integration directly using these samples . Define $g(\underline{\Theta}) = \underline{\mu}_f = -A^{-1}B\underline{W}_p + A^{-1}D\underline{\hat{\epsilon}}_q$ and generate a sequence of M random vectors from $f(\ \underline{\Theta} \mid \underline{W})$, to obtain $\underline{\Theta}_i$ with $i = 1, 2, 3...M$. Evaluate the function $g(\underline{\Theta}_i)\ \forall\ i = 1, 2, 3...M$ and calculate

$$E(\ \underline{W}_f \mid \underline{W}\ ) \cong \frac{1}{M}\sum_{i=1}^{M} g(\underline{\Theta}_i).$$

For the forecasts with restrictions use the same process, just define $g(\underline{\Theta}) = \underline{\mu}_R = \underline{\mu}_f + \Sigma_f \underline{i}(\underline{i}'\Sigma_f\underline{i})^{-1}(W_c - \mu_c)$ and use (8) to calculate

$$E(\ \underline{W}_f \mid W_c, \underline{W}\ ) \cong \frac{1}{M}\sum_{i=1}^{M} g(\underline{\Theta}_i).$$

6.2 VARIANCES

Forecasts without restrictions, the predictive distribution variance is given by

$$Var(\underline{W}_f \mid \underline{W}\ ) = E_{\underline{\Theta}}\left(Var(\underline{W}_f \mid \underline{W}\ , \underline{\Theta})\right) + Var_{\underline{\Theta}}\left(E(\underline{W}_f \mid \underline{W}\ , \underline{\Theta})\right).$$

The first part is calculated using the same algorithm as the means. That is, calculate numerically:

$$E_{\underline{\Theta}}\left(Var(\underline{W}_f \mid \underline{W}\ , \underline{\Theta})\right) = \int \cdots \int Var(\underline{W}_f \mid \underline{W}\ , \underline{\Theta})\ f(\ \underline{\Theta} \mid \underline{W}\ )\ d\underline{\Theta}$$

where $g(\underline{\Theta}) = Var(\underline{W}_f \mid \underline{W}\ , \underline{\Theta}) = \Sigma_f$, to obtain

$$E_{\underline{\Theta}}\left(Var(\underline{W}_f \mid \underline{W}\ , \underline{\Theta})\right) \cong \frac{1}{M}\sum_{i=1}^{M} g(\underline{\Theta}_i).$$

To calculate the second part, define $U(\underline{\Theta}) = E(\underline{W}_f \mid \underline{W}\ , \underline{\Theta})$ so that the variance estimation will be:

$$\widehat{Var}_{\underline{\Theta}}\left(E(\underline{W}_f \mid \underline{W}\ , \underline{\Theta})\right) = \frac{1}{M-1}\sum_{i=1}^{M}\left(U(\underline{\Theta}_i) - \overline{U}(\underline{\Theta})\right)^2$$

with $\overline{U}(\underline{\Theta}) = \frac{1}{M}\sum_{i=1}^{M} U(\underline{\Theta}_i)$.

The variance for the forecasts with restrictions will be calculated the same way using (9). For the first part define $g(\underline{\Theta}) = Var(\underline{W}_f \mid W_c, \underline{\Theta}, \underline{W}\ ) = \not\Sigma_R$, and compute

$$E_{\underline{\Theta}}\left(Var(\underline{W}_f \mid W_c, \underline{\Theta}, \underline{W}\ )\right) \cong \frac{1}{M}\sum_{i=1}^{M} g(\underline{\Theta}_i).$$

For the second part define $U(\underline{\Theta}) = E(\underline{W}_f \mid W_c, \underline{\Theta}, \underline{W}\ )$ and calculate

$$\widehat{Var}_{\underline{\Theta}}\left(E(\underline{W}_f \mid \underline{W}\ , \underline{\Theta})\right) = \frac{1}{M-1}\sum_{i=1}^{M}\left(U(\underline{\Theta}_i) - \overline{U}(\underline{\Theta})\right)^2$$

with $\overline{U}(\underline{\Theta}) = \frac{1}{M}\sum_{i=1}^{M} U(\underline{\Theta}_i)$.

6.3 P-VALUES

The P-value $\Pr\{z - E(z \mid \underline{W}) > |z_0 - E(z \mid \underline{W})\ | |\underline{W}\}$ can be computed under the predictive density of z as follows. Consider the case $z_0 > E(z \mid \underline{W})$ and let $P(z_0 \mid \underline{W}) = \Pr(z > z_0 \mid \underline{W})$. This can be computed as

$$P(z_0 \mid \underline{W}) = \int \cdots \int \Pr(z > z_0 \mid \underline{\Theta}, \underline{W})\ f(\ \underline{\Theta} \mid \underline{W}\ )\ \ d\underline{\Theta}. \tag{10}$$

But since $f(z \mid \underline{\Theta}, \underline{W})$ is a univariate Normal distribution then $\Pr(z > z_0 \mid \underline{\Theta}, \underline{W}) =$ 1-$\Phi(z_0)$,where $\Phi(z)$ is its *cdf.* Thus (10) can be evaluated numerically by Monte Carlo integration.

References

[1] Broemeling, L.D. and M. Land (1984), "On Forecasting with Univariate Autoregressive Processes: A Bayesian Approach," *Communications in Statistics-Theory and Methods*, 13(11), 1305-1320.

