Bayesian Forecasting for Autoregressive Time Series Panel Data

Balgobin Nandram and Joseph D. Petruccelli
Department of Mathematical Sciences
Worcester Polytechnic Institute
100 Institute Road
Worcester, MA 01609-2280

May 29, 1996

Abstract

We use a Bayesian hierarchical model to study the problem of forecasting several commensurate autoregressive time series, stationary or nonstationary. In particular, we use the Metropolis-Hastings algorithm to obtain k-step ahead forecasts. One advantage of this methodology is the natural ability to compute forecast intervals on the original scale when the series are transformed. We applied our methodology to a set of earnings data from 14 California metropolitan areas. We observed substantial differences in the forecast performance of stationary, as opposed to nonstationary models on these data. Surprisingly, we found that pooling did not give clear evidence of forecast improvement.

KEY WORDS: Diagnostic; Hierarchical model; Metropolis-Hastings; Stationarity.

1. INTRODUCTION

We consider the problem of forecasting 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 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.

Chatfield (1993) reviewed methods for calculating interval forecasts, but had little to say about Bayesian methods. Some references on Bayesian forecasting methods are found in Nandram and Petruccelli (1996).

Ledolter and Lee (1993) also considered the problem of forecasting many short, correlated time series. They extended the Bühlmann-Straub model, which assumes all series vary independently about a fixed level, by allowing the level to change as a random walk over time. They reported gains in forecasting for their model when compared with the Bühlmann-Straub model. However, they did not consider forecasting gains for their model when compared with forecasts for the individual series, as we do in this paper.

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. 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 where the posterior distributions are not analytically tractable. We use the

Metropolis-Hastings (M-H) sampler; see Tierney (1994) for a general description, and Chib and Greenberg (1996) for a tutorial.

Two general models for modeling several series simultaneously are the Bayesian vector autoregressive (BVAR) models (see Litterman, 1986) and the seemingly unrelated regression (SUR) models (see Chib and Greenberg, 1995). Among the difficulties posed by these models are the inability to handle large numbers of series due to the large number of parameters in the covariance of the sampling process, and the complexity of incorporating stationarity or nonstationarity restrictions.

Nandram and Petruccelli (1996), proposed an autoregressive model for several commensurate autoregressive time series, which has neither of these drawbacks. They found that pooling produced large gains in precision when estimating autoregressive parameters. This paper also contains a review of past approaches to modeling such series.

In section 2 of the present paper we present the methodology for multi-step forecasting, and in section 3, we describe the computations necessary to implement the methodology. In section 4 we illustrate the methodology with the analysis of a data set on yearly averages of hourly earnings of production workers in fourteen California metropolitan areas. Section 5 has concluding remarks.

2. METHODOLOGY

We briefly describe a methodology for modeling and forecasting any number of time series of possibly varying lengths.

2.1 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 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), 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 i^{th} series. We also assume the vectors of initial observations $\mathbf{y}_i^{(0)} = (y_{i,t_i+p-1}, y_{i,t_i+p-2}, \ldots, y_{i,t_i})', i = 1, \ldots, m$, and $\mathbf{y}^{(0)} = \left(\mathbf{y}_1^{(0)'}, \ldots, \mathbf{y}_m^{(0)'}\right)'$. 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 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

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

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 $\boldsymbol{\varepsilon}_t'$ 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 parameter ψ^2 in the formulation described in (1) and (2) allows us to model contemporaneous correlations among the series. Specifically, the correlation between series i and j at any given time is

$$\rho_{i,j} = \{(1 + 1/(\tau_i \psi^2))(1 + 1/(\tau_j \psi^2))\}^{-1/2}. \tag{3}$$

This feature is not available when the series are modeled individually.

The autoregressive parameters are modeled as

$$\boldsymbol{\phi}_i \mid \boldsymbol{\theta}, \Delta \overset{iid}{\sim} N(\boldsymbol{\theta}, \Delta). \tag{4}$$

Observe that (4) permits pooling of information across series.

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

$$\boldsymbol{\theta} \sim N(\boldsymbol{\theta}_0, C_0), \tag{5}$$

and a Wishart prior

$$\Delta^{-1} \sim W((\nu_0 \Delta_0)^{-1}, \nu_0),\ \nu_0 \geq p+1 \tag{6}$$

where $\boldsymbol{\theta}_0$ and C_0 in (5) and ν_0 and Δ_0 in (6) 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

$$\tau_1, \tau_2, \ldots, \tau_m \overset{iid}{\sim} G(\eta_0/2, \delta_0/2) \tag{7}$$

and

$$\psi^2 \sim \mathrm{IG}(\zeta_0/2, \beta_0/2) \tag{8}$$

where η_0, δ_0 in (7) and ζ_0, β_0 in (8) are to be specified.

