# Bayesian Analysis of Random Coefficient AutoRegressive Models

**Dazhe Wang**

Eli Lilly and Company, Indianapolis, IN 46285 (wang_dazhe@lilly.com)

**Sujit K. Ghosh**

Department of Statistics, North Carolina State University, Raleigh, NC 27695 (sghosh2@stat.ncsu.edu)

**Abstract** - Institute of Statistics Mimeo Series #2566

Random Coefficient AutoRegressive (RCAR) models are obtained by introducing random coefficients to an AR or more generally ARMA model. These models have second order properties similar to that of ARCH and GARCH models. In this article, a Bayesian approach to estimate the first order RCAR models is considered. A couple of Bayesian testing criteria for the unit-root hypothesis are proposed: one is based on the Posterior Interval, and the other one is based on Bayes Factor. In the end, two real life examples involving the daily stock volume transaction data are presented to show the applicability of the proposed methods.

KEY WORDS: Bayes Factor; Gibbs sampling algorithm; Markov Chain Monte Carlo method; Posterior interval; Stationarity parameter; Unit-root testing.

# 1. INTRODUCTION

In finance it is very common to see that the mean corrected return on holding an asset has time dependent variation (i.e. volatility). Researchers have been interested in modeling the time dependent feature of unobserved volatility. A model that is commonly used to model such features, is known as the AutoRegressive Conditional Heteroscedastic (ARCH) model (see Engle, 1982). In an ARCH(1) model, the volatility at time $t$ is expressed in terms of the mean corrected return at time $t-1$. This is similar in spirit to an AR(1) model, the difference being unobserved volatilities are modeled instead of the actual observed series. An extension is to consider Generalized ARCH (GARCH) models (see Bollerslev, 1986), which have been found to be very popular to model the volatility over time. For example, in a GARCH(1,1) model, the volatility at time $t$ is modeled as function of the volatility at time $t-1$ as well as the square of the mean corrected return at time $t-1$. This is similar in spirit to an ARMA(1,1) model. It is desirable to obtain volatility models that preserve the nice properties of ARMA models.

An alternative way to model unobserved volatility is to use the so-called Conditional Heteroscadatic AutoRegressive Moving Average (CHARMA) model (see Tsay, 1987), which allows the coefficients of an ARMA model to be random. Note that the CHARMA models have second-order properties similar to that of ARCH models. However the presence of random coefficients makes it harder to obtain the higher order properties of CHARMA models as compared to that of ARCH models. Another class of models that are very similar in spirit to CHARMA models is known as the Random Coefficient AutoRegressive (RCAR) models studied in detail by Nicholls and Quinn (1982). Historically an RCAR model has been used to model the conditional mean of a time series, but it can also be viewed as a volatility model. Nicholls and Quinn, henceforth, NQ, studied the RCAR models from a frequentist view point. In particular NQ obtained the stationarity condition of the RCAR model for vector valued time series. In case of univariate time series data NQ obtained the Least Square (LS) estimates of the model parameters and showed that under suitable conditions the LS estimates are strongly consistent and obey the central limit theorem. In addition, by assuming the normality of the random coefficients and the error, NQ proposed an iterative method to obtain Maximum Likelihood (ML) estimates of the parameters. They showed that the ML estimates are strongly consistent and satisfy a central limit theorem even when the error processes are not normally distributed.

We consider in this article the Bayesian approach to analyze RCAR models. Bayesian data analysis is becoming more and more appealing because of its flexibility in handling complex models that typically involve many parameters and/or missing observations. The RCAR model can be viewed as a three-level hierarchical model. In the first level of hierarchy, the conditional distribution of the data is specified given the unobserved random coefficients and the parameters. At the second level, the conditional distribution of the unobserved coefficients is specified given the parameters. Finally at the last level, prior densities of the parameters are specified. Prior densities in Bayesian inference play a crucial role. Suitably chosen non-informative priors often yields estimators with good frequentist properties (see Datta and Mukerjee 2004). We find the model building procedure to be quite flexible from a Bayesian perspective but find it more convenient to validate its use from a frequentist perspective. We also illustrate how our proposed methods can be implemented in practice using the software `WinBUGS` (Speigelhalter et al., 2001) available free on the web. From a Bayesian perspective we do not have to consider prior densities with support only on the stationarity region, which is in contrast with the frequentist approach where the asymptotic theory works under the assumption that the RCAR(1) is strictly stationary and ergodic (see Wang, 2003). We propose two Bayesian test criteria to test the unit-root hypothesis and compare their performance using simulation studies.

The rest of the article is organized as follows: Section 2 describes the RCAR model and some alternative assumptions concerning this model. Section 3 describes the parameter estimation procedure and how to implement the Gibbs sampling algorithm via `WinBUGS`. Two model selection criteria are then proposed and their performance evaluated through simulation. Section 4 consists of the discussion of two unit-root testing procedures and the evaluation of their performance through simulation. We apply our proposed methods to a pair of real life stock volume transaction data in Section 5. Finally, in Section 6, we indicate some directions for further research in RCAR models.

# 2. THE RANDOM COEFFICIENT AUTOREGRESSIVE (RCAR) MODELS

For financial time series, it is very common to see that the return of an asset $x_t$ does not have a constant variance, and highly volatile periods tend to be clustered together. In another words, there is a strong dependence of sudden bursts of variability in a return on the series' own past. The conditional variance structure of the RCAR model can pick up this feature of the financial time series. The RCAR models can be described in its simplest form as follows,

$$\begin{aligned} x_t &= \alpha + \sum_{j=1}^{p} \beta_{tj} x_{t-j} + \sigma \epsilon_t, \\ \underline{\beta}_t &= \underline{\mu}_\beta + \Sigma_\beta^{1/2} \underline{u}_t, \end{aligned} \tag{1}$$

where $\underline{\beta}_t = (\beta_{t1}, \ldots, \beta_{tp})^T$ and $\underline{\mu}_\beta = (\mu_{\beta 1}, \ldots, \mu_{\beta p})^T$. It is assumed that $\epsilon_t$'s and $\underline{u}_t$'s are sequences of iid realizations from a distribution with zero mean and unit variance. The parameters $\alpha, \underline{\mu}_\beta, \sigma$ and $\Sigma_\beta$ are to be estimated based on a suitable criterion. In this article the $\underline{u}_t$'s and $\epsilon_t$'s are also assumed to be independent. However several extensions of the above model are possible. For illustrative purposes we will only consider the case with $p = 1$ in the rest of the article. For this simplified case, with slight abuse of notation, we will write the model (1) as,

$$\begin{aligned} x_t &= \alpha + \beta_t x_{t-1} + \sigma \epsilon_t, \\ \beta_t &= \mu_\beta + \sigma_\beta u_t \end{aligned} \tag{2}$$

For the p$^{th}$ order RCAR process (1), Nicholls and Quinn (1982) derived the necessary and sufficient condition for the process to be second-order stationary. In particular, for the RCAR(1) model (2), the condition becomes $\mu_\beta^2 + \sigma_\beta^2 < 1$. For convenience, define $\eta = \mu_\beta^2 + \sigma_\beta^2$ and call $\eta$ the stationary parameter for an RCAR(1) model. We say that an RCAR(1) model has a unit root if $\eta = 1$.

To illustrate the implications of a unit-root, we generate four RCAR(1) series according to Model (2). Figure 1 shows 500 observations for each of the four series by assuming the joint normality of $\beta_t$ and $\epsilon_t$. See the top of each plot to find the specific values of the pair $(\eta, \mu_\beta)$ that generated different series. Notice that only Series 1 is second-order stationary, Series 2 and 3 have unit root in the sense that $\eta = 1$, and Series 4 is a random walk.

For Series 1, it is seen from Figure 1 that it tends to oscillate around its mean value of 0 with approximately finite and equal amount of variability. For Series 2 and 3, they also tend to oscillate around its mean 0, but the variability could be very large due to sudden bursts in values. For the random walk case, the series

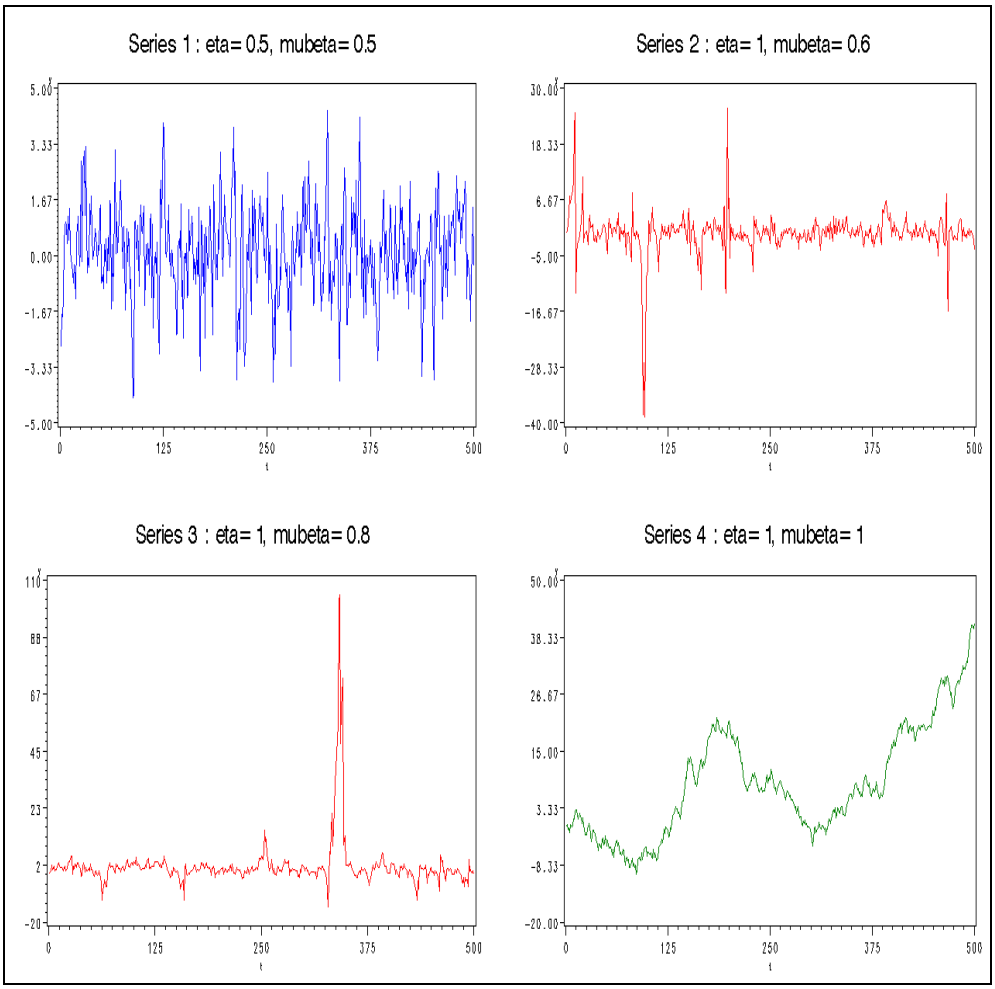


Figure 1: Comparison of RCAR(1) series with $\eta < 1$ and $\eta = 1$ (n=500 observations). The series were generated by fixing $\alpha = 0$ and $\sigma^2 = 1$.

seems to be free to wander, and there is no tendency for the series to be clustered around any fixed value.

