Bayesian Analysis of Autoregressive Time Series Panel Data: A Gibbs Sampler Approach

Balgobin Nandram*
Worcester Polytechnic Institute

and

Joseph D. Petruccelli
Worcester Polytechnic Institute

May 5, 1995

Abstract

We employ the Gibbs sampler to implement the Bayesian paradigm to analyze autoregressive time series panel data which arise naturally in business and economics applications. A hierarchical Bayesian normal linear model is developed for such data, and a Bayesian diagnostic is proposed to assess model fit. Two versions of the Gibbs sampler are used: A restricted version which enforces stationarity or nonstationarity conditions on the series, and an unrestricted version which does not. Two sets of latent variables are used: The first set enables us to use all observations in fitting and forecasting, and the second models contemporaneous correlations between series. We use two examples to show how to implement both versions of the Gibbs sampler to do model checking, estimation, and forecasting. As expected, we find that restricting stationary series to be stationary does not add new information, but restricting nonstationary series to be stationary leads to estimators and predictors which differ substantially from those obtained in the unrestricted case. Compared with inference based on individual series, there are gains in precision for estimation and forecasting when similar series are pooled, with larger gains for shorter series. We validate these conclusions with a small simulation study.

Key Words: Data analysis; diagnostics; hierarchical models; latent variables; Metropolis algorithm; stationary-restricted.

*Supported in part by NSF grant DMS 9204812

1 Introduction

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 model (e.g., an AR(p)). The key advantage of simultaneously modeling several series is the possibility of pooling information from all series. This not only can 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.

There has been a shift in recent years away from modeling such series with structural equations and toward modeling them with Bayesian vector autoregressive (BVAR) models (see, e.g., Sims, 1980, Litterman, 1986 and Kadiyala and Karlsson, 1993), which capture inter-series as well as within-series relations. While BVAR models have proven useful in some applications, when there is a large number of short series, the large number of parameters involved in the BVAR model forces over-reliance on and over-simplification of prior specifications. An example of this over-simplification is found in the so-called "Minnesota prior" (see Litterman, 1980). More complex priors, such as the normal-Wishart (see Kadiyala and Karlsson, 1993), remedy the over-simplification, but result in computational intractability due to the large number of parameters involved. Thus for such situations, an alternate approach is desirable. In this paper, we develop a methodology which can model a large number of simultaneous commensurate series.

A number of alternate approaches to this problem has been previously considered. Using a simpler model than we consider here, Anderson (1978) and Azzalini (1981) considered the pooling of information from several series generated from exactly the same autoregressive model. Other researchers have considered more general models in which the autoregressive parameters vary from series to series, providing greater flexibility in modeling (see, for example, Robinson, 1978).

Several authors have tried empirical Bayes-type approaches to the problem. These include Andrews (1976), Ravishanker et al (1992), Ledolter and Lee (1993), and Li and Hui (1983). One common concern with such approaches is that they tend to overestimate precision.

Bayesian formulations have included those of Chow (1973), who considered multistep forecasting, Pai et al (1993), who proposed a model different from ours, and Liu and Tiao (1980), who, in the paper 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. Also we have evidence that for the stationary AR(1) process, even if the exact computations are executed for the Liu and Tiao model, the posterior distributions of the autoregressive parameters tend to have the undesirable property of being bimodal. (This is due to the beta prior assumption.) For higher order models the task of making this type of analysis accessible to non-expert users appears impossible. 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 et al (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 in Pai et al (1993).

The modeling of stationarity restrictions raises some interesting questions of its own. First, given that one is modeling uncertainty as to the values of the autoregressive parameters, might this not at times include uncertainty as to whether the process is stationary or nonstationary? Especially in the case of panel data where there may be many short series, one might be uncertain if a given series is stationary. If so, the stationarity restrictions could prove a hindrance. Second, given that modeling the stationarity restrictions involves a significant effort, what advantages does it confer, if any, in a practical sense when the series are stationary? Furthermore, what disadvantages, if any, does it confer in a practical sense when at least some of them are not?

We believe the answer to the first question is yes. In this paper, we attempt to answer the second and third questions using two examples; one consists of apparently stationary realizations and the other nonstationary realizations. This permits comparisons between methodology which enforces stationarity or nonstationarity restrictions (which we term the restricted case), and methodology which does not (the unrestricted case) when applied to both stationary and nonstationary series.

Our approach to the restricted case differs from that of Marriott et al (1992) in two respects. First, as it is not possible to use the full likelihood when the series are nonstationary, we use latent variables in both the stationary and nonstationary cases, so as to avoid conditioning on the first p observations. This also facilitates comparison between the restricted and unrestricted cases, and eliminates the bimodality problem mentioned earlier. Second, rather than transform first to partial autocorrelations, and then to normality, we incorporate the stationarity restrictions directly into the modeling procedure.

In this paper we describe a fully Bayesian solution to the problem of estimation and forecasting autoregressive time series panel data. 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.

In such settings, sampling-based approaches to the calculation of marginal posterior densities can provide solutions. We use the Gibbs sampler (see Gelfand et al, 1992, for a general description, and Chib and Greenberg, 1993, for an application to a single time series with ARMA(p, q) errors) to perform integrations over the variance components for both models. In the restricted case, we also apply the Metropolis algorithm (Tierney, 1991, 1994, Müller, 1994) to obtain the otherwise intractable conditional posterior of the hyperparameters of the restricted autoregressive parameters. In particular, we describe computations for both stationary and nonstationary models and compare the performance of the restricted and unrestricted methods on both stationary and nonstationary 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.

In section 2 of the paper we present the model. The computations are described in section 3. The methodology is exemplified in section 4 with the analysis of a data set on yearly averages of hourly earnings of production workers in fourteen California metropolitan areas. We also perform a small simulation study to assess the gains in estimation and forecasting due to pooling. Section 5 has concluding remarks.

2 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 i^{th} 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) and that the last observation occurs at the same time, n, for all series. We let $n_i = n - t_i + 1$ denote the number of observations in the i^{th} 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 $\mathbf{y}^{(0)} = \left(\mathbf{y}_1^{(0)\prime}, \ldots, \mathbf{y}_m^{(0)\prime}\right)'$. For each $1 \leq t \leq n$, we let $I_t = \{1 \leq i \leq m : t_i \leq t\}$ denote the set of series which have observations at time t, and let m_t denote the number of such series.

The defining relation for the i^{th} series, given the parameters $\boldsymbol{\phi}_i$, τ_i, ψ^2 and $\mathbf{y}_i^{(0)}$, is:

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

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 $\boldsymbol{\varepsilon}_t'$ 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

$$\boldsymbol{\varepsilon}_t \mid \boldsymbol{\tau}, \psi^2 \overset{iid}{\sim} N(\mathbf{0}, \Sigma_t), \tag{2}$$

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

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

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

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

where θ_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

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

and

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

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 as we need not condition on the first p observations in each series.

Next we consider stationarity restrictions. Define the stationarity region of series i to be

$$\Phi_p = \{\boldsymbol{\phi}_i : \text{series (1) is stationary}\}.$$

See, e.g., Box and Jenkins (1976) for details on stationarity. Of the two fitted models that we consider in section 4, the first is an ARI(1) (i.e., an AR(1) in the differenced series) and the second is an AR(2). For $p = 1$ and 2 the stationarity regions are $\Phi_1 = \{(\phi_{i,0}, \phi_{i,1}) : |\ \phi_{i,1}\ |< 1\}$, and $\Phi_2 = \{(\phi_{i,0}, \phi_{i,1}, \phi_{i,2}) : \phi_{i,1} + \phi_{i,2} < 1, \phi_{i,2} - \phi_{i,1} < 1,$ $|\ \phi_{i,2}\ |< 1\}$, respectively. Letting $\mathbf{y}_i = (y_{i,t_i}, y_{i,t_i+1} \ldots, y_{i,n})'$ denote the data vector for the i^{th} series and $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_m')'$ the data vector for all m series, we note that the stationary probabilities, a priori $P(\boldsymbol{\phi}_i \in \Phi_p)$ and a posteriori $P(\boldsymbol{\phi}_i \in \Phi_p \mid \mathbf{y})$, are well defined.