We can restrict the parameters $\boldsymbol{\phi}_i$ in the model to be in the stationary region $\Phi_p = \{\boldsymbol{\phi}_i : \text{series (1) is stationary}\}$, or in its complement. See, e.g., Box and Jenkins (1976) for details on stationarity. For each series, we can compute the posterior probabilities of stationarity $P(\boldsymbol{\phi}_i \in \Phi_p \mid \mathbf{y})$.

To evaluate the adequacy of the models, we compute a multivariate predictive diagnostic in the spirit of Gelfand, Dey and Chang (1992). Defining $\tilde{\mathbf{y}}_t = \{y_{i,t}, i \in I_t\}$, $\tilde{\mathbf{y}}_{(t)} = (\tilde{\mathbf{y}}_1', \tilde{\mathbf{y}}_2', \ldots, \tilde{\mathbf{y}}_{t-1}')'$, $\mathbf{u}_t = E(\tilde{\mathbf{y}}_t \mid \tilde{\mathbf{y}}_{(t)}, \mathbf{y}^{(0)})$ and $\mathcal{V}_t = \mathrm{var}(\tilde{\mathbf{y}}_t \mid \tilde{\mathbf{y}}_{(t)}, \mathbf{y}^{(0)}), t = p+1 \ldots, n$, the diagnostic is

$$\mathbf{d}_t = \mathcal{V}_t^{-1/2}(\tilde{\mathbf{y}}_t - \mathbf{u}_t),\ t = t_0, t_0+1, \ldots, n. \tag{9}$$

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.

2.2 Multi-Step Forecasts

We use the model to predict, $y_{i,n+k}$, the value of the i^{th} series k steps ahead. Let $\mathbf{y}_i = (y_{i,t_i}, y_{i,t_i+1}, \ldots, y_{i,n})'$ 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. Let $\Omega = \{\psi^2, \boldsymbol{\tau}, \boldsymbol{\theta}, \Delta\}$, and let $\mathcal{A}_i(E) = \{(\boldsymbol{\phi}_i, \Omega) : \boldsymbol{\phi}_i \in E\}$ denote the space in which the autoregressive parameters of the i^{th} series are restricted to the set E. Then the posterior distribution of $y_{i,n+k} \mid \mathbf{y}$ with $\boldsymbol{\phi}_i$ restricted to E is

$$p(y_{i,n+k} \mid \mathbf{y}) = \int_{\mathcal{A}_i(E)} p(y_{i,n+k} \mid \mathbf{y}, \boldsymbol{\phi}_i, \Omega) p(\boldsymbol{\phi}_i, \Omega \mid \mathbf{y}) d\boldsymbol{\phi}_i d\Omega. \tag{10}$$

Note that in (10), the predictive distribution $p(y_{i,n+k} \mid \mathbf{y}, \boldsymbol{\phi}_i, \Omega)$ does not depend upon $\boldsymbol{\theta}$ or Δ, and its form is well known (see, e.g., Box and Jenkins, 1976). However, $p(\boldsymbol{\phi}_i, \Omega \mid \mathbf{y})$ is intractable, so we use the M-H algorithm to perform the integration. We obtain $(1-\gamma)100\%$ credible k-step forecast intervals for $y_{i,n+k}$ by calculating the $\gamma/2$ and $1-\gamma/2$ quantiles of the posterior distribution of $y_{i,n+k}$. This computation is facilitated by the Metropolis-Hastings algorithm.

3. COMPUTATIONS

While our primary concern is in obtaining multi-step forecast intervals, we must first fit the model and check the adequacy of that fit. Specifically, we need to estimate $\mathbf{d}_t$ in (9) and estimate the distribution of $y_{i,n+k}$ in (10). Both estimates are obtained by drawing from the posterior distribution $p(\boldsymbol{\phi}, \Omega \mid \mathbf{y})$.

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 $\boldsymbol{\theta}_0$, C_0, ν_0, Δ_0, η_0 and δ_0 using the data. However, we take ζ_0 and β_0 to be zero, and therefore, ψ^2 has a noninformative prior.

3.1 The Metropolis-Hastings Algorithm

In this section, we discuss the implementation of the Metropolis-Hastings 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).

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

$$\varepsilon_{it} \mid \alpha_t, \tau_i \overset{ind}{\sim} N(\alpha_t, \tau_i^{-1}) \tag{11}$$

and

$$\alpha_t \mid \psi^2 \overset{iid}{\sim} N(0, \psi^2) \tag{12}$$

where $t = t_0, \dots, n$ and $i \in I_t$. Let $\boldsymbol{\alpha} = (\alpha_{t_0}, \alpha_{t_0+1}, \dots, \alpha_n)'$ denote the vector of all latent variables in (12).