Empirically, we can see that when an RCAR(1) series is second-order stationary, i.e., $\eta = \mu_\beta^2 + \sigma_\beta^2 < 1$, the series is well-behaved in terms of predictability. On the other hand, when the RCAR(1) has a unit root, the series tends to be unpredictable up to the second order (see Section 4 for some theoretical justification). For a financial time series, being predictable is a desirable feature. Ideally, if a stock return is perfectly predictable, then we can make money by buying low and selling high. So it is of great importance to test the unit-root hypothesis of a financial time series against the hypothesis that it is second-order stationary. That is, we

would be interested in developing tests for the null hypothesis $H_0 : \eta = 1$ versus the alternative $H_1 : \eta < 1$. In the literature, continuous densities such as uniform, truncated normal, and beta distributions that are defined over the interval (0,1) have been used as prior densities for $\eta$, see Liu & Tiao (1980) and Liu (1995) for some examples. For estimation purpose such densities might be useful, but for testing the unit-root hypothesis, any continuous distribution with support solely on $(0, 1)$ interval can produce results biased in favor of stationarity. For this reason, we need to broaden the support and consider prior densities that either assigns a positive mass at unity or a density that is defined on an interval that includes the unity (say $(0, a)$, with $a > 1$).

# 3. PARAMETER ESTIMATION FOR RCAR

To perform Bayesian data analysis for the RCAR(1) model, it helps to set up the model as a three-level hierarchical model. At the first level of hierarchy, the conditional distribution of the data $x_t$'s is specified given the unobserved random coefficients $\beta_t$'s and the parameter $\theta = (\alpha, \mu_\beta, \sigma^2_\beta, \sigma^2)^T$; at the second level, the conditional distribution of $\beta_t$'s is specified given the parameter $\theta$; and finally at the last level, the prior distribution of $\theta$ is given. Throughout this section, we will assume that $\beta_t$'s and $\epsilon_t$'s are iid normal, along with the assumptions that they are mutually independent. Consequently, given a sample $x_0, x_1, \ldots, x_n$, we are able to express the RCAR(1) model in the following hierarchical structure,

$$\begin{aligned} x_0|\alpha, \sigma^2 &\sim N(\alpha, \sigma^2) \\ x_t|x_{t-1}, \beta_t, \alpha, \sigma^2 &\sim N(\alpha + \beta_t x_{t-1}, \sigma^2) \text{ for } t \geq 1 \\ \beta_t|\mu_\beta, \sigma^2_\beta &\sim N(\mu_\beta, \sigma^2_\beta) \\ (\alpha, \mu_\beta, \sigma^2_\beta, \sigma^2) &\sim p(\alpha, \mu_\beta, \sigma^2_\beta, \sigma^2) \end{aligned} \tag{3}$$

where $p(.)$ is the prior density of $\theta$ which reflects our prior beliefs about the unknown parameters. From a Bayesian perspective we do not have to consider prior densities with support only on the stationarity region i.e., $\mu^2_\beta + \sigma^2_\beta < 1$. It is well known that suitably chosen non-informative priors can produce Bayes estimates that enjoy several optimality properties even from a frequentist perspective (see Berger (1985) for some theoretical justifications). Following (3), we can express

likelihood function of $\theta$ as,

$$L(\theta) = L(\theta|x_0, x_1, \ldots, x_n) = \phi(x_0, \alpha, \sigma) \prod_{i=1}^{n} \phi(x_t; \alpha + \mu_\beta x_{t-1}, \sqrt{\sigma^2 + \sigma_\beta^2 x_{t-1}^2}) \quad (4)$$

where $\phi(x; \mu, \sigma)$ denotes the density function of a normal distribution with mean $\mu$ and standard deviation $\sigma$. Therefore, the joint posterior density of the parameters is given by,

$$f(\theta|x_0, x_1, \ldots, x_n) \propto L(\theta)p(\theta).$$

At this point, we only know that the posterior joint density of $\theta$ given the observed data $x_t$'s is proportional to $L(\theta)p(\theta)$. It is usually not an easy task to find the normalizing constant of such joint posterior densities. This problem becomes even harder when we want to evaluate the marginal density of a lower dimensional parameter, say $\mu_\beta$, that is, we have to integrate out the above density with respect to $\alpha, \sigma_\beta^2$ and $\sigma^2$,

$$f(\mu_\beta|x_0, x_1, \ldots, x_n) \propto \int f(\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2|x_0, x_1, \ldots, x_n) d\alpha d\sigma_\beta^2 d\sigma^2.$$

As a possible remedy, we propose to use the so-called Markov Chain Monte Carlo (MCMC) methods. MCMC methods consist of algorithms to construct a Markov chain of the parameters such that its stationary distribution is the posterior distribution of the parameters. Hence, under certain regularity conditions, the realization of the Markov chain can be thought of as approximate values sampled from the posterior distribution of $\theta$ given the $x_t$'s. We will carry out the Gibbs sampler, see Gelfand and Smith (1990), a widely used MCMC method, to obtain dependent samples from the posterior distribution using the software `WinBUGS`.

## 3.1 Choice of Prior Densities

In practice, often no reliable information concerning $\theta = (\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)^T$ exists *apriori*, or an inference solely based on the available data is desired. In another words, we need to find a prior distribution which contains 'no information' about $\theta$ in the sense that it doesn't favor one value of $\theta$ over another. This kind of prior distribution is known as '*non-informative*'. By using a non-informative prior, all of the information we obtain from the posterior $f(\theta|x_0, x_1, \ldots, x_n)$ of $\theta$ arises only from the data, and hence all resulting inferences are relatively 'objective'.

For the parameter estimation of $\theta = (\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)^T$ in RCAR(1) model, we use $N(\mu_0, V_0)$ as non-informative prior for $\alpha$ and $\mu_\beta$, and $IG(a, b)$ as non-informative

prior for $\sigma_\beta^2$ and $\sigma^2$, where $\mu_0 = 0$, $V_0 = 10^6$, $a = 2 + 10^{-10}$ and $b = 0.1$. Notice that for $\alpha$ and $\mu_\beta$, we use the normal priors, and we set the variances of the normal distributions to be large, i.e. $10^6$. We employ the inverse gamma priors for $\sigma_\beta^2$ and $\sigma^2$. The variance of a random variable having $IG(a, b)$ is $\frac{b^2}{(a-1)^2(a-2)}$ with $a > 2$. So by choosing $a = 2 + 10^{-10}$ and $b = 0.1$, we can make the variance of the Inverse Gamma distribution large (i.e. $10^8$) hence make the prior non-informative.

## 3.2 Model Selection

A problem associated with any parametric model is that it may not be robust against model misspecification. Also given two competing models, how do we know which one fits the data better? In order to assess the model performance, it is customary to use some type of model selection criteria such as AIC and BIC. We consider in this article the commonly used AIC and a criterion proposed by Gelfand and Ghosh (1998) to choose between models. It is well known that AIC is defined as,

$$AIC(\theta) = -2lnL(\theta) + 2p \tag{5}$$

where $lnL(\cdot)$ is the log-likelihood function, $p$ is the number of parameters in the model, e.g., for an RCAR(1) model $p = 4$ whereas for an AR(1) model $p = 3$. However for models with random effects (such as unobserved coefficients) it is not clear if the penalty function should depend only on $p$ as defined above. Given the wide popularity of this selection criterion we plan to investigate its performance using simulation. Notice that AIC as defined in Equation (5) is a function of $\theta$ and hence we may compute the posterior distribution of AIC. Graphically one may plot the entire posterior density of AIC for different competing models and choose the model that has significantly lower value of the posterior median than the other competing models.

The Gelfand and Ghosh criteria (GGC) is calculated based on posterior predictive distribution of a future return. This criterion can be used to compare across many models including non-nested ones. To see this, let us define $\underline{x}^p = (x_1^p, x_2^p, \ldots, x_n^p)$ as predicted data generated from the following posterior predictive distribution,

$$p(\underline{x}^p|\underline{x}) \;=\; \int p(\underline{x}^p|\theta)p(\theta|\underline{x})d\theta \tag{6}$$

where $p(\underline{x}^p|\theta)$ is the likelihood function evaluated at $\underline{x}^p$, and $p(\theta|\underline{x})$ is the posterior distribution of parameter $\theta$ given $\underline{x}$. Next define the Mean Squared Pre-

dicted Errors (MSPE) as $MSPE = \frac{1}{n}\sum_{j=1}^{n}(x_j^p - x_j)^2$. Using MSPE as the loss function to measure the discrepancy between $\underline{x}$ and $\underline{x}^p$, the GGC is defined as $GGC = E(MSPE|\underline{x})$, where the expectation is taken with respect to the predictive distribution defined in Equation (6). As with AIC, we prefer a model with smaller value of GGC. In Section 3.4 we will study the performance of GGC in selecting different models by simulation.

## 3.3 Posterior Inference via Gibbs Sampling

In Bayesian estimation for RCAR(1) model, we are interested in the properties of the posterior density of $\theta = (\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)^T$. Deriving the joint posterior density for $\theta$ amounts to integrating out the unobserved coefficients $\beta_t$'s which can be difficult to perform both analytically and numerically when the unobserved coefficients are not normally distributed. This makes the posterior inference about the parameters very difficult. However for normal errors, we can perform the integration analytically and obtain the likelihood in a closed form as in (4). Even in this case deriving the joint posterior density of $\theta$ is analytically intractable. So we appeal to MCMC methods. MCMC methods are being increasingly used for the cases where marginal distribution of the random parameters can not be obtained either analytically or by numerical integration. MCMC methods consist of sampling random variates from a Markov chain, such that its stationary distribution is the posterior distribution of the parameter of interest (see Tierney, 1994, for some excellent theoretical properties of MCMC). That is, under some regularity conditions, the realizations of this Markov chain (after some "burn-in" time, say $m$) can be thought of as realizations sampled from the desired posterior distribution.

We use the Gibbs sampler to obtain dependent samples from the posterior distribution $p(\theta|\underline{x})$. To implement the Gibbs sampler one obtains the full conditional densities, i.e. the conditional density of a component of the vector of parameters given the other components and the observed data. Specifically, we derive the conditional densities, $f(\alpha|\mu_\beta, \sigma_\beta^2, \sigma^2, \underline{x})$, $f(\mu_\beta|\alpha, \sigma_\beta^2, \sigma^2, \underline{x})$, $f(\sigma_\beta^2|\alpha, \mu_\beta, \sigma^2, \underline{x})$, and $f(\sigma^2|\alpha, \mu_\beta, \sigma_\beta^2, \underline{x})$ as the full conditional densities of $\alpha$, $\mu_\beta$, $\sigma_\beta^2$ and $\sigma^2$, respectively based on Model (2). The full conditional densities of $\alpha$ and $\mu_\beta$ turn out to be Normal densities. However for the other two parameters, their full conditional densities are non-standard and can not be sampled easily.