In section 3 we describe two versions of the Gibbs sampler. While in the first, or unrestricted, version, the autoregressive parameters $\boldsymbol{\phi}_i$ are sampled from the normal distribution in (3), in the second, or restricted, version, the sampling is done from distribution (3) with the $\boldsymbol{\phi}_i$ restricted to Φ_p or its complement.

3 Computations

Defining $\overline{\hat{\boldsymbol{\phi}}} = \sum_{i=1}^{m} \hat{\boldsymbol{\phi}}_i/m$, we take $\nu_0 = p+2$ and $\Delta_0 = S_{\boldsymbol{\phi}}/\nu_0$, where $S_{\boldsymbol{\phi}} = \sum_{i=1}^{m}(\hat{\boldsymbol{\phi}}_i - \overline{\hat{\boldsymbol{\phi}}})(\hat{\boldsymbol{\phi}}_i - \overline{\hat{\boldsymbol{\phi}}})'/(m-1)$. We also estimate θ_0, C_0, ν_0, Δ_0, η_0 and δ_0 using the data. However, we take α_0 and β_0 to be zero, and therefore, ψ^2 has a noninformative prior. In addition, using the data, we obtain $\mathbf{y}_i^{(0)}$ for the i^{th} series by backcasting the autoregressive process. 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 backcasting is done, and $t_0 = p+1$ otherwise. To facilitate computation, we rewrite (2) in terms of latent variables α_t:

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

and

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

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 data set we run both the restricted and unrestricted versions of the Gibbs sampler to obtain M stationary iterates of the $\boldsymbol{\phi}_i, \psi^2$ and $\Omega = \{\mathbf{y}^{(0)}, \boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta}, \Delta\}$ when backcasting ($\Omega = \{\boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta}, \Delta\}$ when not backcasting), which we denote by $\boldsymbol{\phi}_i^{(j)}, i = 1, \ldots, m, \psi^{2(j)}$ and $\Omega^{(j)}$ respectively, $j = 1, 2, \ldots M$. We use these iterates for estimation, forecasting and model assessment.

3.1 The Unrestricted Gibbs Sampler

To implement the Gibbs sampler, we must be able to sample from the full conditional distributions of the parameters $\{\boldsymbol{\phi}_i\}_{i=1}^m$, ψ^2 and Ω.

First, we describe the conditional posterior distributions of the ϕ_i. Define the following quantities for all $i = 1, \ldots, m$: $\mathbf{d}_i = \sum_{t=t_i}^n y_{i,t}\mathbf{y}_{i,t-1}$, $G_i = \sum_{t=t_i}^n \mathbf{y}_{i,t-1}\mathbf{y}'_{i,t-1}$, $\Lambda_i = (G_i + (\tau_i\Delta)^{-1})^{-1}G_i$, $\hat{\boldsymbol{\phi}}_i = G_i^{-1}\boldsymbol{d}_i$, and $\hat{\hat{\boldsymbol{\phi}}}_i = \Lambda_i^{-1}(\mathbf{y}_{i,t_0-1}, \mathbf{y}_{i,t_0-2}, \ldots, \mathbf{y}_{i,t_0-p})\boldsymbol{\alpha}$. Then the conditional posterior distribution

$$\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega \overset{ind}{\sim} N(\Lambda_i(\hat{\boldsymbol{\phi}}_i - \hat{\hat{\boldsymbol{\phi}}}_i) + (I - \Lambda_i)\boldsymbol{\theta}, (I - \Lambda_i)\Delta), i = 1, \ldots, m. \quad (11)$$

Second, letting $\boldsymbol{\phi} = (\phi'_1, \ldots, \phi'_m)'$, the conditional posterior distribution of τ_i is

$$\tau_i \mid \mathbf{y}, \mathbf{y}^{(0)}, \boldsymbol{\alpha}, \boldsymbol{\phi} \overset{ind}{\sim} G((n_i - p + \eta_0)/2, b_i^{-1}) \quad (12)$$

where $b_i = \frac{1}{2}\left[\sum_{t=t_0}^n e_{i,t}^2 + \delta_0\right]$ and $e_{i,t} = y_{i,t} - (\boldsymbol{\phi}'_i\mathbf{y}_{i,t-1} + \alpha_t)$ are the residuals. Also, the posterior conditional distribution of ψ^2 is

$$\psi^2 \mid \boldsymbol{\alpha} \sim \mathrm{IG}((\alpha_0 + n - t_0 + 1)/2, (\beta_0 + \sum_{t=t_0}^{n} \alpha_t^2)/2). \quad (13)$$

Third,

$$\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta \sim N(\Gamma\overline{\boldsymbol{\phi}} + (I - \Gamma)\boldsymbol{\theta}_0, (I - \Gamma)C_0) \quad (14)$$

where $\overline{\boldsymbol{\phi}} = \sum_{i=1}^m \boldsymbol{\phi}_i/m$, and $\Gamma = [\Delta^{-1} + (mC_0)^{-1}]^{-1}\Delta^{-1}$. Also,

$$\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta} \sim W(\{\sum_{i=1}^{m}(\boldsymbol{\phi}_i - \boldsymbol{\theta})(\boldsymbol{\phi}_i - \boldsymbol{\theta})' + \nu_0\Delta_0\}^{-1}, m + \nu_0). \quad (15)$$

Fourth,

$$\alpha_t \mid \mathbf{y}, \mathbf{y}^{(0)}, \boldsymbol{\phi}, \boldsymbol{\tau}, \psi^2 \overset{ind}{\sim} N\left(\sigma_{\alpha t}^{-2} \sum_{i \in I_t} \tilde{e}_{i,t}\tau_i, \sigma_{\alpha t}^2\right) \quad (16)$$

where $\tilde{e}_{i,t} = y_{i,t} - \boldsymbol{\phi}'_i\mathbf{y}_{i,t-1}$ and $\sigma_{\alpha t}^2 = \left(\psi^{-2} + \sum_{i \in I_t} \tau_i\right)^{-1}$.

Finally, we consider the conditional posterior distribution for $\mathbf{y}_i^{(0)}$. Let $\mathbf{c_i} = (c_{i,1}, \ldots, c_{i,p})'$, where $c_{i,1} = y_{i,1} - (\phi_{i,0} + \alpha_{t_i})$, and $c_{i,k} = y_{i,k} - (\phi_{i,0} + \sum_{j=1}^{k-1} \phi_{i,j} y_{i,j} + \alpha_{t_i+k-1})$, $k = 2, \ldots, p$. Also, let Φ_i be the $p \times p$ lower triangular matrix with the main diagonal having all entries equal to $\phi_{i,p}$ and the j^{th} subdiagonal having all entries equal to $\phi_{i,p-j}$, $j = 1, \ldots, p-1$. Then

$$\mathbf{y}_i^{(0)} \mid \boldsymbol{\phi}_i, \alpha_{t_i}, \alpha_{t_i+1}, \ldots, \alpha_{t_i+p-1}, \tau_i \sim N(\mathbf{b}_0^*, B_0^*) \tag{17}$$

where $\mathbf{b}_0^* = (\tau_i \Phi_i' \Phi_i + B_0^{-1})^{-1}(\tau_i \Phi_i' \Phi_i \mathbf{c}_i + B_0^{-1}\mathbf{b}_0)$ and $B_0^* = (\tau_i \Phi_i' \Phi_i + B_0^{-1})^{-1}$.