We run the M-H sampler to obtain M stationary iterates of the $\boldsymbol{\phi}_i$, ψ^2 and $\Omega = \{\boldsymbol{\alpha}, \boldsymbol{\tau}, \boldsymbol{\theta}, \Delta\}$, which we denote by $\boldsymbol{\phi}_i^{(j)}, i = 1, \dots, m$, $\psi^{2(j)}$ and $\Omega^{(j)}$ respectively, $j = 1, 2, \dots M$. In the M-H algorithm $\boldsymbol{\theta}$ and Δ are drawn from their posterior conditional distributions using dependent rejection sampling, and the $\boldsymbol{\phi}_i$ are drawn using independent rejection sampling. The iterates obtained from the M-H algorithm are used for forecasting and model assessment. For a simple illustration, see Chib and Greenberg (1996).

To implement the Metropolis-Hastings 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 $\boldsymbol{\phi}_i$. Define the following quantities for all $i = 1, \dots, m$: $\mathbf{c}_i = \sum_{t=t_i}^n y_{i,t}\mathbf{y}_{i,t-1}$, $G_i = \sum_{t=t_i}^n \mathbf{y}_{i,t-1}\mathbf{y}_{i,t-1}'$, $\Lambda_i = (G_i + (\tau_i\Delta)^{-1})^{-1}G_i$, $\hat{\boldsymbol{\phi}}_i = G_i^{-1}\mathbf{c}_i$, and $\hat{\hat{\boldsymbol{\phi}}}_i = \Lambda_i^{-1}(\mathbf{y}_{i,t_0-1}, \mathbf{y}_{i,t_0-2}, \dots, \mathbf{y}_{i,t_0-p})\boldsymbol{\alpha}$.

Then the conditional posterior distribution $\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega$ is

$$p^{(r)}(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega) = \frac{p(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega)}{\int_{\boldsymbol{\phi}_i \in C^{p+1}} p(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega) d\boldsymbol{\phi}_i}, \boldsymbol{\phi}_i \in C^{p+1}, \tag{13}$$

$i = 1, \ldots, m$, where $p(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega)$ is given by

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

In practice, $\boldsymbol{\phi}_i$ is obtained from $p^{(r)}(\boldsymbol{\phi}_i \mid \mathbf{y}, \Omega)$ by rejection sampling.

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

$$\tau_i \mid \mathbf{y}, \mathbf{y}^{(0)}, \boldsymbol{\alpha}, \boldsymbol{\phi} \overset{ind}{\sim} G((n_i - p + \eta_0)/2, b_i^{-1}) \tag{14}$$

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

$$\psi^2 \mid \boldsymbol{\alpha} \sim \mathrm{IG}((\zeta_0 + n - t_0 + 1)/2, (\beta_0 + \sum_{t=t_0}^{n} \alpha_t^2)/2) \tag{15}$$

and the posterior conditional distribution of α_t is

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

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

Third, we obtain the posterior conditional distributions of $\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta$ and $\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta}$, which 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).

The posterior conditional distribution of $\boldsymbol{\theta}$ given $\boldsymbol{\phi}, \Delta$ is

$$g^{(r)}(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta) \propto \frac{g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)}{[\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta) d\mathbf{z}]^m}, \boldsymbol{\theta} \in \mathbf{R}^{p+1} \tag{17}$$

where $\eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta)$ is the (p+1)-variate $\mathrm{N}(\boldsymbol{\theta}, \Delta)$ density, and $g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$ is

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

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

The posterior conditional distribution of Δ^{-1} given $\boldsymbol{\phi}, \boldsymbol{\theta}$ is given by

$$d^{(r)}(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta}) \propto \frac{d(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})}{[\int_{\mathbf{z} \in C^{p+1}} \eta(\mathbf{z} \mid \boldsymbol{\theta}, \Delta) d\mathbf{z}]^m}. \tag{18}$$

where Δ^{-1} is positive definite and $d(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$ is the Wishart distribution

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

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 To obtain a random vector $\boldsymbol{\theta}$ from $g^{(r)}(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$, we use $g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$ as a proposal density. Thus, the probability of moving from state 1 to state 2 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 $g(\boldsymbol{\theta} \mid \boldsymbol{\phi}, \Delta)$. To obtain the random matrix Δ^{-1} from $d^{(r)}(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$, we use $d(\Delta^{-1} \mid \boldsymbol{\phi}, \boldsymbol{\theta})$ as the proposal density. The probability of moving from state 1 to state 2 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

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

In our application L=1000 suffices. The Metropolis algorithm for drawing $\boldsymbol{\theta}$ is utilized in a similar manner.

The Metropolis-Hastings sampler is implemented in a conceptually simple manner. Using appropriate starting values, one random deviate is drawn from each of (13)-(18) in turn, repeating this process until convergence.

3.2 Model Adequacy and Forecasting

We use the iterates from the Metropolis-Hastings sampler to evaluate model adequacy and compute forecast intervals.

To assess model adequacy, we compute the diagnostics $\mathbf{d}_t$ given in (9). The stationary iterates are used to compute the expectations by using weighted averages, in a manner similar to that outlined by Gelfand, Dey and Chang (1992). For example, letting $\tilde{\boldsymbol{\phi}}_t = \{\boldsymbol{\phi}_i, i \in I_t\}$, we compute the conditional mean $\mathbf{u}_t$, as

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

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.

