A Bayesian Analysis of Autoregressive Time Series Panel Data

Balgobin NANDRAM and Joseph D. PETRUCCELLI
Department of Mathematical Sciences, Worcester Polytechnic Institute, Worcester, MA 01609-2280
(balnan@wpi.edu)

We describe a Bayesian hierarchical model to analyze autoregressive time series panel data. We develop two algorithms using Markov-chain Monte Carlo methods, a restricted algorithm that enforces stationarity or nonstationarity conditions on the series and an unrestricted algorithm that does not. Two examples show that restricting stationary series to be stationary provides no new information, but restricting nonstationary series to be stationary leads to substantial differences from the unrestricted case. These examples and a simulation study also show that, compared with inference based on individual series, there are gains in precision for estimation and forecasting when similar series are pooled.

KEY WORDS: Forecasting; Hierarchical model; Latent variable; Markov-chain Monte Carlo algorithm; Stationarity.

We consider the problem of parameter estimation and forecasting for autoregressive time series panel data. These data consist of several time series generated by the same type of autoregressive (AR) model [e.g., an AR(p)]. The key advantage of simultaneously modeling several series is the possibility of pooling information from all series. This can not only improve estimation and forecasting performance but also allows analysis of much shorter series (e.g., economic time series) than it would be possible to model effectively as single series. In such situations the Bayesian paradigm is particularly attractive because it offers a natural scheme for combining and weighting data from several similar sources.

We adapt the hierarchical Bayesian normal linear model (Lindley and Smith 1972) to permit borrowing of strength over all series. The pooling takes place as the autoregressive parameters of the series are assumed to arise from the same distribution. Our model is very flexible in that it can accommodate restrictions on the autoregressive parameters of the series. In the sequel we focus on the restrictions of most general interest—namely, restricting the series to be stationary and restricting them to be nonstationary. We use the same model for both the restricted and unrestricted series except that for the restricted series the parameters are constrained to lie in the proper region. One difficulty with implementing our approach is that the posterior distributions do not exist in closed forms.

Sampling-based approaches have been used successfully to perform integrations in situations in which the posterior distributions are not analytically tractable. We use the Markov-chain Monte Carlo (MCMC) sampler; see Tierney (1994) for a general description. See also Gelfand, Smith, and Lee (1992), who described how to implement the Gibbs sampler (Gelfand and Smith 1990) when there are truncation and order restrictions, and Chib and Greenberg (1994) for an application to a single time series with (autoregressive moving average) ARMA (p, q) errors.

Several approaches to pooling time series have previously been considered. Among those using an empirical Bayes approach were Andrews (1976), Ravishanker, Wu, and Dey (1992), Ledolter and Lee (1993), and Li and Hui (1983). One common concern with this approach is that it tends to overestimate precision.

Bayesian formulations have included those of Pai, Ravishanker, and Gelfand (1993), Chow (1973), who considered multistep forecasting, and Liu and Tiao (1980), who, in the article on panel data closest in spirit to ours, performed a full Bayesian analysis for AR(1) models. Even for this simplest of time series models, however, and even using the approximations Liu and Tiao (1980) were forced to make, the intractability of the computations eliminates any hope of routine application of their methodology. Our approach is much simpler than the one given by Liu and Tiao (1980) because it avoids the use of approximations due to the use of the nonconjugate beta prior.

There is another reason that the Liu and Tiao (1980) approach is extremely intractable—namely, the assumptions on the autoregressive parameters required for stationarity of the series. Most Bayesian approaches to time series analysis explicitly or implicitly assume stationarity of the series but ignore the necessary stationarity restrictions; see, for example, Broemeling (1985). Recently Marriott, Ravishanker, and Gelfand (1992) incorporated these stationarity restrictions using the Gibbs sampler (Gelfand and Smith 1990) to obtain a full Bayesian analysis of a single series. Their approach was to transform the autoregressive coefficients into partial autocorrelations and Fisher transform the partial autocorrelations to normality. This approach to incorporating the stationarity restrictions was also used by Pai et al. (1993). Rather than transform first to partial autocorrelations and then to normality, we incorporate the stationarity restrictions directly into the modeling procedure.

An alternative approach to modeling data such as these is to use Bayesian vector autoregressive (BVAR) models

Journal of Business & Economic Statistics
July 1997, Vol. 15, No. 3

(e.g., see Sims 1980; Litterman 1986; Kadiyala and Karlsson 1993), which capture interseries as well as within-series relations. Although BVAR models have proven useful in some applications, when there are many short series, the large number of parameters involved in the BVAR model forces overreliance on and oversimplification of prior specifications.

In contrast, our methodology is appropriate for many simultaneous commensurate series. Our approach differs from the BVAR approach in three major respects. First, in the BVAR model, dependence between series is built into the sampling process; for our model, it arises solely because of the prior specification and the error structure. Second, the BVAR approach models temporal dependence across series as well as within series; our approach models temporal dependence only within series. Third, our approach allows a more flexible prior specification than does the BVAR and can accommodate any number of series.