The Gibbs sampler is implemented in a simple manner. Using appropriate starting values, one random deviate is drawn from each of (11)-(17) in turn, repeating this process until convergence.

3.2 The Restricted Gibbs Sampler

In this section, we discuss the implementation of the Gibbs sampler when the $\boldsymbol{\phi}_i$ are restricted to any subregion, C^{p+1}, of $\mathbf{R}^{p+1}$. However, in our application C^{p+1} will be either Φ_p or its complement (i.e., the region of stationarity or nonstationarity).

The appropriate conditional posterior distributions for $\boldsymbol{\tau}$, ψ^2, $\boldsymbol{\alpha}$ and $\mathbf{y}^{(0)}$ are given by (12), (13), (16) and (17), respectively.

If $\boldsymbol{\phi}_i$ is restricted to C^{p+1}, the conditional posterior distribution of $\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega$ is

$$p^{(r)}(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega) = \frac{p(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega)}{\int_{\boldsymbol{\phi}_i \in C^{p+1}} p(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega) d\boldsymbol{\phi}_i},$$

$\boldsymbol{\phi}_i \in C^{p+1}, i = 1, \ldots, m$, where $p(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega)$ is given by (11). In practice, $\boldsymbol{\phi}_i$ is obtained from $p^{(r)}(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega)$ by rejection sampling.

The conditional distributions of $\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta$ and $\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta}$ are much more complicated. Due to the restriction on the $\boldsymbol{\phi}_i$, the conditional posterior of these hyperparameters is intractable, and therefore difficult to sample from (Gelfand, Smith and Lee, 1992, Section 2).

Let $g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$ denote the conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{\phi}, \Delta$ in the unrestricted case. Then the conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{\phi}, \Delta$ in the restricted case is

$$g^{(r)}(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta) \propto \frac{g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)}{[\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta) d\mathbf{z}]^m},$$

$\boldsymbol{\theta} \in \mathbf{R}^{p+1}$, where $\eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta)$ is the (p+1)-variate $\mathrm{N}(\boldsymbol{\theta}, \Delta)$ density.

Similarly, let $d(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$ denote the conditional distribution of Δ^{-1} given $\boldsymbol{\phi}, \boldsymbol{\theta}$ in the unrestricted case. Then the conditional distribution of Δ^{-1} given $\boldsymbol{\phi}, \boldsymbol{\theta}$ in the restricted case is

$$d^{(r)}(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta}) \propto \frac{d(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})}{[\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta) d\mathbf{z}]^m}.$$

where the support of Δ^{-1} is the same as that of an unrestricted Wishart distribution in (15). Note that while $g^{(r)}(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$ and $d^{(r)}(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$ are both complicated, it is easy to sample from $g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$ and $d(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$. Also note that for each $\boldsymbol{\theta}$ and Δ, $\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta) d\mathbf{z}$ can be obtained easily by Monte Carlo integration.

To obtain a random vector $\boldsymbol{\theta}$ from $g^{(r)}(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$, or Δ^{-1} from $d^{(r)}(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$, we use the Metropolis algorithm with an independence chain (Tierney, 1991, 1995; Müller, 1994). We use $g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$ or $d(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$ as a proposal density. Thus, the acceptance probability in the Metropolis algorithm is

$$A(\Delta) = \left[\frac{\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}_{(1)}, \Delta) d\mathbf{z}}{\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}_{(2)}, \Delta) d\mathbf{z}} \right]^m,$$

where $\boldsymbol{\theta}_{(1)}$ is the current state, and $\boldsymbol{\theta}_{(2)}$ is the proposed state, a value from the proposal density $g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$. Similarly, the acceptance probability for drawing from $d(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$ is

$$A(\boldsymbol{\theta}) = \left[\frac{\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta_{(1)}) d\mathbf{z}}{\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta_{(2)}) d\mathbf{z}} \right]^m.$$

To approximate $A(\boldsymbol{\theta})$, we select L independent samples from $\eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta_{(1)})$, and L independent samples from $\eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta_{(2)})$ for each $\boldsymbol{\theta}$ and count the number of $\mathbf{z}$ falling in C^{p+1} in either case. We estimate $A(\boldsymbol{\theta})$ by using $\hat{A}(\boldsymbol{\theta})$, the ratio of these counts. We choose L by insisting that

$$P\left\{ |\hat{A}(\boldsymbol{\theta}) A(\boldsymbol{\theta})^{-1} - 1| \leq .05 \right\} = .95.$$

The Metropolis algorithm for drawing $\boldsymbol{\theta}$ is utilized in a similar manner.

Thus, we obtain one iterate in the restricted Gibbs sampler by drawing $\boldsymbol{\phi}_i$ from $p^{(r)}(\boldsymbol{\phi}_i \mid \boldsymbol{\theta}, \Delta)$ by rejection sampling, $\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta$ and $\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta}$ by the Metropolis algorithm with an independence chain in each case, and $\boldsymbol{\tau}$, ψ^2,$\boldsymbol{\alpha}$ and $\mathbf{y}^{(0)}$ as in the unrestricted case.

3.3 Checks on Model Adequacy

To evaluate the adequacy of the models, we compute a multivariate predictive diagnostic in the spirit of Gelfand et al (1992).

Defining $\tilde{\mathbf{y}}_t = \{y_{i,t}, i \in I_t\}$, $\tilde{\mathbf{y}}_{(t)} = (\tilde{\mathbf{y}}_1', \tilde{\mathbf{y}}_2', \ldots, \tilde{\mathbf{y}}_{t-1}')'$, $\mathbf{u}_t = E(\tilde{\mathbf{y}}_t \mid \tilde{\mathbf{y}}_{(t)})$ and $\mathcal{V}_t = \text{var}(\tilde{\mathbf{y}}_t \mid \tilde{\mathbf{y}}_{(t)}), t = p+1 \ldots, n$, the diagnostic is

$$\mathbf{d}_t = \mathcal{V}_t^{-1/2}(\tilde{\mathbf{y}}_t - \mathbf{u}_t),\ t = t_0, t_0+1, \ldots, n. \tag{18}$$

Denoting the components of $\mathbf{d}_t$ as $\{d_{i,t}, i \in I_t\}$, we note that if the model is appropriate, the $\{d_{i,t}, i \in I_t,\ t = 1, \ldots, n\}$ will be approximately a random sample from a $N(0,1)$ distribution.