Forecasting k steps ahead in series i is straightforward. Using (10) an estimate of the posterior distribution $p(y_{i,n+k} \mid \mathbf{y})$ is $M^{-1} \sum_{j=1}^{M} p(y_{i,n+k} \mid \mathbf{y}, \boldsymbol{\phi}_i^{(j)}, \Omega^{(j)})$. The posterior mean and variance are obtained in an obvious manner. We note that the approach to forecasting used here makes it very easy to create forecasts on the original scale when the model is fit to differenced or transformed data.

4. AN APPLICATION: CALIFORNIA EARNINGS DATA

We illustrate our methodology by forecasting one and two steps ahead for a set of fourteen short series. 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 beginning in 1945 and the shortest beginning in 1963. The series lengths are given in column 2 of Table 1. We fit the model to all but the last two observations of each series, reserving the last two observations to assess forecast performance. In principle, our methodology can forecast any number of steps ahead, but because these are short series, we choose to evaluate forecasts of length at most two.

The natural logarithm of each series serves to stabilize variances. 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. Our approach differs from theirs in that we restrict the autoregressive parameters to be stationary. 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 forecast performance of the M-H sampler on nonstationary series, we will also fit an AR(2) to the undifferenced series, restricting the autoregressive parameters to be nonstationary. 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 $x_{i,t} = \ln(y_{i,t})$.

Forecasts were obtained on the original scale using the following methodology. The model given by (19) or (20), (2), and (4)-(8) was fit to the transformed data using the M-H sampler. Samples from the predictive distributions (10) for one and two-step-ahead forecasts were obtained. Finally, estimates of the predictive distributions on the original scale were generated by transforming these samples back to the original scale.

4.1 Computations

Using the conditional posterior distributions (13)-(18), we performed the M-H sampler for both the AR(1) and AR(2) models with multiple runs (Gelman and Rubin, 1992). Specifically, to begin the M-H sampler, we drew ten values of the $\boldsymbol{\phi}_i$ from a dispersed distribution. The M-H sampler was run on each of these ten trajectories.

For both the AR(1) model fit to the differenced logged earnings data and the AR(2) model fit to the logged earnings data, we assessed the convergence of the M-H sampler 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 AR(1) model of differenced logged earnings the quartiles for the PSRs for ϕ_0 and ϕ_1 are 1.008, 1.012 and 1.021. The corresponding quartiles for the 97.5 percentile points of the PSRs are 1.012, 1.019 and 1.032. Plots of the trajectories of the model parameters show rapid convergence.

To be conservative, in each run of the M-H 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 M-H sampler for 4000 iterations. 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 M-H sampler performs satisfactorily in all cases studied.

For both models fit to the earnings data, we computed the diagnostics $d_{i,t}$. Figure 1 displays a normal probability plot of the $d_{i,t}$ from the nonstationary restricted AR(2) model fit to the logged earnings data. In this plot, the distribution of the $d_{i,t}$ appears reasonably normal; almost all the $d_{i,t}$ are within the 95% confidence bands. The normal probability plot for the AR(1) model shows similar patterns.

In order to assess stationarity for the AR(1) model and nonstationarity for the AR(2) model, we ran an unrestricted version of the M-H sampler by taking $\mathbf{C}^{p+1} = \mathbf{R}^{p+1}$; see, for example, (13). The posterior probability of stationarity is then estimated by the proportion of iterates for which $\boldsymbol{\phi} \in \Phi_p$. Column 3 of Table 1 contains the stationarity probabilities for the AR(1) model while column 5 contains the stationarity probabilities for the AR(2) model. Of the former, all but series 8 are virtually stationary, while of the latter, all but series 10 are virtually nonstationary.

We also estimated the correlations ρ_{ij} given in (3) using the iterates from the M-H sampler. Table 1 displays ranges of these estimates for both the AR(1) and AR(2) models. So, for example, for the AR(1) model the estimated correlations of series 1 with the other thirteen series range from .26 to .48. The moderate sizes of these estimates suggest that gains can be realized by modeling contemporaneous correlations.

Having obtained samples from the predictive distributions on the transformed scale, the computations used to estimate the one and two-step-ahead predictive distributions on the original scale were the following. For the AR(2) model, assume the j^{th} iterate from the M-H sampler for the k-step-ahead predictive distribution for series i is $\hat{x}_{i,n+k}^{(j)}, k = 1, 2$. Then the transformation back to the original scale yields iterates $\hat{y}_{i,n+k}^{(j)} = \exp\{\hat{x}_{i,n+k}^{(j)}\}, j = 1, \ldots, M$. For the AR(1) model, assume the j^{th} iterate from the M-H sampler for the k-step-ahead predictive distribution for series i is $\hat{z}_{i,n+k}^{(j)}, k = 1, 2$. Then the transformation back to the original scale yields the iterates $\hat{y}_{i,n+1}^{(j)} = y_{i,n}\exp\{\hat{z}_{i,n+1}^{(j)}\}$ and $\hat{y}_{i,n+2}^{(j)} = \hat{y}_{i,n+1}^{(j)}\exp\{\hat{z}_{i,n+2}^{(j)}\}, j = 1, \ldots, M$. The endpoints of the $100(1-\gamma)\%$ forecast credible intervals are computed as the $\gamma/2$ and $1-\gamma/2$ quantiles of the M iterates on the original scale.