Chib and Greenberg (1995) considered hierarchical versions of Zellner's seemingly unrelated regression (SUR) model, one of which includes the model we consider here. They also used the MCMC method to analyze their models, although they did not incorporate restrictions on the autoregressive parameters, which is one of our main contributions. Using Organization for Economic Cooperation and Development gross national product data, Chib and Greenberg (1995) found strong evidence for the pooled, as opposed to the unpooled, model. For our model, we investigate in detail the effects of pooling on estimation and forecasting. One difficulty posed by the greater generality of the SUR model is its inability to handle large numbers of series due to the many parameters in the covariance of the sampling process. Another difficulty with the greater generality of the BVAR and SUR models is the complexity of incorporating stationarity or nonstationarity restrictions.

In Section 1 of the article, we present the methodology. In Section 2, we illustrate the methodology with the analysis of a dataset on yearly averages of hourly earnings of production workers in 14 California metropolitan areas. We also perform a small simulation study to assess the gains in estimation and forecasting due to pooling. Section 3 has concluding remarks.

1. METHODOLOGY

We briefly describe a methodology for modeling and forecasting any number of time series of possibly varying lengths and the associated computations. Details were given by Nandram and Petruccelli (1995).

1.1 The Model

We observe m time series realizations $\{y_{i,t}\}_{t=t_i}^{n}, i = 1,\ldots,m$, possibly of different lengths, with the ith series starting at time t_i and each generated by an autoregressive model of order p. We assume that the minimum of the t_i equals 1 (i.e., the earliest observation is at time 1), that the last observation occurs at the same time, n, for all series, and that there are no missing observations between the first and last observations. We let $n_i = n - t_i + 1$ denote the number of observations in the ith series. We also assume the unobservable vectors $\mathbf{y}_i^{(0)} = (y_{i,t_i-1}, y_{i,t_i-2},\ldots,y_{i,t_i-p})'$, $i = 1,\ldots,m$, and $\mathbf{y}^{(0)} = (\mathbf{y}_1^{(0)'},\ldots,\mathbf{y}_m^{(0)'})'$. For each $1 \le t \le n$, we let $I_t = \{1 \le i \le m : t_i \le t\}$ denote the set of series that have observations at time t and let m_t denote the number of such series.

The defining relation for the ith series, given the parameters $\boldsymbol{\phi}_i, \tau_i, \psi^2$, and $\mathbf{y}_i^{(0)}$, is

$$y_{i,t} = \boldsymbol{\phi}_i'\mathbf{y}_{i,t-1} + \varepsilon_{i,t}, \qquad t \ge t_i, \tag{1}$$

where $\boldsymbol{\phi}_i' = (\phi_{i0}, \tilde{\boldsymbol{\phi}}_i'), \tilde{\boldsymbol{\phi}}_i' = (\phi_{i1},\ldots,\phi_{ip}), \mathbf{y}_{i,t}' = (1, y_{i,t}, y_{i,t-1},\ldots,y_{i,t-p+1})$, and $\varepsilon_{i,t}$ is an error term. Letting ε_t' be the vector whose components are $\{\varepsilon_{i,t}, i \in I_t\}$ and $\boldsymbol{\tau}' = (\tau_1, \tau_2,\ldots,\tau_m)$, we take

$$\varepsilon_t | \boldsymbol{\tau}, \psi^2 \overset{\text{iid}}{\sim} N(\mathbf{0}, \Sigma_t), \tag{2}$$

where $\Sigma_t = \text{diag}\{\tau_i^{-1}, i \in I_t\} + \psi^2 J_t$ and J_t is an $m_t \times m_t$ matrix of ones. The autoregressive parameters are modeled as

$$\boldsymbol{\phi}_i | \boldsymbol{\theta}, \Delta \overset{\text{iid}}{\sim} N(\boldsymbol{\theta}, \Delta). \tag{3}$$

Observe that (3) permits pooling of information across series.

Next we take conjugate priors for $\boldsymbol{\theta}$ and Δ^{-1}. That is, we take a normal prior

$$\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, C_0) \tag{4}$$

and a Wishart prior

$$\Delta^{-1} \sim W((\nu_0\Delta_0)^{-1}, \nu_0), \qquad \nu_0 \ge p + 1, \tag{5}$$

where $\boldsymbol{\theta}_0$ and C_0 in (4) and ν_0 and Δ_0 in (5) are to be specified.

We also take the prior distribution for $\boldsymbol{\tau}$ to be gamma and for ψ^2 to be inverse gamma. Specifically, we assume that

$$\tau_1, \tau_2, \ldots, \tau_m \overset{\text{iid}}{\sim} \text{G}(\eta_0/2, \delta_0/2) \tag{6}$$

and

$$\psi^2 \sim \text{IG}(\alpha_0/2, \beta_0/2), \tag{7}$$

where η_0, δ_0 in (6) and α_0, β_0 in (7) are to be specified.

To complete the specification, we take

$$\mathbf{y}_i^{(0)} \sim N(\mathbf{b}_0, B_0), \tag{8}$$

where $\mathbf{b}_0$ and B_0 are again parameters to be specified. By employing the latent variables $\mathbf{y}_i^{(0)}$, we are able to use all the data because we need not condition on the first p observations in each series. We will refer to the use of latent variables in this way as initialization. Although for stationary series it is possible to obtain the full likelihood, this is not possible when some or all of the series are nonstationary. Initialization allows us to restrict model parameters to lie in any region. Specifically, we can restrict the parameters ϕ_i in the model to be in the stationary region $\boldsymbol{\Phi}_p = \{\boldsymbol{\phi}_i :$ series (1) is stationary$\}$, or in its complement. See, for example, Box and Jenkins (1976) for details on stationarity.