Given an arbitrary choice of starting values for the parameters (for example, the LS estimates proposed by NQ) at the beginning of the algorithm, say, $\alpha^{(0)}$,

$\mu_\beta^{(0)}$, $\sigma_\beta^{2(0)}$ and $\sigma^{2(0)}$, the Gibbs sampling algorithm for RCAR(1) model is given by: Initialize $\mu_\beta^{(0)}, \sigma_\beta^{2(0)}$ and $\sigma^{2(0)}$. For $k = 1, 2, \ldots, m + M$,

1. Draw $\alpha^{(k)}$ from $f(\alpha|\mu_\beta^{(k-1)}, \sigma_\beta^{2(k-1)}, \sigma^{2(k-1)}, \underline{x})$
2. Draw $\mu_\beta^{(k)}$ from $f(\mu_\beta|\alpha^{(k)}, \sigma_\beta^{2(k-1)}, \sigma^{2(k-1)}, \underline{x})$
3. Draw $\sigma_\beta^{2(k)}$ from $f(\sigma_\beta^2|\alpha^{(k)}, \mu_\beta^{(k)}, \sigma^{2(k-1)}, \underline{x})$
4. Draw $\sigma^{2(k)}$ from $f(\sigma^2|\alpha^{(k)}, \mu_\beta^{(k)}, \sigma_\beta^{2(k)}, \underline{x})$

where $m$ is burn-in and $M$ is the number of samples generated after burn-in. Repeating the above sampling steps, we obtain a discrete-time Markov chain $\{(\alpha^{(k)}, \mu_\beta^{(k)}, \sigma_\beta^{2(k)}, \sigma^{2(k)}); k = 1, 2, \ldots\}$ whose stationary distribution is the joint posterior density of the parameters.

We carry out the Gibbs sampling algorithm as proposed above by means of a software package known as `WinBUGS`. In `WinBUGS`, a Markov chain of the parameters $\alpha$ and $\mu_\beta$ is constructed by direct sampling from the corresponding full conditional densities as they are standard distributions. However, the full conditional densities of $\sigma_\beta^2$ and $\sigma^2$ are not standard and their logarithms do not belong to the class of log-concave densities. In this case, `WinBUGS` employs Adaptive Rejection Metropolis Sampling (ARMS) method to sample from such distributions. For more information on ARMS and other sampling methods implemented in `WINBUGS`, see Gilks et al. (1994, 1996) and Spiegelhalter et al (2001).

The above algorithm can be easily modified to obtain samples under a different prior specification. The Ergodic Theorem for Markov chains, guarantees that the random variates $\theta^{(k)} = (\alpha^{(k)}, \mu_\beta^{(k)}, \sigma_\beta^{2(k)}, \sigma^{2(k)})^T, k = 1, 2, \ldots$ from the above Markov chain converges in distribution to the posterior distribution $f(\theta|\underline{x})$. However in order to determine the burn-in time $m$, i.e., the time after which the samples start coming approximately from the posterior distribution, we use the software `CODA` available in `S-Plus` and `R` versions (see Best et al., 1995 on the use of this free ware). It may be noted that `WinBUGS` also has some features like trace-plots, autocorrelation plots and Gelman-Rubin diagnostics to diagnose approximate convergence. For more details on the use of `CODA` to diagnose convergence of the chains see our remarks in the next section.

## 3.4 A Simulation Study

We conducted an extensive simulation study to explore the frequentist properties of the proposed Bayesian estimation method. We present only a few significant findings from our simulation experiments.

In order to explore different properties of the RCAR(1) models, we designed an experiment to simulate the data from the Model (2) by using various sets of parameter values. To reduce the number of combinations, for each data set throughout our simulation experiments, we fixed the true value of the intercept $\alpha = 0$ and variance of the shocks $\sigma^2 = 1$. Various combinations of $\mu_\beta$ and $\sigma^2_\beta$ were used (see Table 1). Notice that only Case 1 is weakly stationary and satisfies $\mu^2_\beta + \sigma^2_\beta < 1$. The other cases explore different choices of unit-root hypothesis. In particular, Case 8 represents a random walk series. We used the non-informative prior densities for $\theta$ as specified in Section 3.1.

The simulation was carried out using `SAS` and `WINBUGS` together; `SAS` was used to generate data and the Gibbs sampling was performed via `WINBUGS`. The resulting output from `WINBUGS` was then analyzed in `SAS` to report the Monte Carlo summary values. In order to determine suitable values of $m$ (size of burn-in) and $M$ (sample size after burn-in) we used the initial plots (e.g. trace and autocorrelation plots) of the chains. Using three dispersed starting values for $\theta$, we obtained three parallel chains and plotted their traces to see the mixing property of the chains. We also looked at the autocorrelation plots from three starting values to check the rate of convergence. We checked the numerical summary values as suggested by Gelman and Rubin (1992) and Raftery and Lewis (1992) to diagnose the convergence. All these diagnostics were performed using the `S-Plus` version of `CODA`. Based on these initial plots and diagnostics we decided to use $m = 1000$ and $M = 5000$ throughout our simulation.

Table 1 provides the Monte Carlo averages of the posterior mean, standard deviation for each parameter based on sample size of n=100 and 500. It also gives the nominal coverage probabilities for a 95% intervals based on the equal tail posterior intervals.

It appears from Table 1 that the parameters are well estimated in most of the cases with very good nominal coverage probabilities even for small sample size such as $n = 100$. For moderate sample sizes, for example n=500, the estimates and the coverage probabilities get even better. It may be noted that for the last case (with $\sigma^2_\beta = 0$), the nominal coverage probability is 0 because the prior distribution for

Table 1: Monte Carlo average of posterior summaries for different RCAR(1) models (sample size n=100 and n=500, 500 MC replications).

| | | n=100 | | | n=500 | | |
|---|---|---|---|---|---|---|---|
| | true | mean | s.d. | cp | mean | s.d. | cp |
| Case 1 | $\alpha = 0$ | 0.00 | 0.11 | 0.95 | 0.00 | 0.05 | 0.94 |
| | $\mu_\beta = 0.5$ | 0.48 | 0.11 | 0.91 | 0.50 | 0.05 | 0.97 |
| | $\sigma^2 = 1$ | 1.13 | 0.23 | 0.93 | 1.02 | 0.10 | 0.93 |
| | $\sigma^2_\beta = 0.25$ | 0.19 | 0.14 | 0.87 | 0.25 | 0.06 | 0.95 |
| Case 2 | $\alpha = 0$ | 0.01 | 0.14 | 0.94 | 0.00 | 0.06 | 0.95 |
| | $\mu_\beta = 0.6$ | 0.58 | 0.12 | 0.95 | 0.59 | 0.06 | 0.96 |
| | $\sigma^2 = 1$ | 1.07 | 0.28 | 0.93 | 1.01 | 0.12 | 0.94 |
| | $\sigma^2_\beta = 0.64$ | 0.64 | 0.21 | 0.92 | 0.65 | 0.08 | 0.94 |
| Case 3 | $\alpha = 0$ | 0.04 | 0.36 | 0.74 | 0.01 | 0.11 | 0.87 |
| | $\mu_\beta = 0.995$ | 0.94 | 0.05 | 0.73 | 0.99 | 0.01 | 0.87 |
| | $\sigma^2 = 1$ | 1.05 | 0.23 | 0.94 | 1.01 | 0.11 | 0.93 |
| | $\sigma^2_\beta = 0.01$ | 0.01 | 0.01 | 0.97 | 0.01 | 0.00 | 0.95 |
| Case 4 | $\alpha = 0$ | 0.02 | 0.15 | 0.94 | 0.00 | 0.06 | 0.96 |
| | $\mu_\beta = 0.1$ | 0.09 | 0.14 | 0.97 | 0.09 | 0.06 | 0.97 |
| | $\sigma^2 = 1$ | 1.08 | 0.30 | 0.94 | 1.02 | 0.12 | 0.95 |
| | $\sigma^2_\beta = 0.99$ | 1.00 | 0.29 | 0.93 | 0.99 | 0.11 | 0.94 |
| Case 5 | $\alpha = 0$ | 0.01 | 0.14 | 0.96 | 0.00 | 0.06 | 0.96 |
| | $\mu_\beta = -0.6$ | -0.60 | 0.13 | 0.94 | -0.60 | 0.06 | 0.94 |
| | $\sigma^2 = 1$ | 1.09 | 0.28 | 0.94 | 1.01 | 0.11 | 0.96 |
| | $\sigma^2_\beta = 0.64$ | 0.63 | 0.20 | 0.94 | 0.65 | 0.08 | 0.97 |
| Case 6 | $\alpha = 0$ | 0.01 | 0.11 | 0.96 | 0.00 | 0.05 | 0.96 |
| | $\mu_\beta = -0.995$ | -0.98 | 0.03 | 0.95 | -0.99 | 0.01 | 0.97 |
| | $\sigma^2 = 1$ | 1.09 | 0.23 | 0.94 | 1.01 | 0.10 | 0.96 |
| | $\sigma^2_\beta = 0.01$ | 0.01 | 0.01 | 0.98 | 0.01 | 0.00 | 0.94 |
| Case 7 | $\alpha = 0$ | 0.01 | 0.15 | 0.94 | 0.01 | 0.06 | 0.95 |
| | $\mu_\beta = -0.1$ | -0.11 | 0.14 | 0.96 | -0.10 | 0.06 | 0.96 |
| | $\sigma^2 = 1$ | 1.10 | 0.30 | 0.93 | 1.02 | 0.11 | 0.96 |
| | $\sigma^2_\beta = 0.99$ | 0.98 | 0.29 | 0.93 | 0.99 | 0.11 | 0.96 |
| Case 8 | $\alpha = 0$ | 0.04 | 0.37 | 0.69 | 0.01 | 0.18 | 0.63 |
| | $\mu_\beta = 1$ | 0.95 | 0.04 | 0.66 | 0.99 | 0.01 | 0.59 |
| | $\sigma^2 = 1$ | 0.94 | 0.15 | 0.91 | 0.94 | 0.07 | 0.86 |
| | $\sigma^2_\beta = 0$ | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 | 0.00 |

$\sigma_\beta^2$ doesn't put a mass at zero and hence the 95% posterior interval never contains 0. However the posterior mean still provides reasonably good point estimate.

Next we turn to model misspecification. It is our goal to see how the model selection criteria defined in Section 3.2 perform in picking up the correct model when a wrong parametric model is fitted. To see this we simulated data 500 times from, say, an RCAR(1) model and fit both RCAR(1) and AR(1) (by assuming $\sigma_\beta^2 = 0$) models and computed the AIC and GGC to see if the criteria have a lower value when an RCAR(1) model is fit. We repeated the same procedure when the data is generated from an AR(1) model instead. Again we fixed $\alpha = 0$ and $\sigma^2 = 1$ and tried different values for the pair $(\mu_\beta, \sigma_\beta^2)$. Table 2 provides the MC average values of AIC and GGC for data of sample size $n = 500$ for several selected cases. It appears that both AIC and GGC are performing reasonably well in picking up the correct model when the data are generated from RCAR(1). It is also evident that GGC performs better in picking up the correct model as seen from the larger difference in GGC values between RCAR(1) fit and AR(1) fit than that in AIC. When the data are generated from AR(1), AIC prefers AR(1) fit to RCAR(1) fit, but the difference in AIC values between the two fits is not substantial. On the other hand, GGC prefers RCAR(1) fit to AR(1) fit for AR(1) data.

In Table 2, we also list the posterior means and standard deviations from fitting RCAR(1) and AR(1) to both RCAR(1) and AR(1) data. Based on this table, it is evident that RCAR models are robust to model misspecification. This can be seen from the following aspects:

(i) The estimators of $\alpha$ and $\mu_\beta$ are very robust to model misspecification. For instance, when the true data is generated from AR(1) with parameters $\mu_\beta = 0.5$ and fitted as RCAR(1), the posterior means for $\alpha$ and $\mu_\beta$ are 0.00 and 0.50 respectively. This is conceivable because the conditional means of both RCAR(1) and AR(1) models are the same.

(ii) The estimate of $\sigma^2$ based on AR(1) when data generated from RCAR(1) is biased. For example, the posterior mean of $\sigma^2$ is 1.51 whereas the true value of $\sigma^2$ is 1 (for the case $\mu_\beta = 0.5$). If we fit RCAR(1) model for AR(1) data, the estimate of $\sigma^2$ seems to be unbiased (=0.97). This is because AR(1) is a special case of RCAR(1).