4.2 Forecasting

We have two main objectives in our study of forecasting performance on the California earnings data: (a) to compare the forecasts obtained by pooling with those obtained from each individual series, and (b) to compare forecasts in original units obtained from the nonstationary AR(2) model fit to the logged data with those obtained from the stationary AR(1) model fit to the differenced logged data.

Here, forecasts obtained from individual series refer to those obtained from either (19) or (20) together with (2) for $\psi^2 = 0$, and the noninformative prior specification $p(\boldsymbol{\phi}_i, \tau_i) \propto \tau_i^{-1}$ only. That is, no "borrowing of strength" is allowed.

Let $\hat{E}(y_{i,n+k} \mid \mathbf{y})$ denote an estimate of the posterior mean of $y_{i,n+k}$ obtained from the M-H iterates. Let $R_{I,S}$ denote the relative forecast error

$$(\hat{E}(y_{i,n+k} \mid \mathbf{y}) - y_{i,n+k})/y_{i,n+k},$$

where the subscript I signifies individual series and the subscript S signifies the stationary restricted AR(1) model. Similarly, $R_{I,N}$, $R_{P,S}$ and $R_{P,N}$ denote this same quantity for individual, nonstationary restricted, pooled stationary restricted and pooled nonstationary restricted respectively. In Table 2 we summarize the distributions of these ratios.

For both horizons, $R_{I,N}$ and $R_{P,N}$ are very similar, as are $R_{I,S}$ and $R_{P,S}$. However, both $R_{I,N}$ and $R_{P,N}$ have more values greater than 0.05 for horizon 2 than for horizon 1. All forecasts have relative errors of magnitude at most 11%. Among the poorest performers at horizon 1 are some of the shortest series: 4, 8, 10, 11 and 12, while the longest series, 1, 2, 5 and 9, are among the best performers. This general pattern holds for horizon 2 as well with the exception of series 8. While the one-step ahead forecast for series 8 severely underpredicts when the stationary model is used, the two-step-ahead forecast for series 8 severely overpredicts when the nonstationary model is used. We believe this is connected with the low posterior probability of stationarity for series 8 under the AR(1) model; see Table 1.

Let $W_{S:\,I,P}$ denote the ratio of the width of the 95% forecast interval for individual series (i.e., no pooling) to the width of the 95% forecast interval for pooled series when the stationary model is used. Let $W_{N:\,I,P}$ denote the same quantity for the nonstationary model. Table 3a displays the $W_{S:\,I,P}$ and $W_{N:\,I,P}$ for the earnings data. Similarly, let $W_{I:\,N,S}$ denote the ratio of the width of the 95% forecast interval for the nonstationary model to the width of the 95% forecast interval for the stationary model for individual series, and $W_{P:\,N,S}$ the same ratio for pooled series. Table 3b displays the $W_{I:\,N,S}$ and $W_{P:\,N,S}$ for the earnings data.

In Table 3a, we see that there is little difference between $W_{S:\,I,P}$ and $W_{N:\,I,P}$ at horizon 1, but at horizon 2 substantial differences emerge: for example, series 1 and 2, the two longest series, have $W_{S:\,I,P}$ much less than 1 and $W_{N:\,I,P}$ slightly greater than 1. Table 3b shows that the $W_{I:\,N,S}$ and $W_{P:\,N,S}$ tend to be less than 1 for horizon 1 and greater than 1 for horizon 2. None of these ratios lie in (0.99, 1.01) for horizon 2, indicating more extreme behavior between nonstationary and stationary forecasts for the longer horizon.

We estimated the probability content of an individual 95% forecast interval relative to pooled forecasts as follows. Having obtained the individual forecast interval, we computed point forecasts for the pooled model for each M-H iterate. The probability content is estimated as the proportion of these pooled forecasts that fall in the individual forecast interval. For stationary series we denote this quantity by P_S, and for nonstationary series we denote it by P_N.

Table 4 shows that 95% forecast intervals for individual series can have quite different coverage probabilities when these intervals are evaluated using the pooled models. For all but two or three series in each case the estimated coverage probabilities of the individual series evaluated using the pooled model lies more than 0.5% above or below the nominal 95% level. Some differences are extreme. In particular the coverages for series 14 for the nonstationary case are 84.5% and 75.7% for horizons 1 and 2, coverages for series 1 for the stationary case are 89.1% and 89.8%, and for series 6, coverages for the stationary case are 83.8% and 84.9% and for the nonstationsry case 85.1% and 89.3%. In addition, we note that there is similarity in patterns of behavior of the interval widths shown in Table 3 and the probability content observed in Table 4. For example, series 1 and 2 show low values of $W_{S:\,I,P}$ and P_S and high values of $W_{N:\,I,P}$ and P_N.