We use the stationary iterates from the Gibbs sampler to compute the expectations by using weighted averages, in a manner similar to that outlined by Gelfand et al (1992). For example, letting $\tilde{\boldsymbol{\phi}}_t = \{\boldsymbol{\phi}_i, i \in I_t\}$, we compute the conditional mean $\mathbf{u}_t$, as

$$\mathbf{u}_t = E_{\tilde{\boldsymbol{\phi}}_t, \Sigma_t \mid \tilde{\mathbf{y}}_{(t)}}[E(\tilde{\mathbf{y}}_t \mid \tilde{\mathbf{y}}_{(t)}, \tilde{\boldsymbol{\phi}}_t, \Sigma_t)] \approx \sum_{j=1}^{M} E(\tilde{\mathbf{y}}_t \mid \tilde{\mathbf{y}}_{(t)}, \tilde{\boldsymbol{\phi}}_t^{(j)}, \Sigma_t^{(j)}) w_t^{(j)}$$

where

$$w_t^{(j)} = \{f(\tilde{\mathbf{y}}_t, \tilde{\mathbf{y}}_{t+1}, \ldots, \tilde{\mathbf{y}}_n \mid \tilde{\mathbf{y}}_{(t)}, \tilde{\boldsymbol{\phi}}_t^{(j)}, \Sigma_t^{(j)})\}^{-1} / \sum_{j=1}^{M} \{f(\tilde{\mathbf{y}}_t, \tilde{\mathbf{y}}_{t+1}, \ldots, \tilde{\mathbf{y}}_n \mid \tilde{\mathbf{y}}_{(t)}, \tilde{\boldsymbol{\phi}}_t^{(j)}, \Sigma_t^{(j)})\}^{-1},$$

$t = p+1, p+2, \ldots, n, j = 1, \ldots, M$, and $f(\cdot \mid \tilde{\mathbf{y}}_{(t)}, \tilde{\boldsymbol{\phi}}_t^{(j)}, \Sigma_t^{(j)})$ is the conditional multivariate normal density of present and future observations given past observations, $\tilde{\boldsymbol{\phi}}_t^{(j)}$ and $\Sigma_t^{(j)}$. The other expectations are computed in a similar manner.

3.4 Estimation and Forecasting

For both the restricted and unrestricted Gibbs sampler, we study the sampling process by computing the posterior distributions of $\phi_i \mid \mathbf{y}$ and $\tau_i \mid \mathbf{y}$. We also obtain the posterior distributions of one-step-ahead forecasts from time n for each series.

First for the unrestricted case, as an estimate of the posterior distribution $p(\phi_i \mid \mathbf{y})$ we compute $\sum_{j=1}^{M} p(\phi_i \mid \mathbf{y}, \Omega^{(j)})/M$, where the distribution of $\phi_i \mid \mathbf{y}, \Omega^{(j)}$ is given by (11). Thus letting $\overline{\mathbf{h}}_i = \sum_{j=1}^{M} \mathbf{h}_i^{(j)}/M$, where $\mathbf{h}_i^{(j)} = \Lambda_i^{(j)}(\hat{\phi}_i^{(j)} - \hat{\hat{\phi}}_i^{(j)}) + (I - \Lambda_i^{(j)})\theta^{(j)}$, we estimate the posterior mean and variance of ϕ_i by $\overline{\mathbf{h}}_i$ and $\{\sum_{j=1}^{M}(I - \Lambda_i^{(j)})\Delta^{(j)} + \sum_{j=1}^{M}(\mathbf{h}_i^{(j)} - \overline{\mathbf{h}}_i)(\mathbf{h}_i^{(j)} - \overline{\mathbf{h}}_i)'\}/M$, respectively.

For the restricted case these simple Rao-Blackwellized estimators do not exist. The posterior density of $\phi_i \mid \mathbf{y}$ is obtained directly from the empirical distribution of the stationary iterates, using nonparametric density estimation (Silverman, 1986). Letting $\overline{\phi}_i = \sum_{j=1}^{M} \phi_i^{(j)}/M$, we estimate the posterior mean and variance by $\overline{\phi}_i$ and $\sum_{j=1}^{M}(\phi_i^{(j)} - \overline{\phi}_i)(\phi_i^{(j)} - \overline{\phi}_i)'/M$, respectively.

For both the restricted and unrestricted cases, as an estimate of the posterior distribution $p(\tau_i \mid \mathbf{y})$, we compute $\sum_{j=1}^{M} p(\tau_i \mid \mathbf{y}, \mathbf{y}^{(0)^{(j)}}, \alpha^{(j)}, \phi_i^{(j)})/M$, where the distribution of $\tau_i \mid \mathbf{y}, \mathbf{y}^{(0)}, \alpha, \phi$ is given by (12). We obtain posterior distributions of other parameters in a similar way.

Forecasting the next observation in series i is straightforward. Noting that $y_{i,n+1} \mid \mathbf{y}, \mathbf{y}^{(0)}, \alpha, \phi_i, \tau_i \sim N(\phi_i' \mathbf{y}_{i,n} + \alpha_{n+1}, \tau_i^{-1})$, we estimate the posterior distribution of $y_{i,n+1}$ by $\sum_{j=1}^{M} p(y_{i,n+1} \mid \mathbf{y}, \mathbf{y}^{(0)^{(j)}}, \alpha^{(j)}, \phi_i^{(j)}, \tau_i^{(j)})/M$. Thus, letting $g_i^{(j)} = \phi_i^{(j)\prime}\mathbf{y}_{i,n} + \alpha_{n+1}^{(j)}$ and $\overline{g}_i = \sum_{j=1}^{M} g_i^{(j)}/M$, we estimate the posterior mean and variance of $y_{i,n+1}$ by $\overline{g}_i$ and $\{\sum_{j=1}^{M} 1/\tau_i^{(j)} + \sum_{j=1}^{M}(g_i^{(j)} - \overline{g}_i)^2\}/M$, respectively.

The Gibbs sampler also allows computation of the estimated posterior probability that a given series is stationary. This is done by counting the number of the 1000 stationary Gibbs iterates that yield autoregressive parameters in the region of stationarity. That is, we estimate $P(\phi_i \in \Phi_p \mid \mathbf{y})$.

4 Empirical Studies

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

4.1 Earnings Data

The data (Liu and Tiao, 1980) consist of yearly averages of the hourly earnings of production workers in fourteen California metropolitan areas. Each of the fourteen series ends in 1977, but the series are of different lengths with the longest (of length 33) beginning in 1945 and the shortest (of length 15) beginning in 1963. 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 fourteen series are nonstationary, but taking a first difference transforms all of them to stationarity. As did Liu and Tiao, we fit an AR(1) to the differenced data. The first stage of the model is:

$$z_{i,t} = \phi_{i0} + \phi_{i1} z_{i,t-1} + \epsilon_{i,t},\ t = t_i + 1, \ldots, 33,\ i = 1 \ldots, 14 \tag{19}$$

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, in order to compare the performance of the restricted and unrestricted Gibbs sampler algorithm on nonstationary series, we will also fit an AR(2) to the undifferenced series. The first stage of this model is:

$$x_{i,t} = \phi_{i0} + \phi_{i1} x_{i,t-1} + \phi_{i2} x_{i,t-2} + \epsilon_{i,t},\ t = t_i, \ldots, 33,\ i = 1 \ldots, 14 \tag{20}$$

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 (19) and (20).

4.1.1 Computations and Model Assessment

Using the conditional distributions (11)-(17) we performed the Gibbs sampler algorithm for each data set and for both the restricted and unrestricted cases using multiple runs (Gelman and Rubin, 1992). Specifically, to begin the Gibbs sampler, we drew ten values of the $\boldsymbol{\phi}_i$ from a dispersed distribution. The Gibbs sampler was run on each of these ten trajectories.

Within each step of the restricted Gibbs sampler the Metropolis algorithm was run five times (Müller, 1994) to obtain values of $\boldsymbol{\theta}$ and Δ from their respective conditional posterior distributions. To obtain an estimate of the acceptance probability when $\boldsymbol{\theta}$ $(or \Delta)$ was selected from its conditional posterior distribution, 1000 values of $\mathbf{z}$ were drawn from $\eta(\mathbf{z} \mid \boldsymbol{\theta}_{(i)}, \Delta)$ (or $\eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta_{(i)})), i = 1, 2$.