Finally, to carry out the analysis, Nandram and Petruccelli (1995) described two MCMC samplers. Although in the first, or unrestricted, version, the autoregressive parameters ϕ_i are sampled from the normal distribution in (3), in the second, or restricted, version, the sampling is done from Distribution (3) with the ϕ_i restricted to ϕ_p or its complement.

1.2 Computations

Defining $\bar{\phi} = \sum_{i=1}^m \hat{\phi}_i/m$, we take $\nu_0 = p+2$ and $\Delta_0 = S_\phi/\nu_0$, where $S_\phi = \sum_{i=1}^m (\hat{\phi}_i - \bar{\phi})(\hat{\phi}_i - \bar{\phi})'/(m-1)$. We also estimate $\theta_0, C_0, \nu_0, \Delta_0, \eta_0$, and δ_0 using the data. We take α_0 and β_0 to be 0, however, and, therefore, ψ^2 has a noninformative prior. In addition, when initialization is done, using the data we obtain $\mathbf{y}_i^{(0)}$ for the ith series by the usual backcasting method applied to the ith series. Then we take $\mathbf{b}_0 = \sum_{i=1}^m \mathbf{y}_i^{(0)}/m$ and $B_0 = \sum_{i=1}^m (\mathbf{y}_i^{(0)} - \mathbf{b}_0)(\mathbf{y}_i^{(0)} - \mathbf{b}_0)'/(m-1)$ in the prior specification (8).

Let $t_0 = 1$ when initialization is done and $t_0 = p+1$ otherwise. To facilitate computation, we rewrite (2) in terms of latent variables α_t:

$$\varepsilon_{it}|\alpha_t, \tau_i \overset{\text{ind}}{\sim} N(\alpha_t, \tau_i^{-1}) \tag{9}$$

and

$$\alpha_t|\psi^2 \overset{\text{iid}}{\sim} N(0, \psi^2), \tag{10}$$

where $t = t_0, \ldots, n$ and $i \in I_t$. Let $\boldsymbol{\alpha} = (\alpha_{t_0}, \alpha_{t_0+1}, \ldots, \alpha_n)'$ denote the vector of all latent variables in (10).

For each dataset, we run both the restricted and unrestricted versions of the MCMC sampler to obtain M stationary iterates of the ϕ_i, ψ^2, and $\Omega = \{\mathbf{y}^{(0)}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta}, \Delta\}$ when initialization is done ($\Omega = \{\boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta}, \Delta\}$ when initialization is not done), which we denote by $\phi_i^{(j)}, i = 1, \ldots, m, \psi^{2(j)}$ and $\Omega^{(j)}$, respectively, $j = 1, 2, \ldots, M$. Whenever the ϕ_i are restricted, the MCMC algorithm is a Metropolis–Hastings algorithm (see Tierney 1994) in which $\boldsymbol{\theta}$ and Δ are drawn from their posterior conditional distributions using dependent rejection sampling and the ϕ_i are drawn using independent rejection sampling. For the unrestricted problem, a straightforward Gibbs sampler is done. The iterates obtained from the MCMC algorithm are used to perform estimation, forecasting, and model assessment.

Details of the computations, which include the conditional posterior distributions, the Metropolis–Hastings algorithm, estimation and forecasting, and model assessment, were given by Nandram and Petruccelli (1995).

2. EMPIRICAL STUDIES

We apply our methodology to earnings data and 12 examples of simulated data.

2.1 Earnings Data

The data (Liu and Tiao 1980) consist of yearly averages of the hourly earnings of production workers in 14 California metropolitan areas. Each of the 14 series ends in 1977, but the series are of different lengths with the longest beginning in 1945 and the shortest beginning in 1963. These series are of six different lengths—15 (series 3 and 10), 16 (series 8, 11, 13, and 14), 20 (series 4, 6, 7, and 12), 26 (series 5), 27 (series 9), and 33 (series 1 and 2). The natural logarithm of each series serves to stabilize variances. The last observation of each series was set aside to assess prediction performance and the model was fit to the remaining data. Most if not all of the 14 series are nonstationary, but taking a first difference transforms all of them to stationarity. As Liu and Tiao did, we fit an AR(1) to the differenced data. The first stage of the model is

$$z_{i,t} = \phi_{i0} + \phi_{i1} z_{i,t-1} + \varepsilon_{i,t},$$
$$t = t_i + 1, \ldots, 33, \quad i = 1, \ldots, 14, \tag{11}$$

where the $y_{i,t}$ is the average hourly earnings in area i during year t and $z_{i,t} = \ln(y_{i,t}) - \ln(y_{i,t-1})$.

In addition, to compare the performance of the restricted and unrestricted MCMC sampler on nonstationary series, we will also fit an AR(2) to the undifferenced series. The first stage of this model is

$$x_{i,t} = \phi_{i0} + \phi_{i1} x_{i,t-1} + \phi_{i2} x_{i,t-2} + \varepsilon_{i,t},$$
$$t = t_i, \ldots, 33, \quad i = 1, \ldots, 14, \tag{12}$$

where $x_{i,t} = \ln(y_{i,t})$.