Table 5 presents 95% forecast intervals on the original scale for four selected series. Series 2 is a long series which is virtually stationary with the AR(1) model and nonstationary with the AR(2) model. Series 14 is a short series with the same stationarity properties. Series 8 and 10 are short series. The posterior probability of stationarity for series 8 is only 0.59 for the AR(1) model, though it is virtually nonstationary for the AR(2) model. The posterior probability of nonstationarity for series 10 is 0.25 for the AR(1) model, though the series is virtually stationary for the AR(2) model. The differences between individual and pooled intervals are small, but there are substantial differences between stationary and nonstationary intervals, particularly at horizon 2.

Finally, we computed the correlations, $\hat{\varrho}_{i,j}(k) = \text{corr}(y_{i,n+k}, y_{j,n+k} \mid \mathbf{y})$ using the M-H iterates. For the stationary model, the correlations range from 0.11 to 0.47 for horizon 1 and are slightly smaller for horizon 2. These moderate correlations suggest that pooling might be beneficial in forecasting using the stationary model. However, for the nonstationary model the correlations do not differ substantially from zero, suggesting that there may be little benefit from pooling using the nonstationary model.

5. CONCLUSIONS

Our main objective has been to use Bayesian methods to obtain forecast intervals for time series panel data. We have accomplished this objective for both stationary and nonstationary series by using a hierarchical Bayesian model via the Metropolis-Hastings algorithm. Further, we developed sampling-based diagnostics to study model fit.

We applied the methodology to California earnings data, using both a nonstationary model fit to the logs of the earnings and a stationary model fit to the differences of those logs. Diagnostics showed a reasonable fit for both models. One advantage of our sampling-based approach is that we were able to compute forecast intervals on the original scale in a natural way.

We found that in terms of relative forecast error individual and pooled forecasts perform similarly for the stationary model and for the nonstationary model at both horizons. However, the forecast is biased upwards at each horizon, but especially at horizon 2, for the nonstationary model for both pooled and individual forecasts. Nonstationary intervals tend to be wider than their stationary counterparts at horizon 2. For stationary intervals there is similarity between individual and pooled interval widths at both horizons, but individual widths tend to be greater than pooled widths for nonstationary series at horizon 2. At horizon 1, stationary pooled interval widths tend to be greater than their nonstationary counterparts. The probability contents of the forecast intervals for individual series are generally different from the nominal value of 95% when evaluated using the pooled series.

One area for further research is the construction of optimal simultaneous multi-step forecast intervals.

References

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

Chatfield, C., 'Calculating interval forecasts' (with discussion), *Journal of Business & Economic Statistics*, **11**, (1993), 121-143.

Chib, S. and Greenberg, E., 'Hierarchical analysis of SUR models with extensions to correlated serial errors and time-varying parameter models', *Journal of Econometrics*, **68**, (1995), 339-360.

———, 'Understanding the Metropolis-Hastings algorithm', *The American Statistician*, **49**, (1996), 327-335.

Gelfand, A. E., Dey, D. K. and Chang, H., 'Model determination using predictive distributions with implementation via sampling-based methods', in *Bayesian Statistics 4*, ed. J. Bernardo: Oxford University Press, 148-167, 1992.

Gelfand A.E., Smith, A.F.M. and Lee, T.M., 'Bayesian analysis of constrained parameter and truncated data problems using Gibbs sampler', *Journal of the American Statistical Association*, **87**, (1992), 523-530.

Gelman, A. and Rubin, D. B., 'Inference from iterative simulation using multiple sequences' (with discussion), *Statistical Science*, **7**, (1992), 457-472 & 483-511.

Ledolter, J. and Lee, C-S., 'Analysis of many short time sequences: forecast improvements achieved by shrinkage', *Journal of Forecasting*, **12**, (1993), 1-11.

Lindley, D. V. and Smith, A. F. M., 'Bayes estimates for the linear model', *Journal of the Royal Statistical Society*, Ser. B, **34**, (1972), 1-41.

Litterman, R. B., 'Forecasting with Bayesian vector autoregressions-five years' of experience', *Journal of Business & Economic Statistics*, **4**, (1986), 25-38.

Liu, L. and Tiao, G. C., 'Random coefficient first-order autoregressive models', *Journal of Econometrics*, **13**, (1980), 305-325.

Nandram, B., and Petruccelli, J. D., 'Bayesian analysis of autoregressive time series panel data', *Journal of Business & Economic Statistics*, in press.