[2] Broemeling, L.D. and S. Shaarawy (1986), "A Bayesian Analysis of Time Series," in: *Bayesian Inference and Decision Techniques*, P. Goel and A. Zellner (eds.), 337-354, Amsterdam: Elsevier Science Pub.

[3] Broemeling, L.D. and S. Shaarawy (1988), "Time Series: A Bayesian Analysis in the Time Domain," in: *Bayesian Analysis of Time Series and Dynamic Models*, J.C. Spall (ed.), 1-22, New York: Marcel Dekker.

[4] Cholette, P.A. (1982), "Prior Information and ARIMA Forecasting," *Journal of Forecasting*, 1,375-383.

[5] Chow, G.C. (1975), "Multiperiod Predictions From Stochastic Difference Equations by Bayesian Methods," in: *Studies in Bayesian Econometrics and Statistics*, S.E. Fienberg and A. Zellner (eds.), Amsterdam: North-Holland.

[6] de Alba, E. (1988), "Disaggregation and Forecasting: A Bayesian Analysis," *Journal of Business and Economic Statistics*, 6, 197-206.

[7] de Alba, E. (1992), "Constrained Forecasting in Regression: A Bayesian Analysis," *ESTADISTICA (IASI)*, Vol. 44, No. 142-143, 27-72.

[8] de Alba, E. (1993), "Constrained Forecasting in Autoregressive Time Series Models: Bayesian Analysis," *International Journal of Forecasting* 9, 95-108.

[9] de Alba, E. and O. Aguilar-Chavez (1995), "Pronosticos Bayesianos con Restricciones en Modelos ARMA", *Revista de Estadistica (INEGI)*, Mexico, forthcoming.

[10] De Groot, M.H. (1970), *Optimal Statistical Decisions*, New York: McGraw-Hill.

[11] Gelfand, A.E. and Smith, A.F.M. (1990), "Sampling based approaches to calculating marginal densities", *Journal of the American Statistical Association* 85, 398-409.

[12] Geman,S. and Geman, D. (1984), "Stochastic Relaxation, Gibbs distributions and the Bayesian restoration of images", *IEEE Trans. Patt. Anal. Mach. Intelligence* 6, 721-740

[13] Geweke, J. (1989), "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, Vol. 57, No. 6, 1317-1340.

[14] Guerrero, V.M. (1989), " Optimal Conditional ARIMA Forecasts," *Journal of Forecasting*, 8, 215-229.

[15] Guerrero, V.M., (1990) "Restricted ARIMA Forecasts which account for parameter changes", *ESTADISTICA* 42, 139, pp. 17-31.

[16] Guerrero, V.M., (1991a) "ARIMA forecasts with restrictions derived from a structural change", *International Journal of Forecasting* 7, 339-347.

[17] Guerrero, V.M. (1991b) *Análisis Estadístico de Series de Tiempo Económicas*, UAM Ixtapalapa, México

[18] Guerrero, V.M., (1993) "Combining historical and preliminary information to obtain timely time series data", *International Journal of Forecasting* 9, 477-485.

[19] Marriott, J.; Ravishanker, N.; Gelfand, A. and Pai, J., (1993) "Bayesian Analysis of ARMA Processes: Complete Sampling Based Inference Under Exact Likelihoods", en: *Bayesian Statistics and Econometrics: Essays in Honor of Arnold Zellner*, D. Barry et. al (eds.), Amsterdam: North-Holland.

[20] Monahan, J.F. (1983), "Fully Bayesian Analysis of ARMA Time Series Models," *Journal of Econometrics*, Vol. 21, 307-331.

[21] McCulloch, R.E. and Tsay, R.S. (1994), "Bayesian Analysis of Autoregressive Time Series Via the Gibbs Sampler," *Journal of Time Series Analysis*, 15, 235- 250.

[22] Naylor, J.C. and Marriott, J.M., (1994) "A Bayesian Analysis of Non-stationary AR Series", Research Report Department of Mathematics, Statistics and Operational Research, Number 11/94

[23] Newbold, P. (1973), "Bayesian Estimation of Box-Jenkins Transfer Function-noise Models," *Journal of the Royal Statistical Society, Series B*, Vol. 35, No. 2, 323- 336.

[24] Shaarawy, S. and L Broemeling (1984), "Bayesian Inferences and Forecasts with Moving Average Processes," *Communications in Statistics-Theory and Methods*, 13(15), 1871-1888.

[25] Shaarawy, S. and L Broemeling (1985), "Inference and Prediction with ARMA Processes," *Communications in Statistics-Theory and Methods*, 14(10), 2531- 2548.

[26] Stewart, L. (1979), "Multiparameter univariate Bayesian analysis", *Journal of the American Statistical Association* 74, 684-693.

[27] Thompson, P.A. and R.B. Miller (1986), "Sampling the Future: A Bayesian approach to Forecasting from Univariate Time Series Models," *Journal of Business and Economic Statistics*, Vol. 4, No. 4, 427-436.

[28] West, M. and J. Harrison (1989), *Bayesian Forecasting and Dynamic Models*, New York: Springer-Verlag.

[29] Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley and Sons, Inc.

[30] Zellner, A. and Rossi, P.E. (1984), " Bayesian analysis of dichotomous quantal response models", *Journal of Econometrics* 25,365-393.