In what follows, we obtain estimates of the posterior distributions of the autoregressive parameters and one-step-ahead predictors for Models (11) and (12).

Using the conditional posterior distributions (see Nandram and Petruccelli 1995), we performed the MCMC sampler for each dataset and for both the restricted and unrestricted cases with multiple runs (Gelman and Rubin 1992). Specifically, to begin the MCMC sampler, we drew 10 values of the ϕ_i from a dispersed distribution. The MCMC sampler was run on each of these 10 trajectories.

Within each step of the restricted MCMC sampler (i.e., the Metropolis–Hastings algorithm), the fifth iterated values of $\boldsymbol{\theta}$ and Δ from their respective conditional posterior distributions were taken. For a similar discussion of the Metropolis algorithm, see Müller (in press).

For both the AR(1) model fit to the differenced logged earnings data and the AR(2) model fit to the logged earnings data, as well as for all variations of the MCMC sampler (unrestricted, nonstationary restricted, stationary restricted, with or without initialization), we assessed the convergence of the MCMC sampler by studying the potential scale reductions (PSR's) and their 97.5 percentile points as suggested by Gelman and Rubin (1992). To do this we ran 500 iterations and used the last 250 to compute the PSR values. (PSR values near 1 are indicative of convergence.) For the earnings data, we obtained reasonable PSR values. For example, for the stationary restricted AR(1) model of differenced logged earnings, the quartiles for the PSR's for ϕ_0 and ϕ_1 are 1.011, 1.014, and 1.023 with initialization and 1.008, 1.012, and 1.021 without initialization. The corresponding quartiles for the 97.5 percentile points of the PSR's are 1.016, 1.021, and 1.031 and 1.012, 1.019, and

1.032, respectively. Plots of the trajectories of the model parameters show rapid convergence.

To be conservative, in each run of the MCMC sampler we used 500 iterates as a "burn-in." We then used a single sequence rather than multiple sequences for inference. Specifically, we ran the MCMC sampler for 2,000 iterations and selected every other one to give 1,000 "stationary" iterates. For all models that we fit to the earnings data, there is no indication of serial correlation in the iterates as indicated by the sample autocorrelations. From these convergence diagnostics, we conclude that the MCMC sampler performs satisfactorily in all cases studied.

For all models fit to the earnings data, we computed diagnostics (Gelfand, Dey, and Chang 1992) to assess model fit. In all cases, normal probability plots of standardized residuals showed consistency with appropriate fitted models and inconsistency with inappropriate fitted models (e.g., stationary restricted models fit to nonstationary data). Details were given by Nandram and Petruccelli (1995).

2.2 Numerical Results

We consider the effect of pooling on the autoregressive parameters, ϕ_i, the precision, τ_i, and on the one-step predictor of the last observation, $\hat{y}_{i,n+1}, i = 1, \ldots, m$. With this object in mind, we study two ratios. The first is the ratio of the posterior expectations for each series when only the data for the individual series are used versus the case when all the series are pooled.

For pooling under stationarity the ratio is

$$R_{\mathrm{ES}} = E(\cdot|\mathbf{y},\ \text{Individual})/E(\cdot|\mathbf{y},\ \text{Pooled}),$$

where the expectation in the denominator is taken with respect to the stationary restricted model and $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_m')'$ the data vector for all m series. We let R_{ENS} denote the ratio in which the expectation in the denominator is taken with respect to the nonstationary restricted model. We use the analogous ratios of posterior standard deviations, R_{SS} and R_{NSS}, to study the gain in precision obtained by using the hierarchical model.

The second ratio, which for each series compares the actual one-step prediction performance of the pooled stationary-restricted AR(1) or nonstationary-restricted AR(2) to the predictor based on the individual series, is

$$R_{\mathrm{PE}} = \frac{|y_{i,n+1} - E(y_{i,n+1}|\mathbf{y},\ \text{Individual})|}{|y_{i,n+1} - E(y_{i,n+1}|\mathbf{y},\ \text{Pooled})|}. \tag{13}$$

Finally, we limit our discussion to the unrestricted pooled case because our results suggest that for stationary (or nonstationary) series the estimated posterior distributions of the requisite parameters computed under the unrestricted model are nearly indistinguishable from those computed under the stationary (nonstationary) model.

2.2.1 Differenced Logged Earnings Data. For the differenced logged earnings data, the posterior probabilities of stationarity, $P(\phi_i \in \mathbf{\Phi}_p|\mathbf{y})$, computed from the MCMC sampler with initialization, except for series 6 and series 8 (posterior probabilities of stationarity .75 and .40, respectively) are all near 1, and thus the series may be considered stationary.

From columns 2–5 of Table 1, it can be seen that compared with individual estimation of ϕ_0 there is little difference in pooled estimation of ϕ_0 for all except series 6 and 8 under the stationarity restriction, but there are large differences in estimating ϕ_0 for all series under the nonstationarity restriction. Similarly, except for series 6, 8, 10, and 13, there is little difference in pooled estimation of ϕ_1 compared with individual estimation, but large differences in estimating ϕ_1 for nearly all series under the nonstationarity restriction.