(iii) For $\sigma^2$, the efficiency loss against AR(1) misspecification is much greater than the efficiency loss against RCAR(1) misspecification. For example, for

Table 2: Posterior summary for fitting RCAR(1) and AR(1) when data are generated from RCAR(1) and AR(1) (sample size n=500, 500 MC replications).

| DGP | RCAR(1) | | | | AR(1) | | | |
|---|---|---|---|---|---|---|---|---|
| FIT | RCAR(1) | | AR(1) | | RCAR(1) | | AR(1) | |
| | mean | s.e. | mean | s.e | mean | s.e. | mean | s.e. |
| $\alpha = 0$ | 0.00 | 0.05 | 0.00 | 0.06 | 0.00 | 0.05 | 0.00 | 0.04 |
| $\mu_\beta = 0.5$ | 0.50 | 0.05 | 0.50 | 0.05 | 0.48 | 0.04 | 0.50 | 0.04 |
| $\sigma^2_\beta = 0.25$ | 0.25 | 0.06 | | | 0.03 | 0.02 | | |
| $\sigma^2 = 1$ | 1.02 | 0.10 | 1.51 | 0.18 | 0.97 | 0.06 | 1.01 | 0.06 |
| AIC | 1577.58 | | 1621.54 | | 1422.37 | | 1420.05 | |
| GGC | 1015.28 | | 1511.20 | | 972.28 | | 1004.42 | |
| $\alpha = 0$ | 0.00 | 0.06 | 0.01 | 0.15 | 0.00 | 0.05 | 0.01 | 0.04 |
| $\mu_\beta = 0.6$ | 0.59 | 0.06 | 0.53 | 0.11 | 0.58 | 0.04 | 0.60 | 0.04 |
| $\sigma^2_\beta = 0.64$ | 0.65 | 0.08 | | | 0.02 | 0.01 | | |
| $\sigma^2 = 1$ | 1.01 | 0.12 | 12.54 | 63.38 | 0.97 | 0.06 | 1.01 | 0.06 |
| AIC | 1865.48 | | 2311.60 | | 1422.35 | | 1420.02 | |
| GGC | 1002.83 | | 12518.86 | | 972.59 | | 1004.81 | |
| $\alpha = 0$ | 0.01 | 0.11 | 0.01 | 0.15 | -0.02 | 0.24 | -0.01 | 0.18 |
| $\mu_\beta = 0.995$ | 0.99 | 0.01 | 0.98 | 0.01 | 0.98 | 0.01 | 0.98 | 0.01 |
| $\sigma^2_\beta = 0.01$ | 0.01 | 0.00 | | | 0.00 | 0.00 | | |
| $\sigma^2 = 1$ | 1.01 | 0.11 | 3.00 | 8.75 | 0.16 | 0.29 | 1.19 | 0.27 |
| AIC | 1672.37 | | 1777.31 | | 1434.94 | | 1431.42 | |
| GGC | 1006.75 | | 2994.68 | | 1155.03 | | 1183.73 | |

the case with $\mu_\beta = 0.995$, the relative efficiency for AR(1) misspecification is $\frac{8.75}{0.11} = 79.5$ whereas the relative efficiency for RCAR(1) misspecification is $\frac{0.29}{0.27} = 1.07$.

(iv) The AIC increase is not substantial when RCAR(1) is fitted to AR(1) data, whereas the AIC increase substantially when AR(1) is fitted to RCAR(1) data. For the case with $\mu_\beta = 0.5$, the AIC increase by $1422.37 - 1420.05 = 2.32$ due to RCAR(1) misspecification as compared to the AIC increase due to AR(1) misspecification which is $1621.54 - 1577.58 = 43.96$.

# 4. UNIT-ROOT TESTING PROCEDURES FOR RCAR

## 4.1 Motivation for Unit-root Testing in RCAR(1)

One of the important problems in financial time series analysis is to forecast the future observations of a stock price series based on the realizations already observed. Consider the forecasting for an AR(1) process. When a stationary process is inadequately modeled as a unit-root process, the long term prediction interval will be wider compared to the one obtained from when it is correctly modeled. Similar situation happens for RCAR(1) models. In this section, we will consider the impact of a unit root in RCAR(1) models by means of examining the conditional variance of prediction errors.

Suppose at time $t$ we have $t+1$ observations $x_0, x_1, x_2, \ldots, x_t$ from a series an RCAR(1) series. For simplicity, let us assume the true model parameter values $\theta = (\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)^T$ are known. The Minimum Mean Square Error (MMSE) 1-step-ahead predictor based on $x_0, x_1, x_2, \ldots, x_t$ is $\hat{x}_{t+1} = E(x_{t+1}|x_t) = \alpha + \mu_\beta x_t$ and the conditional variance is given by $v_{t+1|t} = Var(x_{t+1}|x_t) = \sigma_\beta^2 x_t^2 + \sigma^2$. For $j \geq 2$, the $j$-step-ahead MMSE predictor has the following form,

$$\hat{x}_{t+j} = E(x_{t+j}|x_t) = \alpha + \mu_\beta \hat{x}_{t+j-1} = \alpha + \mu_\beta(\alpha + \alpha\mu_\beta + \ldots + \alpha\mu_\beta^{j-1} + \mu_\beta^j x_t)$$

$$= \alpha\left(\frac{1-\mu_\beta^j}{1-\mu_\beta}\right) + \mu_\beta^j x_t, \text{ for } \mu_\beta < 1.$$

The j-step ahead conditional variance of predication error has the form

$$v_{t+j|t} = Var(x_{t+j}|x_t)$$

$$
\begin{aligned}
&= E[Var(x_{t+j}|x_{t+j-1}, x_t)] + Var[E(x_{t+j}|x_{t+j-1}, x_t)] \\
&= \eta v_{t+j-1|t} + \sigma_\beta^2 \hat{x}_{t+j-1}^2 + \sigma^2.
\end{aligned}
$$

Notice that the $j$-step-ahead conditional variance $v_{t+j|t}$ depends on both its lag value $v_{t+j-1|t}$ and the $(j-1)$-step-ahead prediction value $\hat{x}_{t+j-1}$. An evaluation of $v_{t+j|t}$ gives that

$$
v_{t+j|t} = \eta^{j-1} v_{t+1|t} + \sigma^2 \sum_{k=1}^{j-1} \eta^{k-1} + \sigma_\beta^2 \sum_{k=1}^{j-1} \eta^{k-1} \hat{x}_{t+j-k}^2 \tag{7}
$$

If the true RCAR(1) process has a unit root, that is $\eta = \mu_\beta^2 + \sigma_\beta^2 = 1$, then

$$
v_{t+j|t} = v_{t+1|t} + \sigma^2 (j-1) + \sigma_\beta^2 \sum_{k=1}^{j-1} \hat{x}_{t+j-k}^2.
$$

It is obvious that $v_{t+j|t} \to \infty$ as $j \to \infty$.

On the other hand, if the true RCAR(1) process is weakly stationary, that is $\eta = \mu_\beta^2 + \sigma_\beta^2 < 1$, then the first term in Equation (7) converges to zero as $j$ goes to infinity and the second term converges to $\frac{\sigma^2}{1-\eta}$. The third term can be rewritten in the following form:

$$
\begin{aligned}
\sigma_\beta^2 \sum_{k=1}^{j-1} \eta^{k-1} \hat{x}_{t+j-k}^2 &= \sigma_\beta^2 \sum_{k=1}^{j-1} \left[ \frac{\alpha}{1-\mu_\beta} + \left( x_t - \frac{\alpha}{1-\mu_\beta} \right) \mu_\beta^{j-k} \right]^2 \eta^{k-1} \\
&= \sigma_\beta^2 \sum_{k=1}^{j-1} (a + b\mu_\beta^{j-k})^2 \eta^{k-1} \\
&\leq \sigma_\beta^2 a^2 \sum_{k=1}^{j-1} \eta^{k-1} + 2\sigma_\beta^2 |ab| \sum_{k=1}^{j-1} |\mu_\beta|^{j-k} + \sigma_\beta^2 b^2 \sum_{k=1}^{j-1} \mu_\beta^{2(j-k)},
\end{aligned}
$$

where $a = \frac{\alpha}{1-\mu_\beta}$ and $b = x_t - \frac{\alpha}{1-\mu_\beta}$. Each term in the right hand side of the above inequality converges to a finite number as $j$ goes to infinity, since $\eta < 1$ implies $|\mu_\beta| < 1$. Furthermore, since the third term in Equation (7) is a non-negative increasing series indexed by $j$, it also converges as $j \to \infty$. Therefore, we have the conditional prediction error term $v_{t+j|t}$ to be convergent as $j \to \infty$ when $\eta < 1$.

In practice, we need to estimate the parameter $\theta = (\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)^T$ instead of knowing its true value. The difference between the prediction error obtained this way and the one obtained based on knowing the true values is negligible. An implication of the above derivation is that the predication interval for future observation for a unit-root RCAR(1) process tends to be wider than that of a

stationary process, hence it is of importance to test for the unit root in an RCAR(1) process. In the following sections, we will introduce two different Bayesian testing procedures to test the unit-root hypothesis of RCAR(1) model, that is, we want to test the null hypothesis $H_0 : \eta = 1$ versus the alternative hypothesis $H_1 : \eta < 1$.

## 4.2 Choice of Prior Densities

For the convenience of assigning prior density for $\eta$, we reparameterize $\theta$ to $\psi = (\alpha, \sigma^2, \mu_\beta, \eta)^T$. We will specify priors in terms of $\psi$ rather than $\theta = (\alpha, \mu_\beta, \sigma_\beta^2, \sigma^2)^T$ as was used in the estimation problem. The prior we use for $\alpha$ is $N(\mu_0, V_0)$, where $\mu_0$ and $V_0$ are assumed to be known quantities. We consider both informative priors and non-informative priors for $\alpha$. For non-informative prior, we set $\mu_0$, the mean of the normal distribution to be 0 and $V_0$, the variance of the normal distribution to be large, say $10^6$. For informative prior, we used $\mu_0 = 0$ and $V_0 = 10$ in our simulation study. We employed the Inverse Gamma prior distribution $IG(a, b)$ for $\sigma^2$. Same as for $\alpha$, we considered both informative and non-informative priors. In our simulation, we chose $IG(a = 2 + 10^{-10}, b = 0.1)$ as non-informative prior and $IG(a = 2.1, b = 1.1)$ as informative prior.

The central concern in testing unit root of RCAR(1) is $\eta$. We will focus on the choice of its prior in the remaining part of this section.

(i) The mixture prior.

Let $B$ be a Bernoulli random variable with probability of success $Pr(B = 1) = p$. In addition, let $U$ be a continuous random variable defined on the interval $(0, 1)$ with density function $f_U(.)$. Furthermore, assume $B$ and $U$ are independent. We set the stationarity parameter $\eta = B + (1 - B)U$. Here, $p$ can be viewed as the mixing probability. And $p$ itself can be regarded as either a constant or an unknown random quantity. When $p$ is assumed to be a constant, the marginal prior density of $\eta$ is

$$f_\eta(\eta) = pI(\eta = 1) + (1 - p)f_U(\eta)I(0 < \eta < 1).$$

However, when $p$ is assumed to be a random quantity, with hyper-prior density $f_p(p)$, then

$$f(\eta|p) = pI(\eta = 1) + (1 - p)f_U(\eta)I(0 < \eta < 1).$$

and

$$f_\eta(\eta) = E(p)I(\eta = 1) + [(1 - E(p)]f_U(\eta)I(0 < \eta < 1).$$

that is, the marginal density of $\eta$ is also a mixture distribution with mixing probability $E(p)$. In our simulation study, we choose $U \sim \text{Unif}(0,1)$, and $p = 0.5$ and $0.95$ as the constant mixing probabilities.