Tierney, L., 'Markov chains for exploring posterior distributions' (with discussion), *Annals of Statistics*, **22**, (1994), 1701-1762.

Table 1: *Characteristics of the 14 series: series length, posterior probability of stationarity, and range of estimated correlations between each series and all other series*

		AR(1)		AR(2)	
Series	Length	Prob	Range	Prob	Range
1	33	1.00	.26-.48	0.05	.18-.39
2	33	1.00	.28-.48	0.03	.20-.39
3	15	0.97	.23-.42	0.01	.17-.35
4	20	1.00	.20-.36	0.01	.14-.29
5	26	1.00	.21-.39	0.05	.14-.30
6	20	0.89	.25-.46	0.00	.19-.39
7	20	1.00	.20-.35	0.00	.14-.30
8	16	0.59	.18-.32	0.01	.12-.25
9	27	1.00	.23-.42	0.00	.17-.36
10	15	1.00	.16-.28	0.25	.10-.20
11	16	0.99	.18-.31	0.00	.13-.26
12	20	0.98	.16-.30	0.00	.11-.24
13	16	1.00	.16-.29	0.02	.10-.21
14	16	0.98	.20-.36	0.00	.15-.32

Table 2: *Distribution of the relative errors of the point forecasts by horizon and type of model (individual–stationary, pooled–stationary, individual–nonstationary or pooled–nonstationary)*

		Ranges of Relative Errors				
Horizon	Measure	$(-0.11, -0.05]$	$(-0.05, -0.01]$	$(-0.01, 0.01]$	$(0.01, 0.05]$	$(0.05, 0.09)$
1	$R_{I,S}$	8,12,11	13,6,5,1	3,7,9,2	14,10,4	–
	$R_{P,S}$	12,8	11,13,6,1	5,3,2	7,9,14,10,4	–
	$R_{I,N}$	12,8	11,13,6	1,5,3	9,7,2,14	10,4
	$R_{P,N}$	12	8,11,13,6	1,5,3	9,7,2,14	10,4
2	$R_{I,S}$	12,13	6,11,1,8,5	9,3	7,2,14,4,10	–
	$R_{P,S}$	12,13	6,11,1,8	5,9,3	2,14,7,4,10	–
	$R_{I,N}$	12	13,6,1	–	11,5,9,8,3	2,7,14,10,4
	$R_{P,N}$	12	13,6,1	11	5,9,8,3,2	14,7,10,4

NOTE: Entries are series identifiers; see Table 1. In each row of the table the series identifiers are in order of increase of the measure for that row.

Table 3: *Distribution of ratio of forecast interval widths by horizon*

		Ranges of Ratios				
Horizon	Measure	$(0.53, 0.94]$	$(0.94, 0.99]$	$(0.99, 1.01]$	$(1.01, 1.05]$	$(1.05, 1.45)$
a. Individual versus pooled by stationarity						
1	$W_{S:\,I,P}$	6,2,1,11	14,12	13,3	4	9,5,10,7,8
	$W_{N:\,I,P}$	6,14,11,2,1	13	4	12,3	9,7,8,10,5
2	$W_{S:\,I,P}$	6,1,11,2	12,13,14	3,5	10,4	9,7,8
	$W_{N:\,I,P}$	14,6,11	13	–	2,4,1,8	12,7,3,9,10,5
b. Nonstationary versus stationary by degree of pooling						
1	$W_{I:\,N,S}$	14,7,9,4,8	6,3	11	2,13,5,12	1,10
	$W_{P:\,N,S}$	7,9,14,4,2	3,1,5,6,11	8,10,12	13	–
2	$W_{I:\,N,S}$	14,7,9	4,8	–	11	6,3,12,13,5,2,1,10
	$W_{P:\,N,S}$	14,9,7	4,6,3	–	11,12,5	8,1,2,10,13

NOTE: Entries are series identifiers; see Table 1. In each row of the table the series identifiers are in order of increase of the measure for that row.

Table 4: *Distribution of probability content of individual 95% forecast intervals with respect to forecasts for the pooled model, by horizon and stationarity*

		Ranges of Probability Contents				
Horizon	Measure	$(0.75, 0.84]$	$(0.84, 0.90]$	$(0.90, 0.945]$	$(0.945, 0.955]$	$(0.955, 0.99)$
1	P_S	6	1,2	11,14,3	12,4,13	5,9,10,7,8
	P_N	–	14,6	2,11,1	13,4	12,3,9,7,8,10,5
2	P_S	–	6,1	11,2,12,14,13	3,5	4,10,9,7,8
	P_N	14	6,11	2,13	8,3	1,4,12,7,9,10,5

NOTE: Entries are series identifiers; see Table 1. In each row of the table the series identifiers are in order of increase of the measure for that row.