Figure 1 displays the first 50 iterates of the trajectories for ϕ_0 and ϕ_1 for series 6 of the differenced logged earnings data. Plots (a) and (b) use backcasting while (c) and (d) do not. We note that the trajectories converge rapidly. Plots for the other series and for other parameters showed similar rapid convergence, as did plots for the AR(2) model fit to the logged earnings data.

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 Gibbs sampler algorithm (unrestricted, nonstationary restricted, stationary restricted, with or without backcasting) we assessed the convergence of the Gibbs sampler algorithm by studying the potential scale reductions (PSR) 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 PSRs for the ϕ_0 and ϕ_1 are 1.011, 1.014 and 1.023 with backcasting and 1.008, 1.012 and 1.021 without backcasting.

The corresponding quartiles for the 97.5 percentile points of the PSRs are 1.016, 1.021 and 1.031 and 1.012, 1.019 and 1.032 respectively.

To be conservative, in each run of the Gibbs 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 Gibbs sampler for 2000 iterations and selected every other one to give 1000 "stationary" iterates. For all models 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 Gibbs sampler algorithm performs satisfactorily in all cases studied.

For all models fit to the earnings data, we computed the diagnostics $d_{i,t}$. Figure 2 displays a normal probability plot of the three $d_{i,t}$ from AR(2) models fit to the logged earnings data: the unrestricted, stationary restricted and nonstationary restricted. In this plot, the distributions of $d_{i,t}$ appear reasonably normal. As expected, the $d_{i,t}$ for the unrestricted and nonstationary restricted cases stay within or close to the 95% confidence bands. For the inappropriate stationary restricted case, the values of the diagnostic stray well outside the 95% confidence bands indicating an overestimate in the residual variance.

Normal probability plots for the AR(1) models showed similar patterns.

4.1.2 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

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

where the expectation in the denominator is taken with respect to the stationary restricted model. We let R_{ENS} denote the ratio where 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

$$R_{PE} = |y_{i,n+1} - E(y_{i,n+1}|\mathbf{y},\ \text{Individual})|/|y_{i,n+1} - E(y_{i,n+1}|\mathbf{y},\ \text{Pooled})|. \qquad (21)$$

Finally, we limit our discussion on 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.

Differenced Logged Earnings Data Table 1(a) shows that for the differenced logged earnings data the posterior probabilities of stationarity computed from the Gibbs sampler with backcasting, except for series 8, are all near 1, and thus may be considered stationary.

From Table 2(a)(i), 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 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.

Table 2(a)(ii) shows substantial improvement in the stationary-restricted case when the series are pooled. For ϕ_0 the unpooled posterior standard deviations range from 22%

Table 1
Series Lengths and Posterior Probabilities of Stationarity

a. AR(1) Earnings Data

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Length	31	31	13	18	24	18	18	14	25	13	14	18	14	14
PP_SB	1.00	1.00	.90	.98	.99	.75	.99	.40	.99	.98	.94	.92	.97	.95
PP_SNB	1.00	1.00	.92	1.00	1.00	.77	1.00	.38	1.00	1.00	.98	.95	.99	.96

b. AR(2) Earnings Data

Series	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Length	32	32	14	19	25	19	19	15	26	14	15	19	15	15
PP_SNB	.12	.09	.04	.04	.11	.02	.00	.02	.00	.44	.00	.01	.07	.02

NOTE: PP_SB denotes the posterior probability of stationarity computed with backcasting, PP_SNB denotes the posterior probability of stationarity computed without backcasting.

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

Table 2(b) shows 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, while the unpooled posterior standard deviations of precision range from 8% smaller to 287% greater than the pooled posterior standard deviations.

Table 2(c) 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 (21). 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(c) which show comparable means and prediction errors for the pooled and unpooled cases.

We consider the plots of the estimated bivariate posterior distributions of ϕ_0, ϕ_1. Figure 3 presents these plots for series 6 in the unrestricted and stationary restricted case with and without backcasting. These plots are typical of the plots for the other series. As expected, the plots reveal negative correlation.

We compare the estimates and predictors for the pooled stationary-restricted model with and without backcasting. There is virtually no difference in the estimates and predictors with and without backcasting. Figure 3 shows this for the estimates of ϕ_0 and ϕ_1 for series 6.

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

$$\rho_{i,j} = \{(1 + 1/(\tau_i \psi^2))(1 + 1/(\tau_i \psi^2))\}^{-1/2}. \tag{22}$$

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

Logged Earnings Data Table 1(b) shows that, except for series 10, all series have near zero posterior probabilities of being stationary.

From Table 3(a)(i), 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.

Table 3(a)(ii) compares 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.

From Table 3(b) we see 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, 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) 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 backcasting. 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 backcasting to those without backcasting range from 0.63 for series 5 to 11.2 for series 10. For

ϕ_1 the range is 0.99 to 1.18 and for ϕ_2 the range is from 0.86 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 for both the backcasting and no backcasting cases (for series 14, for example, the actual value is 1.792 while the backcast predictor is 1.808 and the no backcast predictor is 1.815.) The range of ratios of backcast predictors to no backcast predictors is 0.996-1.002. 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 are difficulties involved in estimation and prediction when nonstationary series are backcast.

Finally, we note that the contemporaneous correlations between series i and j, given by (22) are much smaller than those seen in the AR(1) model, ranging from 0.012 to 0.108 for the backcasting case and from 0.041 to 0.144 for the no backcasting case.

4.2 A Small Simulation Study

In order 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 $\theta_0 = 0$, $\tau_i = 100, i = 1, \ldots, 10$, and we took the diagonal elements of the 2×2 matrix Δ to be 0.01 and the off-diagonal element to be 0.005. We varied θ_1 at two levels, 0.5 and 0.8, and ψ^2 at three levels, 0.00, 0.01 and 0.10, corresponding to correlations of 0.00, 0.50 and 0.91 respectively (see (22)). 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.

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

The results of the simulations show little difference in the two levels of θ_1, so we will report the results for $\theta_1 = 0.5$. For all examples, while there is little bias in either pooled or individual estimation of ϕ_1, the pooled and unpooled estimates of ϕ_0, are considerably different from the actual ϕ_0 values. Further, the estimates are substantially worse at the largest ψ^2 level. The accuracy of pooled and individual one-step-ahead predictors is comparable at all factor levels.

To assess the performance of pooled versus individual estimators, for each series we looked at the ratios of the posterior expectations (R_E) and standard deviations (R_S) for the individual versus the pooled for the autoregressive parameters and the one-step-ahead predictors; see Table 4. While the R_E values vary substantially for ϕ_0, for ϕ_1, $R_E \approx 1$. For prediction, there are large differences in R_E. The R_S values indicate that there are substantial gains in precision for pooled estimation of ϕ_0 and ϕ_1. Further these gains increase with decreasing series length and with increasing between-series correlation. In terms of prediction, the picture is more mixed. All series of length 10 have R_S values considerably larger than 1 except one series at $\psi^2 = 0.01$. Further, the gains increase with increasing correlation. However, for series of length 20, 4 series at $\psi^2 = 0.01$ and three series at $\psi^2 = 0.10$ show decreases in precision. Thus the benefits of pooling on one-step-ahead prediction are less clear with longer series.

5 Conclusions

Assuming a hierarchical Bayesian linear model, we have accomplished five main tasks. First, we have shown that it is feasible to use sampling-based methods to analyze stationary autoregressive time series panel data. In particular, we have successfully implemented the Gibbs sampler algorithm to obtain reasonable estimates and forecasts. Second, we have also developed and implemented a Bayesian diagnostic for the difficult problem of assessing model fit of autoregressive time series. Third, we have obtained estimates and forecasts for nonstationary autoregressive series without transforming them to stationary ones. Fourth, 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. Fifth, we have also used latent variables to incorporate 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 is 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. However, restricting nonstationary series to be stationary (or stationary series to be nonstationary) results in biased estimators with artificially low variances.