Now define $\gamma = \mu_\beta^2$. Since it is true that $\gamma \leq \mu_\beta^2 + \sigma_\beta^2 = \eta$, given $\eta$, we assume the prior distribution of $\gamma$ is $\text{Unif}(0, \eta)$. As a consequence, the marginal prior density of $\gamma$ is $f_\gamma(\gamma) = -ln(\gamma),$ for $0 < \gamma < 1$. The prior density of $\mu_\beta$ can be obtained using standard probability density transformation technique.

(ii) The flat prior.
In addition to the above defined mixture prior for $\eta$, we suggest flat prior which has a support on a continuous region that includes unity, for example, $\text{Unif}(0, 1+\epsilon)$, for Bayesian unit-root testing based on Posterior Interval method. In this case, $\epsilon > 0$ can be chosen arbitrarily. We choose $\epsilon$ such that the expected value of $\eta$ having this density is equal to the expected value of $\eta$ having a mixture prior density with mixing probability p. In our simulation study, we chose $\epsilon = 1$, i.e. a $\text{Unif}(0,2)$ distribution as the flat prior for $\eta$, which is corresponding to the case of mixture prior with mixing probability $p = 0.5$ and mixture density $\text{Unif}(0,1)$. For the unit-root testing based on Bayes Factor method, we used $\text{Unif}(0,1)$ as the flat prior for $\eta$.

## 4.3 Unit-root Testing Based on Posterior Interval

In frequentist analysis, one way to perform hypothesis testing is to use confidence interval as the acceptance region for the corresponding testing problem. Similar approach is also available in Bayesian inference. One usual way is to present a central interval of 0.95 posterior probability with lower and upper bound corresponding to 2.5% and 97.5% percentiles of the posterior distribution respectively. It should be noted that when using a continuous density which is defined on an interval that doesn't include 1 as a prior for $\eta$, 1 can never be picked by the Gibbs sampling procedures. This problem can be avoided by using the mixture prior or the continuous prior with support including unity as we proposed in Section 4.2. One can obtain a 95% equal-tailed posterior interval for $\eta$ based on the posterior distribution of $\eta$ from the Markov chain we construct via the Gibbs sampling algorithm. In order to test the unit-root hypothesis $H_0 : \eta = 1$, we may simply use the rule of rejecting the null hypothesis if 1 is not included in the 95% posterior interval. From our simulation study, we found this simple rule performs reasonably

well for properly chosen priors of $\eta$ in the sense of maintaining good frequentist properties, such as the high power with low total error rates (Type I + Type II). An even more attracting feature of this rule is that it is relatively easier to implement by constructing posterior intervals via MCMC methods compared to the use of Bayes Factor method.

## 4.4 Unit-root Testing Based on Bayes Factor

The problem of testing null hypothesis $H_0 : \eta = 1$ versus alternative hypothesis $H_1 : \eta < 1$ can be regarded as comparing two proposed models and selecting a better one based on the observed data $\underline{x}$. In this sense, it might be convenient to switch notation from 'hypotheses' $H_0$ and $H_1$ to 'models' $M_0$ and $M_1$. The Bayes Factor is defined as the ratio of the posterior odds $\frac{P(M_0|\underline{x})}{P(M_1|\underline{x})}$ to the prior odds $\frac{P(M_0)}{P(M_1)}$. We can interpret the Bayes Factor as the odds in favor of $M_0$ against $M_1$ given the data $\underline{x}$. In most of the data analysis, the prior odds is taken to be 1 to indicate the prior ignorance. Following the Bayes' Theorem, Bayes Factor can be written as,

$$BF = \frac{m_0(\underline{x})}{m_1(\underline{x})}$$

where $m_i(\underline{x})$ denotes the marginal likelihood of the data under model $i$ $(i = 0, 1)$, that is,

$$m_i(\underline{x}) = P(\underline{x}|M_i) = \int f(\underline{x}|\psi_i, M_i)\pi_i(\psi_i)d\psi_i$$

where $\psi_i$ is the parameter vector under model $i$, $f(\underline{x}|\psi_i, M_i)$ is the likelihood of the data under model $i$ and $\pi_i(\psi_i)$ is the prior density for $\psi_i$ under model $i$.

Computing marginal likelihood, hence Bayes Factor, turns out to be extremely challenging for most practical problems due to the difficulty of high-dimension integration. In literature, there are numerous approaches for computing the Bayes Factor. For a nice review on this topic, readers are referred to Kass and Raftery (1995) and Han and Carlin (2001).

When using a continuous prior for $\eta$, for example Unif$(0, 1)$, we applied the method proposed by Gelfand and Dey (1994) to calculate the marginal likelihood $m_i(\underline{x})$, $i = 1, 2$, and hence the Bayes Factor. Their idea follows from the fact that the identity

$$[m_i(\underline{x})]^{-1} = \int \frac{h(\psi_i)}{f(\underline{x}|\psi_i)p(\psi_i)} p(\psi_i|\underline{x})d\psi_i$$

holds for any proper density $h(.)$. Given sample $\psi_i^{(k)}, k = 1, 2, \ldots, M$, from the posterior distribution, this suggests the following estimator of the marginal density $m_i(\underline{x})$,

$$\hat{m}_i(\underline{x}) = \left[ \frac{1}{M} \sum_{k=1}^{M} \frac{h(\psi_i^{(k)})}{f(\underline{x}|\psi_i^{(k)}) p(\psi_i^{(k)})} \right]^{-1}.$$

Gelfand and Dey proposed choosing $h(.)$ to roughly match the posterior density $p(\psi_i|\underline{x})$, in particular, they suggested using a multivariate normal or $t$ distribution with mean and variance estimated from the $\psi_i^{(k)}$ samples. We used multivariate normal density as $h(.)$ for our simulation study.

If the mixture prior density is used for $\eta$, we can estimate the Bayes Factor from the Markov chain much more easily. Recall that if we assume the prior odds to be 1, i.e. the mixing probability $p$ of the mixture prior to be 0.5, then the Bayes Factor is equal to the posterior odds given by

$$\frac{P(M_0|\underline{x})}{P(M_1|\underline{x})} = \frac{P(\eta = 1|\underline{x})}{P(0 < \eta < 1|\underline{x})}$$

It is easy to see that when there is a positive mass assigned to point 1 *apriori*, the posterior density of $\eta$ will have a positive mass at 1 as well. Therefore, the numerator and the denominator of the posterior odds can be easily estimated from the Markov chain generated by Gibbs sampler algorithm. This estimator is simply,

$$\frac{\sum_{m+1}^{m+M} I(\eta^{(i)} = 1)}{\sum_{m+1}^{m+M} I(0 < \eta^{(i)} < 1)}$$

where $I(.)$ is the indicator function, $m$ is the burn-in point, $M$ is the number of points after burn-in and $\eta^{(i)}$ is the $i^{th}$ element of the Markov chain for $\eta$ that is used for the inferential purpose. The Bayesian unit-root test based on the Bayes Factor (note that prior odds is 1) rejects the null hypothesis if this estimator is less than 1.

## 4.5 A Simulation Study

In this section, we display the results of a simulation experiment we conducted in order to explore the frequentist properties of the Posterior Interval (PI) method as well as the Bayes Factor (BF) method to test the unit-root hypothesis of RCAR(1) model. To simulate data from Model (2), we fixed $\alpha = 0$, $\sigma^2 = 1$ and $\mu_\beta = 0.6$ and 1, and tried several values for $\eta$ (e.g. $\eta = 1$, 0.98, 0.95 etc.). Tables 3 lists the different combinations of $\mu_\beta$ and $\eta$ that were used in our simulation.

We applied non-informative prior densities for $\alpha$ and $\sigma^2$ for our experiment related to Posterior Interval method whereas we used informative priors of these parameters for the experiment related to BF method. For $\eta$, we considered both mixture priors (with constant mixing probabilities ($p = 0.5$ and 0.95) and mixture distribution Unif$(0, 1)$) and flat priors, (Unif$(0, 2)$ for PI method and Unif$(0, 1)$ for BF method).

The second column of Table 3 provides the results of a pilot study we conducted in order to assess the performance of different priors used in BF method by means of comparing the Proportion Correct Decision (PCD). We chose the sample size $n = 300$ and the Monte Carlo sample size to be 100. The PCD is defined as follows: for a series with a unit root, that is, $\eta = 1$, the PCD is the proportion that the estimated BF is greater or equal to 1, while for a stationary series, that is, $\eta < 1$, the PCD is the proportion that the estimated BF is less than 1. In the table, the highest PCDs are highlighted. It is observed that when the true value of $\eta$ is 1, the MU(0.95) prior performs better than the other two priors, which is conceivable because it assigns more prior mass at point 1 hence favors the null hypothesis. However, it should be noted that the other two priors also perform comparably well as their PCDs are reasonably close to 1. When the true value of $\eta$ moves away from 1, the PCDs for using MU(0.5) prior performs consistently better than that of the other two. In all, we can see when using the MU(0.5) prior for $\eta$, the test maintains a relatively higher power and at the same time keeps the Type-I error rate low. As a consequence, we decided to study further the performance of using MU(0.5) as the prior in BF method. In the second column of Table 4 we list the PCDs obtained from using MU(0.5) for different sample sizes. We chose the sample sizes $n = 1000, 2500$ and 5000, which are the typical sizes one would normally observe from the empirical financial time series data. The simulation is based on 500 Monte Carlo replications. It is observed that for moderately large sample size, the test achieves very high power when the true value of $\eta$ is less than 0.8. However the Type II error rates committed by the test are quite high when $\eta$ is close to 1.