Columns 6–9 of Table 1 show substantial improvement in the stationary-restricted case when the series are pooled. For ϕ_0, the unpooled posterior standard deviations range from 22% (for series 1, the longest series) to 101% (for series 8, one of the shortest series) greater than the pooled posterior standard deviations. For ϕ_1 there are comparable increases. In the nonstationary-restricted case the unpooled posterior standard deviations are lower than the pooled posterior standard deviations for series 10 for both ϕ_0 and ϕ_1.

Table 1. Unpooled to Pooled Ratios of Posterior Means and Standard Deviations for Each Series for the Autoregressive Coefficients, ϕ_0,ϕ_1, for the AR(1) Earnings Data

	Means				Standard deviations			
Series	$R0_{ES}$	$R1_{ES}$	$R0_{ENS}$	$R1_{ENS}$	$R0_{SS}$	$R1_{SS}$	$R0_{SNS}$	$R1_{SNS}$
1	.97	1.00	−8.18	.56	1.22	1.23	1.26	3.27
2	.95	1.02	−6.51	.65	1.33	1.31	1.51	3.70
3	.91	1.07	−3.58	.56	1.70	1.76	1.73	2.14
4	1.04	.86	−7.43	.39	1.54	1.55	1.83	3.03
5	1.06	.85	−19.32	.37	1.48	1.46	1.98	3.06
6	.54	1.21	−1.19	.73	1.48	1.48	1.36	1.44
7	1.10	.84	−10.17	.35	1.71	1.65	1.88	3.12
8	.35	1.42	−2.72	.91	2.01	2.20	1.45	1.73
9	1.08	.92	−13.07	.41	1.55	1.52	2.19	3.44
10	1.40	−.14	22.24	−.04	1.80	1.61	.69	.60
11	1.27	.78	−13.81	.32	1.72	1.72	1.90	2.31
12	1.24	.95	−12.62	.47	1.86	1.74	1.90	2.27
13	1.44	.54	−15.12	.21	1.78	1.66	1.61	1.71
14	1.07	1.02	−9.15	.57	1.67	1.55	1.67	2.14

NOTE: The quantities indexed ES denote the ratio $E(\cdot|\mathbf{Y}, \text{Individual})/E(\cdot|\mathbf{Y}, \text{Pooled})$ under the restriction of stationarity. Those indexed ENS denote the same quantities under the restriction of nonstationarity. The ratios indexed SS and SNS denote the corresponding ratios of standard deviations.

For all other series the increases for ϕ_0 are comparable to those in the stationary-restricted case, but those for ϕ_1 are much larger, ranging from 44 to 244%.

We used these same ratios to study the precisions τ_i of the process. We found that in the stationary-restricted case the unpooled posterior mean precisions range from 21% smaller to 98% greater than the pooled posterior means. The unpooled posterior standard deviations of precision range from 5% smaller to 123% greater than the pooled posterior standard deviations. On the other hand, for the nonstationary-restricted case the unpooled posterior mean precisions range from 16% to 190% greater than the pooled posterior means, whereas the unpooled posterior standard deviations of precision range from 8% smaller to 287% greater than the pooled posterior standard deviations.

Table 2 reveals that pooled prediction is comparable to unpooled prediction under the stationarity restriction for all series except series 8 and 10. The corresponding pooled standard deviations are also comparable to the unpooled. For pooled predictors under the nonstationarity restriction, the difference is more pronounced, especially for series 10. The unpooled standard deviations are anywhere from 14% smaller to 48% larger than the pooled standard deviations.

The models were fit to all but the last observation of the series, and that last observation was used as an out-of-sample value to evaluate the prediction error ratio R_{PE} given by (13). The results show that the unpooled predictor performs worse than the pooled predictor on series 1–4, 8, 9, 12, and 13 with absolute prediction error from 5% to 200% larger. On the remaining series the absolute prediction error of the unpooled predictor ranges from 9% to 91% smaller. This is not surprising in light of the results from Table 2 that show comparable means and prediction errors for the pooled and unpooled cases.

There is virtually no difference in the estimates and predictors for the pooled stationary-restricted model with and without initialization.

Finally, we study the contemporaneous correlations between series i and j,

$$\rho_{i,j} = \{(1 + 1/(\tau_i \psi^2))(1 + 1/(\tau_j \psi^2))\}^{-1/2}. \quad (14)$$

Table 2. Unpooled to Pooled Ratios of Posterior Means and Standard Deviations for Each Series for the One-Step Predictor, $y_{i,n+1}$, for the AR(1) Earnings Data

Series	R_{ES}	R_{ENS}	R_{SS}	R_{SNS}
1	.99	.82	1.00	.91
2	1.00	.91	1.00	.87
3	1.02	.83	1.01	1.04
4	.96	1.07	.99	1.09
5	.94	.72	.98	1.10
6	1.07	.87	1.02	.86
7	.96	.79	.99	1.19
8	1.30	.95	1.06	1.48
9	.99	.81	.99	1.08
10	1.19	2.41	1.08	1.21
11	.95	.63	.98	1.19
12	1.01	.66	.98	1.30
13	.92	.63	.97	1.13
14	1.04	.88	1.00	1.07

NOTE: The quantities indexed ES, ENS, SS, and SNS are defined in the note to Table 1.