Table 4
Ranges of R_E and R_S for ϕ_0, ϕ_1
and One-Step-Ahead Prediction for the Twelve Simulated Examples

		Estimation				Prediction	
		ϕ_0		ϕ_1			
Length	ψ^2	R_E	R_S	R_E	R_S	R_E	R_S
10	0.00	0.69-1.82	1.14-1.36	0.99-1.01	1.10-1.22	0.86-1.36	1.10-1.18
	0.01	-0.55-1.73	1.14-3.17	0.98-1.01	1.09-3.10	-0.01-1.69	0.87-1.44
	0.10	0.36-1.17	2.52-8.73	0.99-1.01	2.50-8.46	0.92-1.34	1.11-1.58
20	0.00	0.71-3.16	0.97-1.16	1.00-1.01	1.02-1.14	0.92-1.06	1.01-1.05
	0.01	0.63-1.83	0.92-1.76	1.00-1.01	0.92-1.69	0.81-1.13	0.68-1.20
	0.10	-9.73-1.10	2.66-5.38	0.99-1.00	2.68-5.18	0.89-2.53	0.85-1.17

References

Anderson, T.W. (1978), Repeated measurements on autoregressive processes, Journal of the American Statistical Association 73, 371-378.

Andrews, R.W. (1976), Multiperiod predictions from an autoregressive model using empirical Bayes methods (Graduate School of Business Administration, University of Michigan, Ann Arbor, MI).

Azzalini, A. (1981), Replicated observations of low order autoregressive time series, Journal of Time Series Analysis 2, 63-70.

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

Box, G.E.P. and Tiao, G.C. (1992), Bayesian Inference in Statistical Analysis, New York: Wiley.

Broemeling, L. (1985), Bayesian Analysis of linear models, New York: Marcel Dekker.

Chib, S. and Greenberg, E. (1993), Bayes inference in regression models with ARMA(p,q) errors, Journal of Econometrics (to appear).

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. Bayesian Statistics 4, Eds. J. Bernardo, Oxford University Press, 148-167.

Gelfand A.E. and Smith, A.F.M. (1990), Sampling based approaches for calculating marginal densities 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, 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 B 34, 1-41.

Litterman, R.B. (1980) A Bayesian procedure for forecasting vector autoregressions (Department of Economics, Massachusetts Institute of Technology, Cambridge, MA).

——— (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 (Department of Statistics, University of Connecticut, Storrs, CT).

Müller, P. (1994), Metropolis based posterior integration schemes, (Department of Statistics, Purdue University, West Lafayette, IN).

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

Ravishanker, N., Wu, L.S-Y and Dey, D.K. (1992), Shrinkage estimation for business planning and forecasting of regional IBM revenue (Department of Statistics, University of Connecticut, Storrs, CT).

Robinson, P.M. (1978), Statistical inference for a random coefficient autoregressive model, Scandinavian Journal of Statistics 5, 163-168.

Sims, C.A. (1980), Macroeconomics and reality, Econometrica 48, 1-48.

Tierney, L. (1991), Exploring posterior distributions using Markov chains, in computer science and statistics: Proceedings of the 23^{rd} Symposium on the Interface (E. M. Keramidas, ed.), Interface Foundation, Fairfax Station, Va, 563-570.

Tierney, L. (1994), Markov chains for exploring posterior distributions, Annals of Statistics (to appear).

Table 2
Unpooled to Pooled Ratios of Posterior Means and Standard Deviations for Each Series for Characteristics Associated with the AR(1) Earnings Data

a. Autoregressive Coefficients, ϕ_0, ϕ_1

	i. Means				ii. Standard Deviations			
Series	$R0_{ES}$	$R1_{ES}$	$R0_{ENS}$	$R1_{ENS}$	$R0_{SS}$	$R1_{SS}$	$R0_{SNS}$	$R1_{SNS}$
1	0.97	1.00	-8.18	0.56	1.22	1.23	1.26	3.27
2	0.95	1.02	-6.51	0.65	1.33	1.31	1.51	3.70
3	0.91	1.07	-3.58	0.56	1.70	1.76	1.73	2.14
4	1.04	0.86	-7.43	0.39	1.54	1.55	1.83	3.03
5	1.06	0.85	-19.32	0.37	1.48	1.46	1.98	3.06
6	0.54	1.21	-1.19	0.73	1.48	1.48	1.36	1.44
7	1.10	0.84	-10.17	0.35	1.71	1.65	1.88	3.12
8	0.35	1.42	-2.72	0.91	2.01	2.20	1.45	1.73
9	1.08	0.92	-13.07	0.41	1.55	1.52	2.19	3.44
10	1.40	-0.14	22.24	-0.04	1.80	1.61	0.69	0.60
11	1.27	0.78	-13.81	0.32	1.72	1.72	1.90	2.31
12	1.24	0.95	-12.62	0.47	1.86	1.74	1.90	2.27
13	1.44	0.54	-15.12	0.21	1.78	1.66	1.61	1.71
14	1.07	1.02	-9.15	0.57	1.67	1.55	1.67	2.14

	b. Precision of the Process, τ				c. One-Step Predictor, $y_{i,n+1}$			
Series	R_{ES}	R_{ENS}	R_{SS}	R_{SNS}	R_{ES}	R_{ENS}	R_{SS}	R_{SNS}
1	1.33	1.95	1.39	1.59	0.99	0.82	1.00	0.91
2	1.41	1.97	1.49	1.64	1.00	0.91	1.00	0.87
3	1.24	1.87	1.53	1.85	1.02	0.83	1.01	1.04
4	0.93	1.45	1.09	1.19	0.96	1.07	0.99	1.09
5	0.95	1.45	1.06	1.21	0.94	0.72	0.98	1.10
6	1.98	2.90	2.23	2.87	1.07	0.87	1.02	0.86
7	0.79	1.38	0.95	1.06	0.96	0.79	0.99	1.19
8	0.88	1.16	1.05	1.28	1.30	0.95	1.06	1.48
9	0.95	1.45	1.06	1.19	0.99	0.81	0.99	1.08
10	0.84	1.60	1.09	1.08	1.19	2.41	1.08	1.21
11	1.03	1.71	1.40	1.37	0.95	0.63	0.98	1.19
12	0.90	1.33	1.17	1.13	1.01	0.66	0.98	1.30
13	0.97	1.63	1.24	1.29	0.92	0.63	0.97	1.13
14	1.04	1.65	1.30	1.51	1.04	0.88	1.00	1.07