A similar pilot study was run to evaluate the performance of using different priors of $\eta$ in PI method, and its result is given in the third column of Table 3. In this case, PCD is defined as follows: for a series with a unit root, the PCD is the proportion of cases in which 1 falls into the 95% equal-tail posterior interval of $\eta$, whereas for a stationary series, the PCD is the proportion of cases that 1

Table 3: Average values of Proportion of Correct Decision (PCD) for unit-root testing using BF and PI methods (sample size n=300 for BF and n=1500 for PI, 100 MC replications).

| case | | BF method | | PI method | |
|---|---|---|---|---|---|
| $\eta$ | $\mu_\beta$ | Prior for $\eta$ | PCD | Prior for $\eta$ | PCD |
| 1 | 0.6 | MU(0,1) p=0.95 | **0.99** | MU(0,1) p=0.95 | **1** |
| | | MU(0,1) p=0.5 | 0.93 | MU(0,1) p=0.5 | **1** |
| | | Unif(0,1) | **0.99** | Unif(0,2) | 0.95 |
| 1 | 1 | MU(0,1) p=0.95 | **0.96** | MU(0,1) p=0.95 | **0.95** |
| | | MU(0,1) p=0.5 | 0.85 | MU(0,1) p=0.5 | 0.92 |
| | | Unif(0,1) | 0.95 | Unif(0,2) | 0.69 |
| 0.98 | 0.6 | MU(0,1) p=0.95 | 0.04 | MU(0,1) p=0.95 | 0 |
| | | MU(0,1) p=0.5 | **0.18** | MU(0,1) p=0.5 | 0.02 |
| | | Unif(0,1) | 0.1 | Unif(0,2) | **0.08** |
| 0.95 | 0.6 | MU(0,1) p=0.95 | 0.03 | MU(0,1) p=0.95 | 0 |
| | | MU(0,1) p=0.5 | **0.18** | MU(0,1) p=0.5 | 0.01 |
| | | Unif(0,1) | 0.07 | Unif(0,2) | **0.15** |
| 0.9 | 0.6 | MU(0,1) p=0.95 | 0.05 | MU(0,1) p=0.95 | 0 |
| | | MU(0,1) p=0.5 | **0.31** | MU(0,1) p=0.5 | 0.05 |
| | | Unif(0,1) | 0.12 | Unif(0,2) | **0.37** |
| 0.8 | 0.6 | MU(0,1) p=0.95 | 0.08 | MU(0,1) p=0.95 | 0.20 |
| | | MU(0,1) p=0.5 | **0.59** | MU(0,1) p=0.5 | 0.53 |
| | | Unif(0,1) | 0.36 | Unif(0,2) | **0.96** |
| 0.7 | 0.6 | MU(0,1) p=0.95 | 0.31 | MU(0,1) p=0.95 | 0.94 |
| | | MU(0,1) p=0.5 | **0.86** | MU(0,1) p=0.5 | **1** |
| | | Unif(0,1) | 0.72 | Unif(0,2) | **1** |
| 0.6 | 0.6 | MU(0,1) p=0.95 | 0.77 | MU(0,1) p=0.95 | **1** |
| | | MU(0,1) p=0.5 | **0.99** | MU(0,1) p=0.5 | **1** |
| | | Unif(0,1) | 0.96 | Unif(0,2) | **1** |
| 0.5 | 0.6 | MU(0,1) p=0.95 | 0.98 | MU(0,1) p=0.95 | **1** |
| | | MU(0,1) p=0.5 | **1** | MU(0,1) p=0.5 | **1** |
| | | Unif(0,1) | **1** | Unif(0,2) | **1** |

falls outside the 95% equal-tail posterior interval of $\eta$. The sample size used in the simulation is 1500, and the Monte Carlo sample size is 100. From the table we see that when the series has a unit root, the priors MU(0.95) and MU(0.5) perform better, while when series is stationary and the true value of $\eta$ is greater than 0.8, Unif$(0, 2)$ is more powerful than the other two. When the true value of $\eta$ is less than 0.8, all three tests perform very well in terms of maintaining high power. It should also be noted that MU(0.5) performs consistently better than MU(0.95) except for the case when $\eta = 1$ and $\mu_\beta = 1$. However, even for this case, the difference in PCD between these two is only about 0.03. As a result, we studied further the performance of unit-root testing based on PI using Unif$(0, 2)$ and MU(0.5) as priors for $\eta$. We used the same sample sizes as those used in Bayes Factor method, i.e. n=1000, 2500 and 5000, and the Monte Carlo sample size of 500. The third and fourth columns of Table 4 list the simulated PCDs for using these two priors. It is seen that when using the MU(0.5) prior there is less Type-I error committed, while when using Unif$(0, 2)$ as prior there is more power achieved. For practitioner, one way to choose between these two priors is to compare the so-called '*total error rate*', which is calculated as the sum of Type I and Type II error rates. For more discussion about the total error rate for the unit-root testing and for comparison with frequentist test criteria, see Wang (2003).

# 5. APPLICATIONS

In this section, we will consider the application of RCAR(1) models using the Bayesian procedures that we developed in the previous sections. The data sets we will use are the daily transaction volume for NASDAQ stock index and for IBM stock. These data were obtained from Yahoo Finance, `http://finance.yahoo.com/`.

## 5.1 NASDAQ Stock Index Data

This data consists of 3279 records of the daily transaction volume of NASDAQ index ranging from January 1990 to December 2002. Top panel of Figure 2 is the time series plot of the daily data in log scale. One feature observed from the plot is that there is a distinct upward trend in the data. Let $x_t$ denote the logarithm of the daily transaction volume at time $t$. We considered the following model which

will capture the linear trend observed in the original series,

$$x_t = \beta_0 + \beta_1 t + y_t. \tag{8}$$

That is, we regressed $x_t$ on time $t$ and obtained the residual $y_t$. The bottom panel in Figure 2 is the plot of the detrended data $y_t$. It is seen that the series is fairly stable around its mean value, except for short-term bursts of high volatility. It is therefore reasonable to model the detrended series $y_t$ as an RCAR(1).

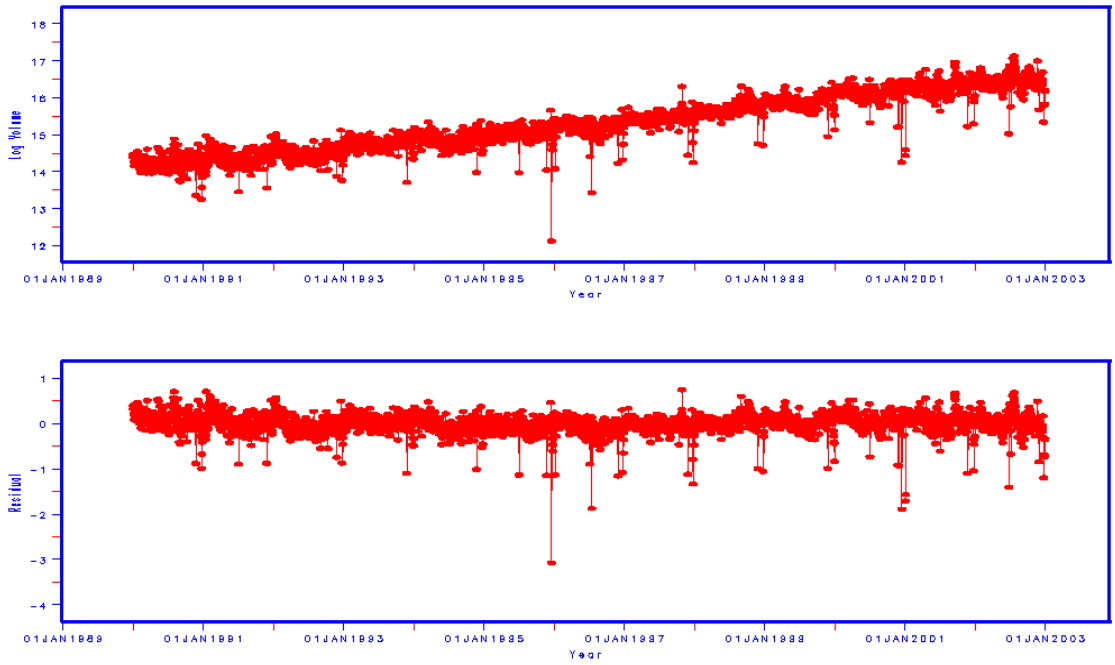


Figure 2: Plot of the NASDAQ daily transaction volume (on log scale) and its detrended version (n=3279)

### 5.1.1 Parameter Estimation

We fitted an RCAR(1) model to the detrended series and considered the parameter estimation using non-informative priors for the parameters in the model. In particular, we used $\mathrm{N}(0, 10^6)$ as prior for $\alpha$ and $\mu_\beta$ and $\mathrm{IG}(2 + 10^{-10}, 0.1)$ as prior for $\sigma_\beta^2$ and $\sigma^2$. We constructed two Markov chains with two sets of initial values and executed the Gibbs sampling iterations 7500 times. For each chain, the first 5000 iterations were discarded and the last 2500 iterations were used to obtain

the posterior distributions of the parameters. Table 5 lists the posterior mean, standard deviation as well as the 95% posterior interval of the parameters in the model.

### 5.1.2 Model Selection

Other than fitting an RCAR(1) model to the detrended log volume transaction series for NASDAQ data, we also considered using the GARCH type of volatility models. One model we used to fit the data is GARCH(1,2) given as the following:

$$\begin{aligned} y_t &= \mu + r_t \\ r_t &= \sigma_t u_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2, \end{aligned} \tag{9}$$

where $u_t$ is a sequence of iid N(0,1) random variables. As an extension of the GARCH(1,2), we also fitted an AR(1) model for data $y_t$ with a GARCH(1,2) error structure, that is,

$$\begin{aligned} y_t &= \mu + \rho y_{t-1} + r_t \\ r_t &= \sigma_t u_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 r_{t-1}^2 + \beta_1 \sigma_{t-1}^2 + \beta_2 \sigma_{t-2}^2, \end{aligned} \tag{10}$$

We used `SAS Proc Autoreg` to fit the above GARCH type models and obtained the AIC associated with each model. We also calculated the AIC from fitting an RCAR(1) to the data using Bayesian approach. Table 6 contains the summary results which we used to select among the three proposed models. It is observed that the detrended data $y_t$ prefer RCAR(1) fit to either GARCH(1,2) fit or AR(1)-GARCH(1) fit based on AIC.

### 5.1.3 Forecasting

From Table 6, it is seen that both RCAR(1) and AR(1)-GARCH(1,2) are competitive models for fitting the residual series. Therefore, we considered the forecasting problem based on fitting these two models. Figure 3 illustrates up to 15 days ahead forecasts and the corresponding forecasting intervals. In the legend, 'lclm', 'pred' and 'uclm' refer to the 95% lower forecasting limit, the point forecast and the 95% upper forecasting limit obtained from fitting an RCAR(1) to the data.

Table 4: Proportion of Correct Decision (PCD) from selected unit-root test criteria using different sample sizes (500 MC replications).

| case | | BF using MU(0.5) | | | PI using Unif(0,2) | | | PI using MU(0.5) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | sample size | | | sample size | | | sample size | | |
| $\eta$ | $\mu_\beta$ | 1000 | 2500 | 5000 | 1000 | 2500 | 5000 | 1000 | 2500 | 5000 |
| 1 | 0.6 | 0.97 | 0.97 | 0.99 | 0.95 | 0.93 | 0.94 | 1 | 1 | 1 |
| 1 | 1 | 0.83 | 0.68 | 0.67 | 0.66 | 0.68 | 0.59 | 0.96 | 0.89 | 0.91 |
| 0.98 | 0.6 | 0.12 | 0.10 | 0.10 | 0.11 | 0.13 | 0.14 | 0.02 | 0.02 | 0.01 |
| 0.95 | 0.6 | 0.15 | 0.20 | 0.27 | 0.12 | 0.22 | 0.38 | 0.01 | 0.02 | 0.04 |
| 0.9 | 0.6 | 0.33 | 0.58 | 0.83 | 0.28 | 0.62 | 0.89 | 0.03 | 0.12 | 0.35 |
| 0.8 | 0.6 | 0.86 | 0.99 | 1 | 0.82 | 0.99 | 1 | 0.30 | 0.87 | 1 |
| 0.7 | 0.6 | 1 | 1 | 1 | 1 | 1 | 1 | 0.88 | 1 | 1 |
| 0.6 | 0.6 | 1 | 1 | 1 | 1 | 1 | 1 | 0.98 | 1 | 1 |
| 0.5 | 0.6 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |

Table 5: Posterior summaries of an RCAR(1) fit to detrended NASDAQ data

| Parameter | Posterior Mean | Standard Deviation | 95% Posterior Interval |
|---|---|---|---|
| $\alpha$ | -0.004 | 0.003 | (-0.011, 0.002) |
| $\mu_\beta$ | 0.557 | 0.020 | (0.518, 0.597) |
| $\sigma^2_\beta$ | 0.106 | 0.017 | (0.077, 0.141) |
| $\sigma^2$ | 0.032 | 0.001 | (0.030, 0.033) |
| $\eta$ | 0.417 | 0.032 | (0.353, 0.487) |

Table 6: Model selection based on AIC and fitting different volatility models to the detrended NASDAQ data

| Model | AIC |
|---|---|
| RCAR(1) based on Bayesian posterior mean | -1687.93 |
| GARCH(1,2) | -1089.65 |
| AR(1)-GARCH(1,2) | -1643.79 |

Whereas 'lclm2', 'pred2' and 'uclm2' refer to the 95% lower forecasting limit, the point forecast and the 95% upper forecasting limit obtained from fitting an AR(1)-GARCH(1,2) to the data. It is seen that the forecasting interval from RCAR(1) model is narrower than that from AR(1)-GARCH(1,2) model when forecasting of the near future, i.e. up to about 5 days ahead; after the 6th day, the forecasting interval obtained by using the two models are almost identical and the length of the interval approaches to a fixed number. Recall that for RCAR(1) model, we have shown in Section 4.1 that when the true value of the stationary parameter is less than 1, the long term forecasting interval converges. For the detrended NASDAQ series, the estimated stationary parameter is about 0.4, therefore the long term behavior of the forecasts conforms with the theoretical result. Another feature observed from Figure 3 is that the point forecasts from fitting two models are very close. The solid points are the actual residual values observed from January 2nd, 2003 to January 22nd, 2003. It is seen that point forecasts reasonably predict what happened in real world. It is also seen that the actual observed values all fall into the 95% forecasting intervals from using both two models.