For the stationary-restricted model with initialization, the $\rho_{i,j}$ range between .12 and .42. Without initialization, estimates of the $\rho_{i,j}$ are very similar.

2.2.2 Logged Earnings Data. Except for series 10, which had posterior probability of stationarity .44, all series have near zero posterior probabilities of being stationary.

From columns 2–7 of Table 3, we see that, with the exception of series 10, the overall difference for the stationary-restricted pooled estimators of ϕ_0 relative to the unpooled estimators is much greater than that for the nonstationary-restricted case. For ϕ_1, the difference is less for all series. Except for series 11, 12, and 14, the difference in estimating ϕ_2 is smaller for the nonstationary-restricted estimators.

Columns 8–13 of Table 3 compare the pooled and unpooled standard deviations. We note that for virtually all series there are large gains from pooling for the nonstationary-restricted case. There are larger increases in standard deviations of the pooled to the unpooled series for the stationary-restricted case.

Analyzing the precisions, we found that the nonstationary-restricted AR(2) model is on average more precise for the pooled than for the unpooled case except for series 6. This is the reverse of what happens with the stationary-restricted AR(2) for all but series 5 and 10. With the exception of series 5 and 10, pooling reduces the standard deviation of τ_i.

As for prediction, Table 4 shows that there is surprisingly little difference in predicting $y_{i,n+1}$. For the stationary-restricted case, the standard errors of prediction are virtually identical for the pooled and unpooled cases. For the nonstationary case, there are small to moderate gains from pooling for 7 of the 14 series. Notice that the two longest series, series 1 and 2, show losses of 17% and 24%.

The R_{PE} ratios for the AR(2) series reveal much the same mixed performance in actual one-step prediction as was seen in the AR(1) model.

We compare the estimates and predictors for the pooled nonstationary-restricted model with and without initialization. In contrast to the AR(1) case, there are large differences in estimates. For ϕ_0, the ratios of the estimates from the model with initialization to those without initialization range from .63 for series 5 to 11.2 for series 10. For ϕ_1, the range is .99 to 1.18 and for ϕ_2 the range is from .86 for series 4 to 38.6 for series 13. The ratios of the standard deviations are in general much larger than 1, with ratios of 1.089–1.665, 1.121–1.702, and 1.122–1.705 for ϕ_0, ϕ_1, and ϕ_2, respectively. Surprisingly, the one-step-ahead predictors are extremely close to the actual observations for all series, both with and without initialization (for series 14, for example, the actual value is 1.792, but the predictor with initialization is 1.808 and the predictor without initialization is 1.815). The range of ratios of predictors with initialization to those without initialization is .996–1.002. Much different behavior is exhibited by the standard deviations of the predictors whose ratios range from 1.12 to 2.82. The fact that these increases in standard deviations did not occur for the AR(1) model fit to stationary series suggests that there

Table 3. Unpooled to Pooled Ratios of Posterior Means and Standard Deviations for Each Series for the Autoregressive Coefficients ϕ_0,ϕ_1,ϕ_2 for the AR(2) Earnings Data

	Means						Standard deviations					
Series	$R0_{ES}$	$R1_{ES}$	$R2_{ES}$	$R0_{ENS}$	$R1_{ENS}$	$R2_{ENS}$	$R0_{SS}$	$R1_{SS}$	$R2_{SS}$	$R0_{SNS}$	$R1_{SNS}$	$R2_{SNS}$
1	.32	1.04	1.09	.97	1.01	1.03	1.46	1.05	1.04	1.39	1.33	1.32
2	.30	1.04	1.07	.89	1.01	1.02	1.25	.96	.96	1.28	1.25	1.25
3	−1.22	.77	.25	1.15	.98	.81	1.94	1.59	1.71	1.37	1.62	1.60
4	−.98	.68	−.11	1.20	.86	−.48	2.50	1.69	1.83	1.44	1.57	1.57
5	.32	.93	.72	1.62	.99	.98	1.77	1.33	1.36	1.58	1.56	1.55
6	−.60	.84	.57	1.03	1.05	1.21	1.23	1.57	1.63	.96	1.28	1.27
7	−1.27	.61	−.47	1.16	.91	1.88	2.23	1.48	1.60	1.40	1.52	1.51
8	−4.72	.86	.50	1.71	1.00	.90	5.26	2.91	3.44	1.87	1.83	1.91
9	.30	.70	.03	1.18	.96	.22	1.16	1.27	1.32	1.31	1.42	1.42
10	.83	.86	−.12	−35.03	.85	−.21	2.65	1.65	1.73	2.34	1.75	1.75
11	−1.28	.52	−.72	1.21	.80	2.49	2.02	1.50	1.63	1.28	1.49	1.48
12	−.67	.58	−.27	1.30	.88	7.87	1.81	1.76	1.86	1.25	1.47	1.46
13	−.33	.77	.02	.82	.94	.12	2.57	1.59	1.68	1.71	1.65	1.63
14	−2.47	.37	−.82	1.58	.64	3.32	2.80	2.39	2.66	1.42	1.50	1.50

NOTE: The quantities indexed ES, ENS, SS, and SNS are defined in the note to Table 1.

are difficulties involved in estimation and prediction when initialization is used for nonstationary series.