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

Table 3
Unpooled to Pooled Ratios of Posterior Means and Standard Deviations for Each Series for Characteristics Associated with the AR(2) Earnings Data

a. Autoregressive Coefficients, ϕ_0, ϕ_1, ϕ_2

	i. Means						**ii. 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	0.32	1.04	1.09	0.97	1.01	1.03	1.46	1.05	1.04	1.39	1.33	1.32
2	0.30	1.04	1.07	0.89	1.01	1.02	1.25	0.96	0.96	1.28	1.25	1.25
3	-1.22	0.77	0.25	1.15	0.98	0.81	1.94	1.59	1.71	1.37	1.62	1.60
4	-0.98	0.68	-0.11	1.20	0.86	-0.48	2.50	1.69	1.83	1.44	1.57	1.57
5	0.32	0.93	0.72	1.62	0.99	0.98	1.77	1.33	1.36	1.58	1.56	1.55
6	-0.60	0.84	0.57	1.03	1.05	1.21	1.23	1.57	1.63	0.96	1.28	1.27
7	-1.27	0.61	-0.47	1.16	0.91	1.88	2.23	1.48	1.60	1.40	1.52	1.51
8	-4.72	0.86	0.50	1.71	1.00	0.90	5.26	2.91	3.44	1.87	1.83	1.91
9	0.30	0.70	0.03	1.18	0.96	0.22	1.16	1.27	1.32	1.31	1.42	1.42
10	0.83	0.86	-0.12	-35.03	0.85	-0.21	2.65	1.65	1.73	2.34	1.75	1.75
11	-1.28	0.52	-0.72	1.21	0.80	2.49	2.02	1.50	1.63	1.28	1.49	1.48
12	-0.67	0.58	-0.27	1.30	0.88	7.87	1.81	1.76	1.86	1.25	1.47	1.46
13	-0.33	0.77	0.02	0.82	0.94	0.12	2.57	1.59	1.68	1.71	1.65	1.63
14	-2.47	0.37	-0.82	1.58	0.64	3.32	2.80	2.39	2.66	1.42	1.50	1.50

	b. Precision of the Process, τ				**c. One-Step Predictor, $y_{i,n+1}$**			
Series	R_{ES}	R_{ENS}	R_{SS}	R_{SNS}	R_{ES}	R_{ENS}	R_{SS}	R_{SNS}
1	1.12	0.70	1.33	1.15	1.01	1.00	1.00	0.83
2	1.24	0.67	1.43	1.27	1.00	1.00	1.00	0.76
3	1.75	0.80	1.84	2.23	1.01	1.00	1.00	0.79
4	1.05	0.64	1.08	1.18	1.02	1.00	1.00	1.15
5	0.84	0.62	0.98	0.93	1.01	1.00	1.00	1.04
6	2.39	1.19	2.37	2.84	1.01	1.00	1.00	0.64
7	1.22	0.63	1.06	1.39	1.03	1.00	1.00	0.97
8	1.13	0.71	1.11	1.40	1.02	1.01	1.00	1.16
9	1.21	0.65	1.14	1.27	1.01	1.00	1.00	0.85
10	0.66	0.56	0.80	0.93	1.01	0.99	1.00	1.46
11	1.70	0.89	1.35	2.17	1.01	1.00	1.00	0.82
12	1.25	0.80	1.08	1.51	0.99	1.00	1.00	1.01
13	1.02	0.73	1.03	1.33	1.01	1.00	1.00	1.09
14	1.87	0.88	1.61	2.34	1.04	1.01	1.01	1.03

NOTE: The quantities labelled R_{ES}, R_{ENS}, R_{SS} and R_{SNS} are defined in the note to Table 2.

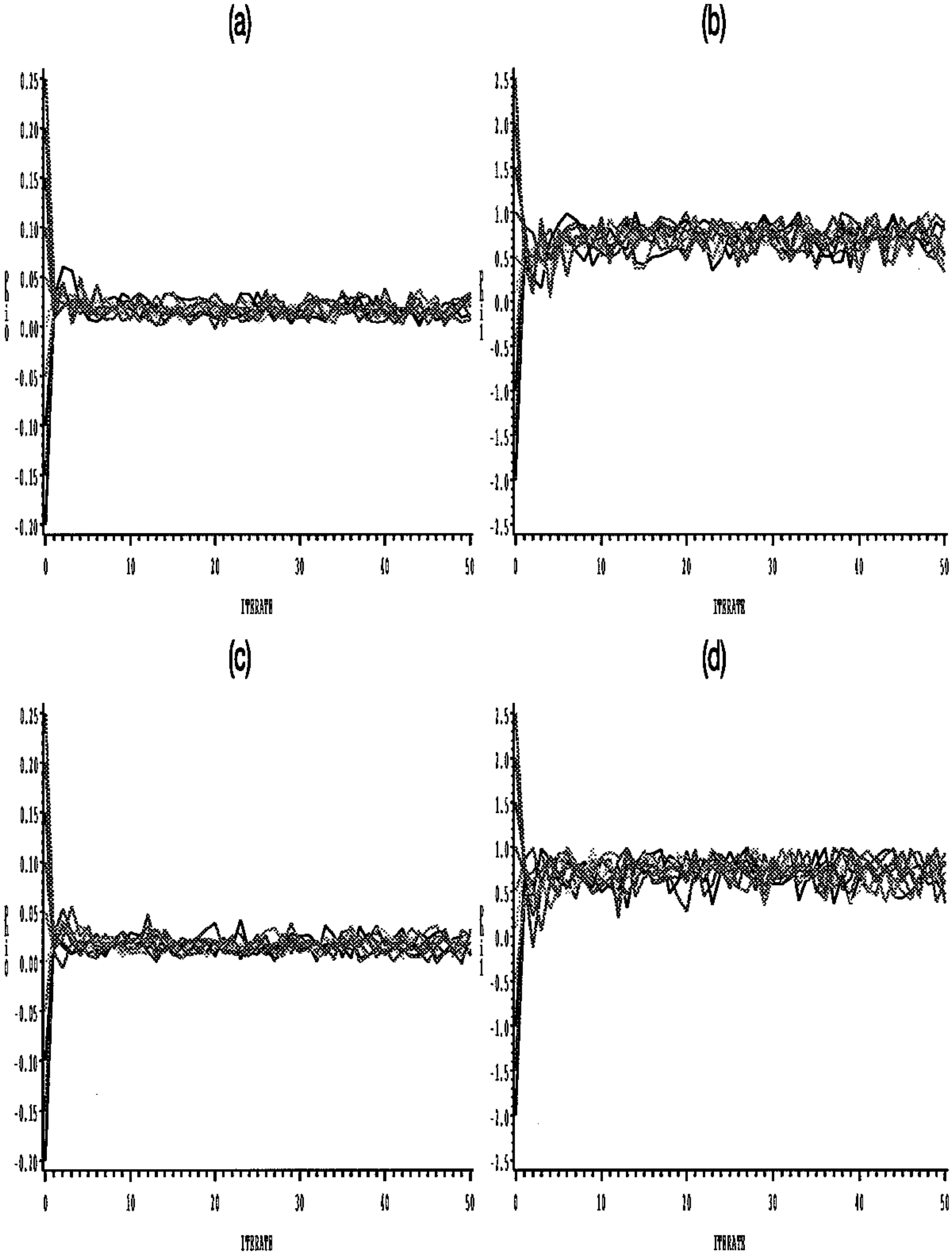

Figure 1: *Trajectories of Ten Sequences for Series 6 of the Logged Earnings Data under the AR(1) Model: (a) ϕ_0 with Backcasting; (b) ϕ_1 with Backcasting; (c) ϕ_0 with No Backcasting; (d) ϕ_1 with No Backcasting;*