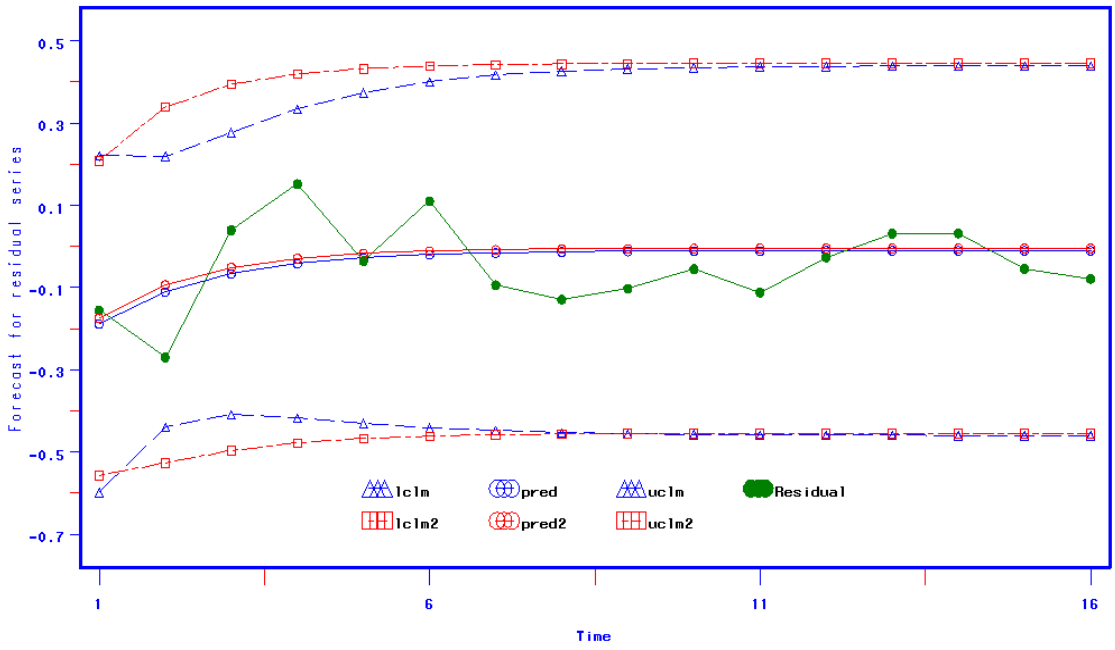


Figure 3: Forecasts for detrended NASDAQ data based on RCAR(1) and AR(1)-GARCH(1,2) models

Instead of using the residual series $y_t$, it is more intuitive to perform forecasting based on the original series $x_t$. To do so, we added the linear time trend back into the point forecasts and adjust the forecasting interval accordingly. Figure 4 gives the plot of the forecasts for the original $x_t$ series.

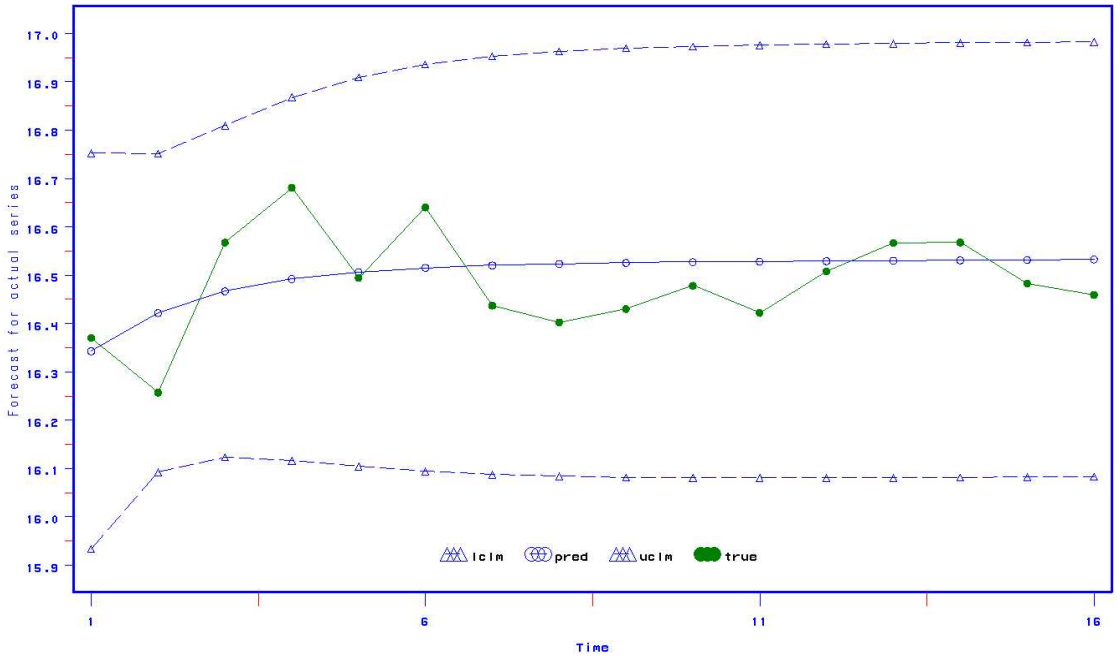


Figure 4: Forecasts for NASDAQ daily transaction volume data based on RCAR(1) model

## 5.2 IBM Data

The second example considered is the IBM stock daily transaction volume data. The data consists of ten years records between 1993 and 2002. Let $x_t$ denote the daily transaction volume at time $t$. Figure 5 displays the time series plot for $x_t$. The sample size for this data set is n=2518.

### 5.2.1 Parameter Estimation

We performed a Bayesian estimation by fitting an RCAR(1) to the IBM daily volume data. The same non-informative priors for the parameters were used as

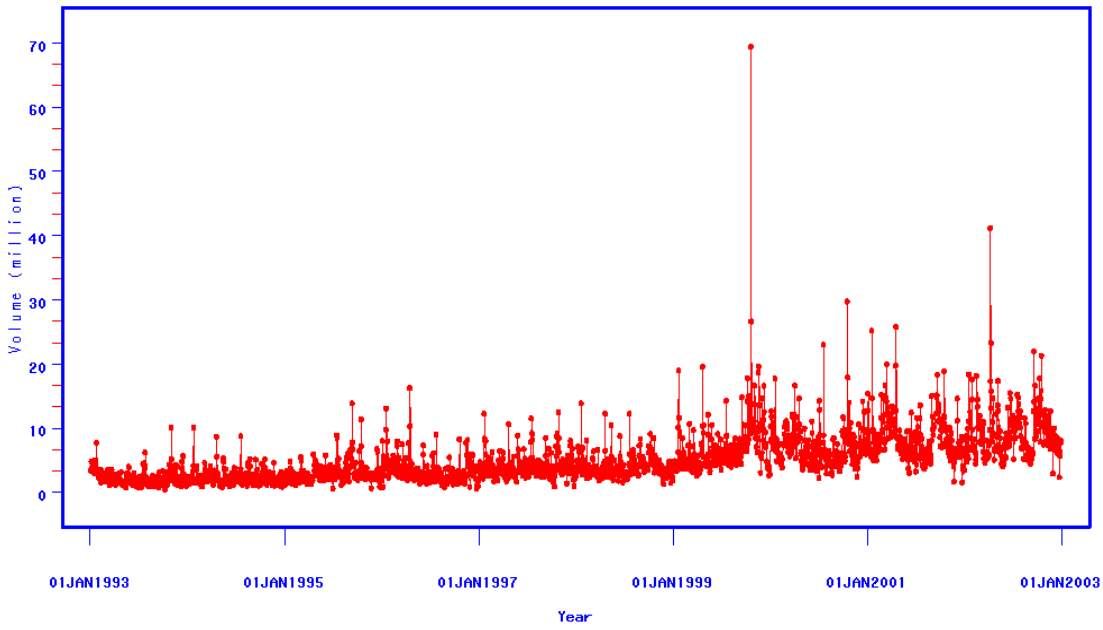


Figure 5: Time series plot of the IBM daily transaction volume data (volume in million), n=2518

in the NASDAQ stock index example. In Table 7, the posterior mean, standard deviation and the 95% posterior interval of the parameters are given. It is seen from the table that the estimate for the stationary parameter $\eta$ in RCAR(1) is about 0.924. It is therefore reasonable to test for the unit-root hypothesis of the series.

### 5.2.2 Unit-root Testing

We performed the unit-root testing based on Bayes Factor and Posterior Interval using various priors for the stationarity parameter of RCAR(1) model. We applied the MU(0.5) prior for $\eta$ in BF method, whereas we used both MU(0.5) and U(0,2) prior for $\eta$ in PI method. The same priors for $\alpha$ and $\sigma^2$ were chosen as in the simulation study. Based on the results presented in Table 8, it is seen that the unit-root hypothesis should be rejected by all the testing criteria except the one based on PI using MU(0.5) as prior for $\eta$. For the similar sample size, i.e. n=2500, our simulation results in Section 4.4 show that for $\eta$ is around 1, using PI method

with MU(0.5) as prior for $\eta$ has the largest total error rate compared to the other two testing criteria. Therefore, we should be more convinced that the unit-root hypothesis should be rejected for the IBM daily transaction volume data.

### 5.2.3 Forecasting

Figure 6 gives up to 21 days ahead forecasts for the IBM daily transaction volume data based on fitting an RCAR(1) model to the original series. There are a few features observed from the graph which seem to be questionable about the model fitting. One is the systematical bias between the point forecasts (denoted by circles) and the actual series (denoted by dots); the other one is that the lower forecasting limits (denoted by triangles) fall below zero; Also, the long term forecasting interval seems to converge very slowly. These features motivated us to seek alternative ways to model the data. We took a logarithm transformation to the original daily volume series. Given that we observed an linear upward trend in the transformed series, we first detrended the data similar to what we did for the NASDAQ data.