Table 5: *Comparison of 95% forecast intervals for selected series by horizon, pooling and stationarity*

		Individual		Pooled	
Series	Horizon	Stationary	Nonstationary	Stationary	Nonstationary
2	1	(6.36, 6.65)	(6.45, 6.82)	(6.32, 6.75)	(6.41, 6.80)
	2	(6.65, 7.46)	(6.82, 7.93)	(6.57, 7.52)	(6.74, 7.84)
8	1	(5.06, 5.81)	(5.16, 5.86)	(5.19, 5.78)	(5.26, 5.85)
	2	(5.20, 6.84)	(5.59, 7.21)	(5.34, 6.75)	(5.55, 7.10)
10	1	(4.15, 4.73)	(4.21, 4.84)	(4.21, 4.73)	(4.26, 4.78)
	2	(4.25, 5.19)	(4.29, 5.66)	(4.29, 5.21)	(4.36, 5.49)
14	1	(5.30, 5.82)	(5.48, 5.83)	(5.28, 5.82)	(5.38, 5.87)
	2	(5.54, 6.80)	(6.16, 6.84)	(5.48, 6.78)	(5.86, 6.89)

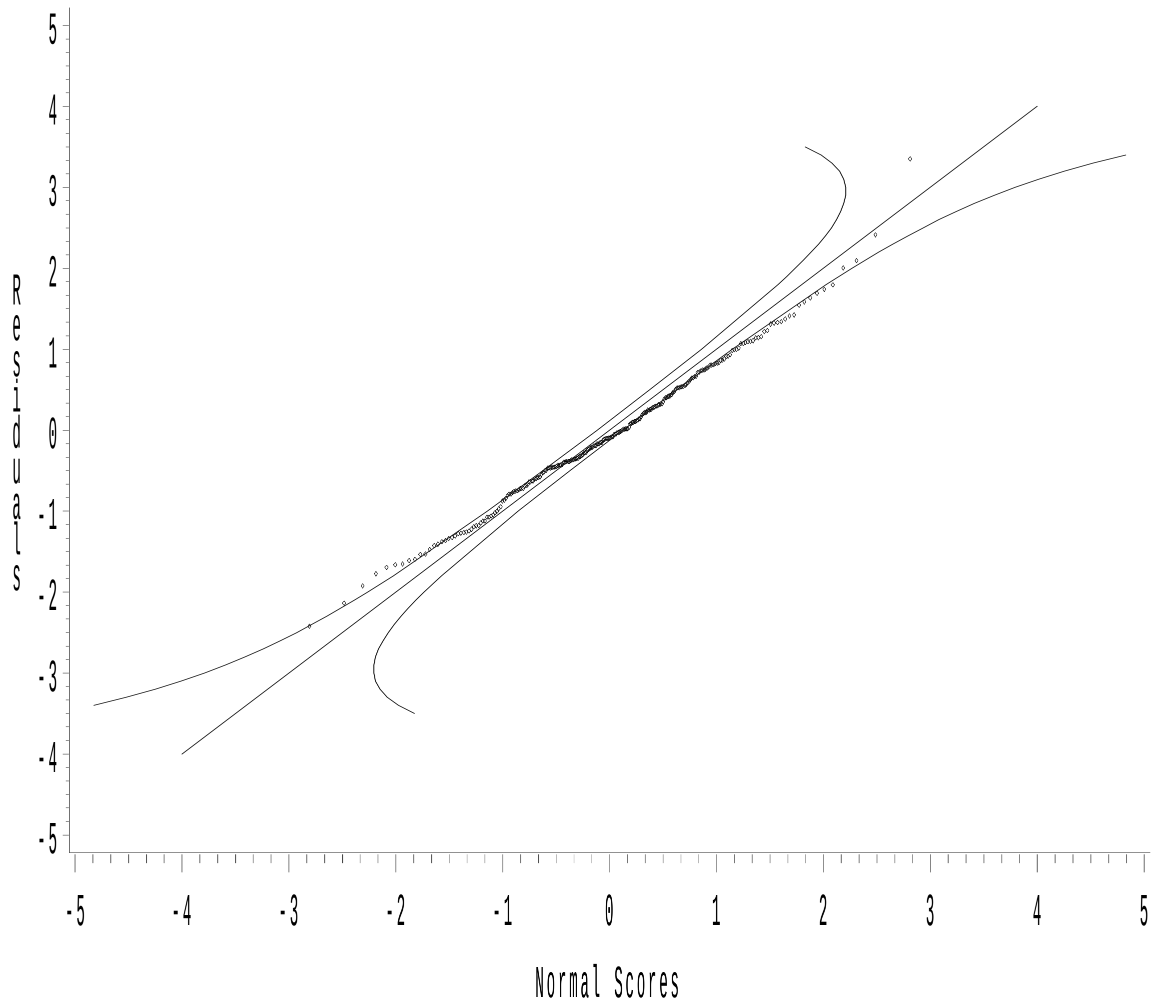

Figure 1: *Normal probability plot of the* $d_{i,t}$ *from the nonstationary restricted AR(2) model fit to the logged earnings data, with 45 degree line and 95% pointwise critical bands*