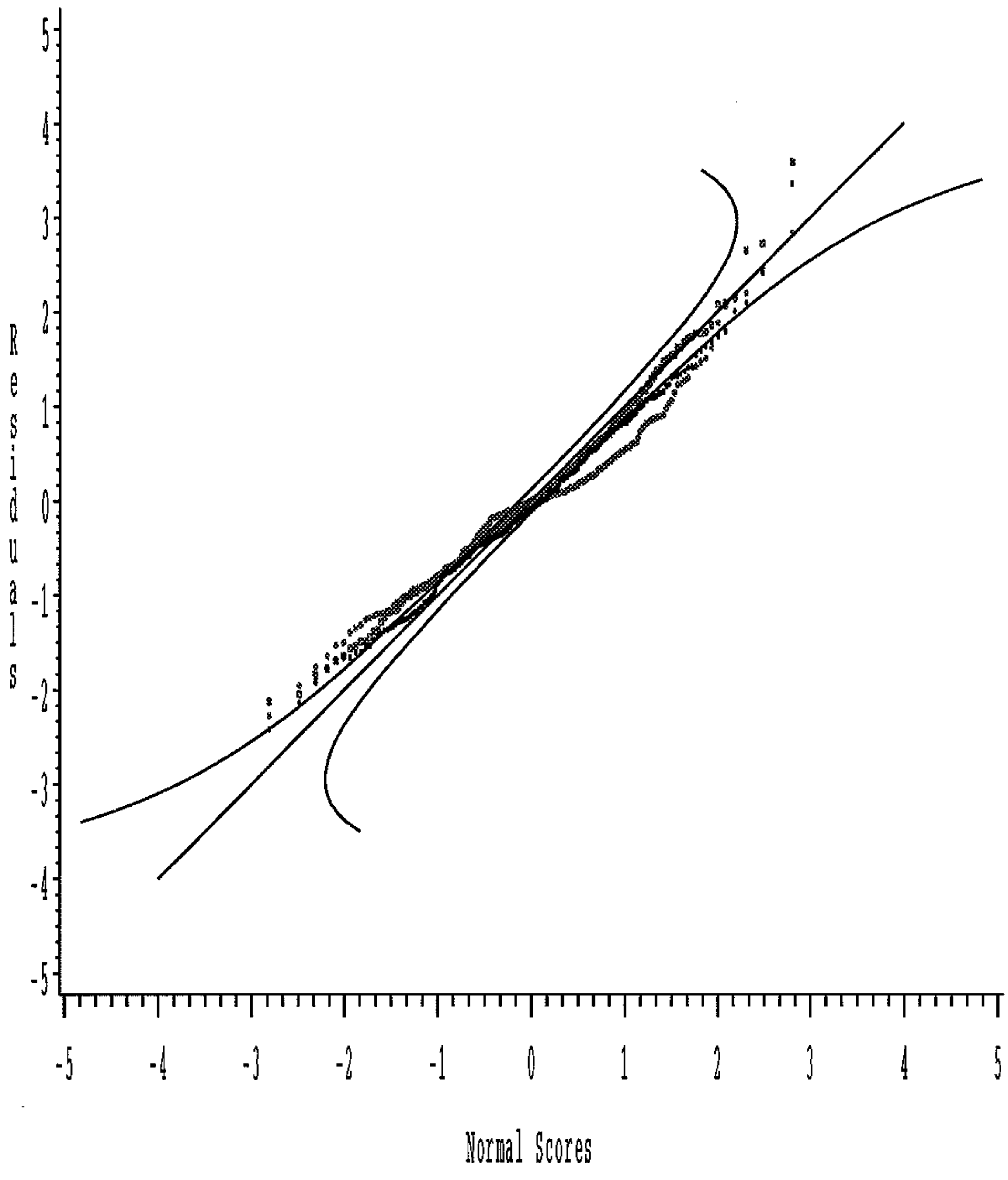

Figure 2: *Normal Probability Plot of $d_{i,t}$ Diagnostic: Logged Earnings Data with Expected 45 Degree Line and 95% Pointwise Critical Bands; Observed Values: Unrestricted (Diamond), Restricted Stationary (Square), Restricted Nonstationary (Plus)*

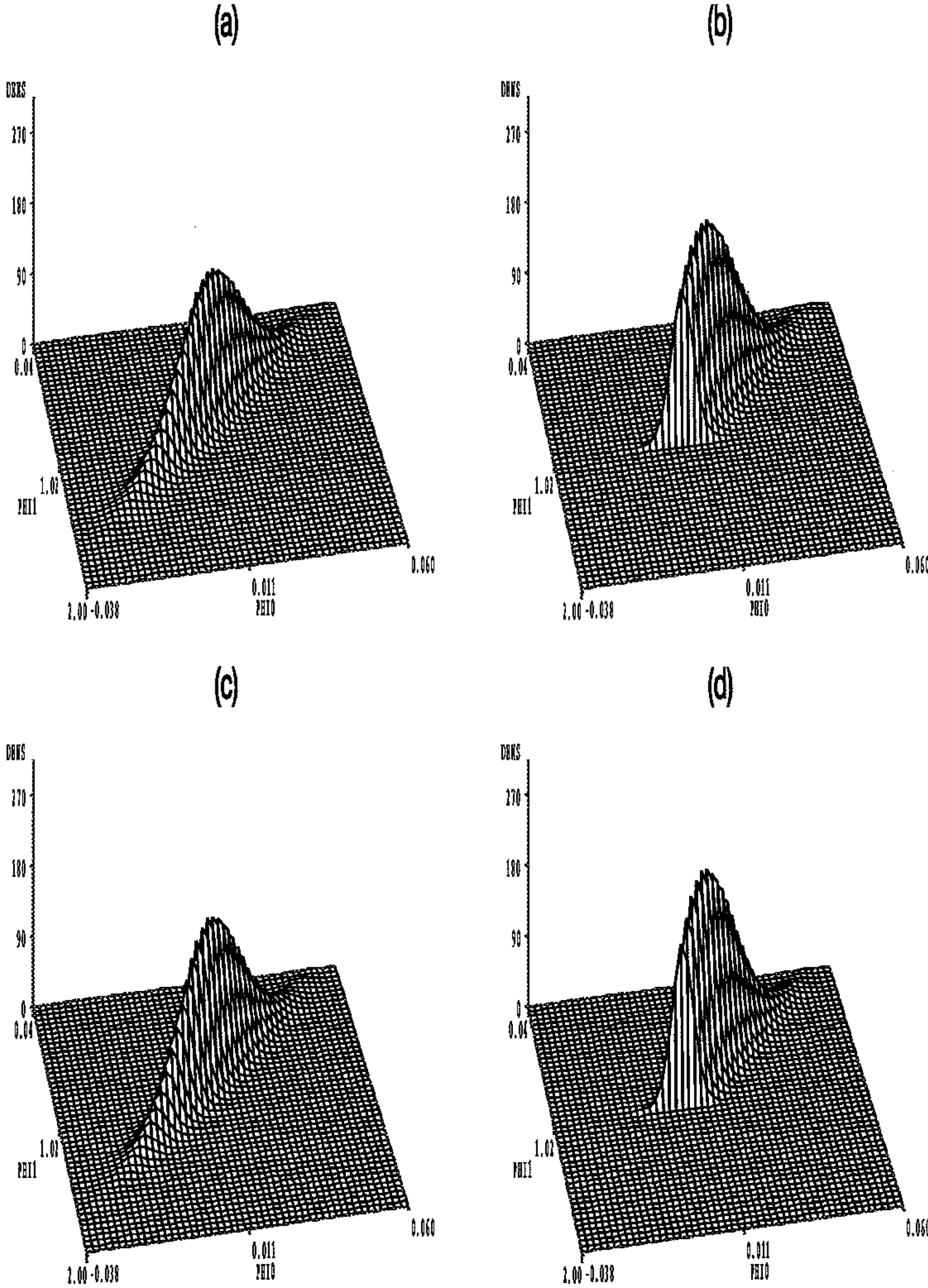

Figure 3: *Bivariate Densities of* ϕ_0 *and* ϕ_1 *for the AR(1) Model Fit to the Logged Earnings Data: (a) Unrestricted with Backcasting; (b) Stationary Restricted with Backcasting; (c) Unrestricted with No Backcasting; (d) Stationary Restricted with No Backcasting*