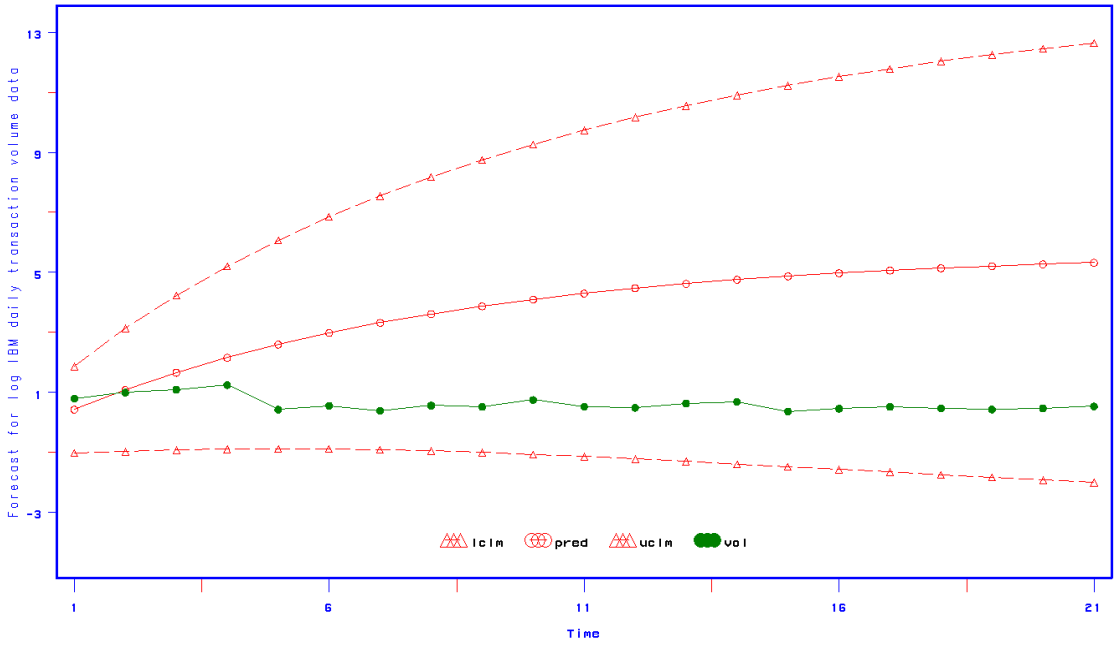


Figure 6: Forecasts for IBM daily transaction volume data based on RCAR(1)

Based on fitting an RCAR(1) model to the detrended IBM daily volume data,

Table 7: Posterior summaries of an RCAR(1) fit to IBM daily transaction volume data

| Parameter | Posterior Mean | Standard Deviation | 95% Posterior Interval |
|---|---|---|---|
| $\alpha$ | 0.726 | 0.066 | (0.594, 0.853) |
| $\mu_\beta$ | 0.871 | 0.019 | (0.835, 0.910) |
| $\sigma^2_\beta$ | 0.165 | 0.007 | (0.151, 0.179) |
| $\sigma^2$ | 0.565 | 0.071 | (0.434, 0.710) |
| $\eta$ | 0.924 | 0.034 | (0.857, 0.993) |

Table 8: Bayesian unit-root tests for IBM daily transaction volume data

| Testing Procedure | Test Statistic | Conclusion |
|---|---|---|
| Bayesian using Bayes Factor | $BF^0 = 0.801$ | Rej |
| Bayesian using Posterior Interval | 95% $PI^1 = (0.879, 1)$ | Don't Rej |
| | 95% $PI^2 = (0.862, 0.987)$ | Rej |

(0 & 1: using MU(0.5) as prior for $\eta$; 2: using U(0,2) as prior for $\eta$)

we calculated the point forecasts as well as the forecasting interval. Figure 7 gives the plot of the forecasts. Compared with Figure 6, it is seen there is no systematic pattern of bias between the forecast series and the observed series. Also, the point forecasts is fairly close to the actual observed series. Another feature seen from the graph is that the forecasting interval get fairly stable after about 5 or 6 days. These also suggest that our second approach to model the IBM data is better than directly modeling the original series as an RCAR(1).

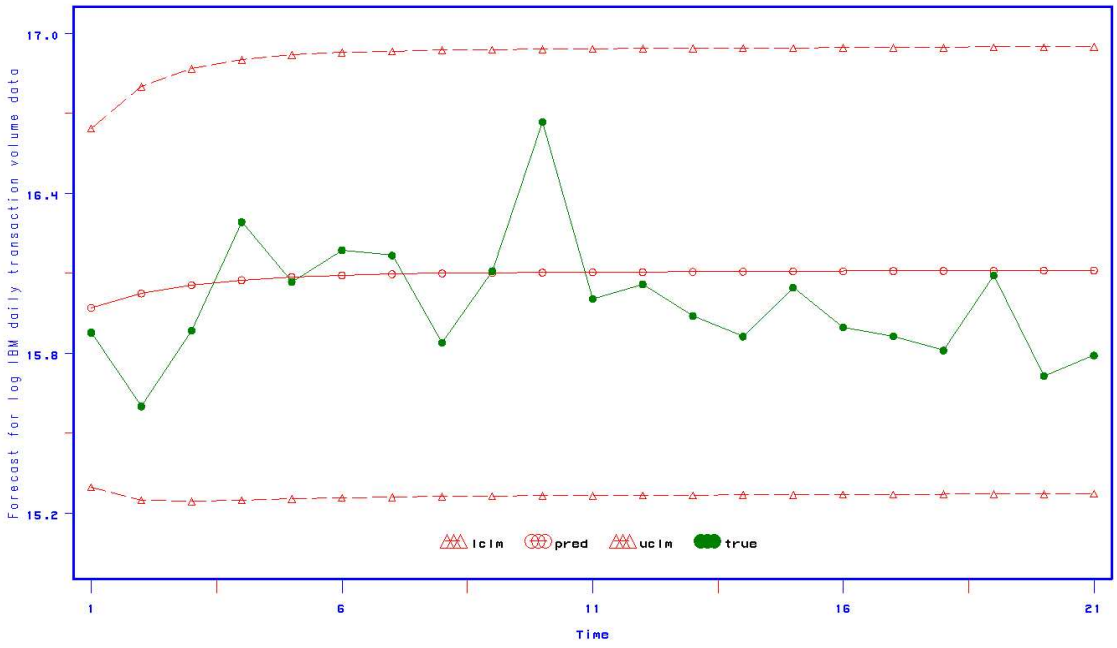


Figure 7: Forecasts for IBM daily data based on fitting an RCAR(1) model to the detrended data (volume in log scale)

# 6. CONCLUSIONS

In this article, we have developed Bayesian estimation and unit-root testing procedures for RCAR(1) Models. Using extensive simulation experiments we showed that the Bayesian estimation procedure using non-informative priors works well for small to moderate sample sizes. The coverage probabilities for both stationary and non-stationary cases (excluding random walk) are close to the nominal level. We

also investigated the performance of the model selection criteria using simulation studies. It appears that both AIC and GGC performs well in picking up the correct model. In general we find that RCAR models are robust to model misspecification and the loss in efficiency due to more parameters is not severe as compared to AR models. We would recommend users to fit the RCAR models to time series data where volatility is expected. For unit-root testing, we introduced a mixture prior density for the stationarity parameter $\eta$, which assigns a non-zero mass to $\eta$ at the null value. By using the mixture prior, we can apply the posterior interval of $\eta$ to test for the unit root in an RCAR(1) model. For Bayesian testing based on the posterior interval, we also considered continuous densities for $\eta$ which include the null value of the parameter in the support. If we use a continuous density whose support doesn't include the null value of $\eta$, the posterior interval will never include the value being tested. Therefore, we can't apply the Posterior Interval method for this kind of prior. However, unit-root testing based on Bayes Factor can be applied in this case. We used both continuous prior and mixture prior for $\eta$ in the study of unit-root test based on Bayes Factor. Estimating Bayes Factor usually requires computationally extensive algorithms and informative priors, therefore it may not be practical. However, if we use the mixture prior for $\eta$, the computation of Bayes Factor can be readily done by means of counting 1's in the posterior sample of $\eta$. On the other hand, the computation in Posterior Interval method is very simple and can be easily carried out. Using a Monte Carlo study, we found that for samples of moderate to large sizes, our proposed tests maintain relatively high power, which indicates that one can obtain reliable conclusions from applying them to test for a unit root in RCAR(1). Some possible extension to our current research might include:

(i) analysis of random coefficient autoregression with auxiliary variables,

(ii) unit-root testing for RCAR(1) model with correlated random coefficient and error, and

(iii) unit-root testing for higher order RCAR models.

# REFERENCES

Berger, J. O. (1985), Statistical decision theory and Bayesian analysis. New York: Springer-Verlag

Best, N. G., Cowles, M. K., and Vines, S. K. (1995), CODA Menu version 0.30.

Bollerslev T. (1986), "Generalized autoregressive conditional heteroscedasticity," *Journal of Econometrics*, 31 , 307-327.

Datta, G. S. and Mukerjee, R. (2004), "Probability Matching Priors: Higher Order Asymptotics," Lecture Notes in Statistics, Vol. 178, Springer-Verlag, New York.

Engle, R F. (1982), "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflations," *Econometrica*, 50, 987-1007.

Gelfand, A. and Dey, D. K. (1994), "Bayesian model choice: Asymptotics and exact calculations," *Journal of the Royal Statistical Society, Series B, Methodological*, 56, 501-514.

Gelfand, A. and Ghosh, S. K. (1998), "Model choice: A minimum posterior predictive loss approach," *Biometrika*, 85, 1-11.

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

Gelman, S. and Rubin, D. (1992), "Inference from iterative simulation using multiple sequences (with discussion)," *Statistical Science*, 7, 457-511.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), Markov Chain Monte Carlo in Practice. Chapman & Hall Ltd (London; New York).

Gilks, W. R., Thomas, A., and Spiegelhalter, D. J. (1994), "A Language and a Program for Complex Bayesian Modeling," *The Statistician*, 43, 169-177.

Gilks, W. R., Wild, P. (1992), "Adaptive Rejection Sampling for Gibbs Sampling," *Appl. Statist.*, 41, 337-348.

Han, C. and Carlin, B. P. (2001), "Markov chain Monte Carlo methods for computing Bayes factors: A comparative review," *Journal of the American Statistical Association*, 96, 1122-1132.

Kass, R. E. and Raftery, A. E. (1995), "Bayes factors," *Journal of the American Statistical Association*, 90, 773-795.

Liu, S. I. (1995), "Comparison of forecasts for ARMA models between a random coefficient approach and a Bayesian approach," *Communications in Statistics, Theory and Methods*, 24(2), 319–333.

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

Nicholls, D. F. and Quinn, B. G. (1982), Random coefficient autoregressive models: An introduction. Springer-Verlag Inc (Berlin; New York)

Raftery, A. E. and Lewis, S. (1992), "How many iterations in the Gibbs sampler?" in *Bayesian Statistics 4*, ed. J. M. Bernardo et al, Oxford University press, pp 179-191.

Spiegelhalter, D. J., Thomas, A., Best and N., Lunn, D. (2001), WinBUGS Version 1.4. User Manual.

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

Tsay, R. (1987), "Conditional heteroscedastic time series models," *Journal of the American Statistical Association*. 82, 590-604.

Wang, D. (2003), "Frequentist and Bayesian analysis of Random Coefficient Autoregressive models," North Carolina State University Ph.D. dissertation.