Finally, we note that the contemporaneous correlations between series i and j, given by (14) are much smaller than those seen in the AR(1) model, ranging from .012 to .108 with initialization and from .041 to .144 without initialization.

2.3 A Small Simulation Study

To further assess the improvement in estimation and forecasting due to pooling, we conducted a small simulation study using 10 AR(1) series. We used the model specified by (1)–(3) with all parameters fixed except θ_1 and ψ^2. Specifically, we fixed $\theta_0 = 1, \tau_i = 100, i = 1, \ldots, 10$, and we took the diagonal elements of the 2×2 matrix Δ to be .01 and the off-diagonal elements to be .005. We varied θ_1 at two levels, .3 and .8, and ψ^2 at three levels, .00, .01, and .10, corresponding to correlations of .00, .50, and .91, respectively; see (14). All 10 series were taken to be of equal length, and that length was varied at two levels, 10 and 20. Thus we studied 12 examples.

Table 4. Unpooled to Pooled Ratios of Posterior Means and Standard Deviations for Each Series for the One-Step Predictor, $y_{i,n+1}$, for the AR(2) Earnings Data

Series	R_{ES}	R_{ENS}	R_{SS}	R_{SNS}
1	1.01	1.00	1.00	.83
2	1.00	1.00	1.00	.76
3	1.01	1.00	1.00	.79
4	1.02	1.00	1.00	1.15
5	1.01	1.00	1.00	1.04
6	1.01	1.00	1.00	.64
7	1.03	1.00	1.00	.97
8	1.02	1.01	1.00	1.16
9	1.01	1.00	1.00	.85
10	1.01	.99	1.00	1.46
11	1.01	1.00	1.00	.82
12	.99	1.00	1.00	1.01
13	1.01	1.00	1.00	1.09
14	1.04	1.01	1.01	1.03

NOTE: The quantities indexed ES, ENS, SS, and SNS are defined in the note to Table 1.

For each example we generated the data using (1)–(3). We then ran the MCMC sampler in exactly the same way as described previously except that we took $\eta_0 = \delta_0 = 0$ (i.e., an improper prior on the τ_i) in (6).

First we assess the accuracy of estimation and prediction using the ratios of the estimated autoregressive parameters and one-step-ahead predictors to the true values. We denote the ratios for the unpooled case as R_U and those for the pooled case as R_P. For both the pooled and unpooled cases, the estimates of ϕ_1 tend to be biased downward (i.e., $R_P, R_U < 1$), but the bias disappears for the longer series and the larger value of θ_1. For the short series, variability of R_P and R_U increases with ψ^2. There are a few extremely low outliers in the unpooled case for series of length 20, and $\theta_1 = .8$. There are also outliers in the pooled case, but they are far less extreme. The situation for ϕ_0 is just the reverse, with upward bias and extremely high outliers in the unpooled case for series of length 20 and $\theta_1 = .8$. The ratios in the pooled case exhibit much less variability, however, than in the unpooled case for series length 10 and $\theta_1 = .8$. The accuracy of pooled and individual one-step-ahead predictors is comparable for all 12 examples. For the short series, there is upward bias that increases with ψ^2, whereas for the long series there is downward bias that increases with ψ^2.

We assess the performance of pooled versus individual estimators and one-step-ahead predictors by using two ratios, R_{E}, the ratio of the posterior expectations for the individual versus the pooled, and R_{S}, the ratio of the posterior standard deviations for the individual versus the pooled. For $\theta_1 = .3, R_{\mathrm{E}} \approx 1$ for both series lengths and all values of ψ^2 for both autoregressive parameters and the one-step-ahead predictors. For both estimation and prediction, there are large gains in precision when the series are pooled, and these gains increase with ψ^2. For example, for series of length 10 with $\psi^2 = .01, R_{\mathrm{S}}$ for ϕ_0 ranges between 1.07 and 2.03 and R_{S} for one-step-ahead prediction ranges between 1.02 and 1.76, whereas for $\psi^2 = .1, R_{\mathrm{S}}$ for ϕ_0 ranges between 1.47 and 2.44 and R_{S} for one-step-ahead prediction ranges between 1.26 and 1.94. The gains are less for the longer series; for example, for series length 20 and $\psi^2 = .1$, R_{S} for ϕ_0 ranges between 1.19 and 1.87 and R_{S} for one-step-ahead prediction ranges between 1.06 and 1.46.

For $\theta_1 = .8$ and series of length 20, $R_{\mathrm{E}} \approx 1$ for ϕ_0. For series of length 10, however, R_{E} is far more variable, and this variability increases with ψ^2. For ϕ_1 and the one-step-ahead predictors, $R_{\mathrm{E}} \approx 1$. The behavior of R_{S} is as described for $\theta_1 = .3$.

To sum up, we find little difference between unpooled and pooled point estimators and predictors for all examples and substantial gains in precision from pooling.

3. CONCLUSIONS

We have presented a summary of our methodology for estimation and forecasting any number of commensurate time series of possibly different lengths.

By using a hierarchical Bayesian model, we have accomplished four main tasks. First, we have shown that it is feasible to use sampling-based methods to analyze stationary or nonstationary autoregressive time series panel data. Second, we have obtained estimates and forecasts for nonstationary autoregressive series without transforming them to stationary ones. Third, by using latent variables for the AR(p) model, we have shown that it is feasible to use all the data without conditioning on the first p observations whether the series are stationary or not. Fourth, we have also incorporated contemporaneous correlations among series.

Application of these methods to panel data reveals two main findings. First, overall, the benefits of pooling could effect substantial improvements in estimation and forecasting. As expected, relative to inference based on individual series, when all the series are pooled there is considerable improvement for shorter series with smaller improvement in longer series. Second, there are differences in performance when the same series are unrestricted, stationary restricted, or nonstationary restricted. It is not surprising that there is no substantial difference in performance between unrestricted and stationary-restricted fits for stationary series. On the same note, there is no substantial difference in performance between unrestricted and nonstationary-restricted fits for nonstationary series. Restricting nonstationary series to be stationary (or stationary series to be nonstationary), however, results in biased estimators with artificially low variances.

ACKNOWLEDGMENTS

We are grateful to the editor and four referees for their insightful comments, which led to substantial improvements in this article. The research of Balgobin Nandram was supported in part by National Science Foundation grant DMS 9204812.

[Received September 1993. Revised March 1996.]

REFERENCES

Andrews, R. W. (1976), "Multiperiod Predictions From an Autoregressive Model Using Empirical Bayes Methods," technical report, University of Michigan, Graduate School of Business Administration.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis: Forecasting and Control*, San Francisco: Holden-Day.

Broemeling, L. (1985), *Bayesian Analysis of Linear Models*, New York: Marcel Dekker.

Chib, S., and Greenberg, E. (1994), "Bayes Inference in Regression Models With ARMA(p, q) Errors," *Journal of Econometrics*, 64, 183–206.

——— (1995), "Hierarchical Analysis of SUR Models With Extensions to Correlated Serial Errors and Time-Varying Parameter Models," *Journal of Econometrics*, 68, 339–360.

Chow, G. C. (1973), "Multiperiod Predictions From Stochastic Difference Equations by Bayesian Methods," *Econometrica*, 41, 109–117.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), "Model Determination Using Predictive Distributions With Implementation via Sampling-Based Methods," in *Bayesian Statistics 4*, ed. J. Bernardo: Oxford, U.K.: Oxford University Press, pp. 148–167.

Gelfand, A. E., and Smith, A. F. M. (1990), "Sampling Based Approaches for Calculating Marginal Densities," *Journal of the American Statistical Association*, 85, 398–409.

Gelfand, A. E., Smith, A. F. M., and Lee, T. M. (1992), "Bayesian Analysis of Constrained Parameter and Truncated Data Problems Using Gibbs Sampler," *Journal of the American Statistical Association*, 87, 523–530.

Gelman, A., and Rubin, D. B. (1992), "Inference From Iterative Simulation Using Multiple Sequences" (with discussion), *Statistical Science*, 7, 457–472, 483–511.

Kadiyala, K. R., and Karlsson, S. (1993), "Forecasting With Generalized Bayesian Vector Autoregressions," *Journal of Forecasting*, 12, 365–378.

Ledolter, J., and Lee, C. (1993), "Analysis of Many Short Sequences: Forecast Improvements Achieved by Shrinkage," *Journal of Forecasting*, 12, 1–11.

Li, W. K., and Hui, Y. V. (1983), "Estimation of Random Coefficient Autoregressive Process: An Empirical Bayes Approach," *Journal of Time Series Analysis*, 4, 89–94.

Lindley, D. V., and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society*, Ser. B, 34, 1–41.

Litterman, R. B. (1986), "Forecasting With Bayesian Vector Autoregressions—Five Years' of Experience," *Journal of Business & Economic Statistics*, 4, 25–38.

Liu, L., and Tiao, G. C. (1980), "Random Coefficient First-Order Autoregressive Models," *Journal of Econometrics*, 13, 305–325.

Marriott, J., Ravishanker, N., and Gelfand, A. E. (1992), "Bayesian Inference in Stationary Autoregressive Models Using Gibbs Sampling," technical report, University of Connecticut, Dept. of Statistics.

Müller, P. (in press), "Metropolis Based Posterior Integration Schemes," *Journal of Computational and Graphical Statistics*, 6.

Nandram, B., and Petruccelli, J. D. (1995), "Bayesian Analysis of Autoregressive Time Series Data: A Gibbs Sampler Approach," technical report, Worcester Polytechnic Institute, Dept. of Mathematical Sciences.

Pai, J., Ravishanker, N., and Gelfand, A. E. (1993), "Bayesian Analysis of Concurrent Time Series With Application to Regional IBM Revenue Data," technical report, University of Connecticut, Dept. of Statistics.

Ravishanker, N., Wu, L. S.-Y., and Dey, D. K. (1992), "Shrinkage Estimation for Business Planning and Forecasting of Regional IBM Revenue," technical report, University of Connecticut, Dept. of Statistics.

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

Tierney, L. (1994), "Markov Chains for Exploring Posterior Distributions" (with discussion), *The Annals of Statistics*, 22, 1701–1762.