Bayesian Analysis of ARMA models using Noninformative Priors

Frank Kleibergen* Henk Hoek†

Econometric Institute Report 9553/b

November 1996

Abstract

Parameters in AutoRegressive Moving Average (ARMA) models are locally nonidentified, due to the problem of root cancellation. Parameters can be constructed which represent this identification problem. We argue that ARMA parameters should be analyzed conditional on these identifying parameters. Priors exploiting this feature result in regular posteriors, while priors which neglect it result in posteriori favor of nonidentified parameter values. By considering the implicit AR representation of an ARMA model a prior with the desired proporties is obtained. The implicit AR representation also allows to construct easily implemented algorithms to analyze ARMA parameters. As a byproduct, posteriors odds ratios can be computed to compare (nonnested) parsimonious ARMA models. The procedures are applied to two datasets, the (extended) Nelson-Plosser data and monthly observations of US 3-month and 10 year interest rates. For approximately 50% of the series in these two datasets an ARMA model is favored above an AR model.

1 Introduction

Auto Regressive Moving Average (ARMA) models are a cornerstone of time series analysis, see a.o. Harvey (1981), and are commonly used in applied work. They do however possess

*Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, the Netherlands, Email:kleiberg@few.eur.nl. Part of this research was carried out while the author was hosted at the department of Econometrics at Tilburg University and during his PPS fellowship of NWO.

†Tinbergen Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, email: hoek@tir.few.eur.nl. We thank participants of the NBER/NSF Time Series Seminar, Rotterdam 1996, and ESEM, Istanbul 1996, and seminar participant at the Tinbergen Institute, CenTER and the University of Limburg for helpful comments and suggestions. The paper is available online at 'http:\\www.cs.few.eur.nl\kleiberg\publications.html'.

some well known problems. Maybe the best known problem is the problem of root cancellation, *i.e.* the autoregressive polynomial and the moving average polynomial have one or more roots in common. If root cancellation occurs, some AR and MA parameters are redundant as they do not affect the model. These parameters are said to be locally nonidentified. The problem of local nonidentification is common to many models in statistics and econometrics, see for example the Simultaneous Equation Model which is discussed in a.o. Phillips (1989). In the ARMA model, the local nonidentification problem implies certain conditionalization rules on the parameters. More specifically, hyper parameters can be constructed which, when restricted to zero, represent the presence of common factors. Therefore, ARMA parameters should be analyzed conditional on these hyper parameters.

The focus of the paper is on the conditional identification of ARMA parameters and its consequences for priors and posteriors in a Bayesian analysis. In this respect our approach differs from earlier Bayesian analysis of ARMA models like in, *e.g.*, Chib and Greenberg (1994), Monahan (1983) and Zellner (1971). Note that in a classical statistical setting the conditional identification problem is, although important, less of a problem than in a Bayesian analysis. This results from the fact that a classical researcher is mainly interested in a single point of the parameter space (*e.g.* the maximum likelihood estimate), while a Bayesian researcher analyzes the complete parameter space and is therefore confronted with all parameter values for which the identification problem occurs.

The paper is organized as follows. In section 2 the conditional identification of ARMA parameters is discussed. In section 3 we show that ignoring the identification problem may result in ill-behaved posterior densities. To overcome this problem priors are suggested which recognize that the ARMA parameters should be analyzed conditional on being identified, as represented by certain hyper parameters. A class of priors, which possess this property, leads to diffuse priors for the parameters of the implicit AR(∞) representation of the ARMA model. In order to analyze the ARMA model using these priors both Importance Sampling as well as Gibbs/Metropolis algorithms are developed in section 4. These sampling algorithms also yield posteriors of the hyper parameters which represent possible identification problems. A posterior odds ratio is proposed to compare different (parsimonious) ARMA models which have an equal number of parameters. Note that due to the identification problem, a general to specific modelling approach is not applicable. Section 5 contains an application of the developed procedures to two data sets, *i.e.* the extended Nelson-Plosser data and a data set consisting of monthly observations of U.S. 3-month and 10-year interest rates. For almost 50% of the series under consideration an ARMA model is favored above a pure AR model. In particular for price and interest rate series, there is strong evidence in favor of the ARMA model. Finally, section 6 summarizes and concludes.

2 Conditional Identification in ARMA Models

The problem of root cancellation (or common factors) is well known in the analysis of ARMA models, see, *e.g.*, Harvey (1981). Root cancellation leads to simplification of the ARMA model and to local nonidentification of redundant AR and MA parameters. To show this phenomenon, consider the "simplest" ARMA model, the ARMA$(1,1)$ model,

$$(1 \Leftrightarrow \rho L)y_t = (1 \Leftrightarrow \alpha L)\varepsilon_t, \tag{1}$$

where L is the lag-operator, $L^j y_t = y_{t-j}$ and ε_t is independently and identically distributed according to a Normal distribution with mean zero and variance σ^2, ε_t i.i.d. $N(0, \sigma^2)$, $t = 1, \ldots, T$. By considering the implicit AR(∞) and MA(∞) representations of this model,

$$\text{AR}(\infty) \quad : \quad y_t = \sum_{i=1}^{t+1} \alpha^{i-1}(\rho \Leftrightarrow \alpha)y_{t-i} + \varepsilon_t \quad \Leftrightarrow \quad (1 \Leftrightarrow \alpha L)(y_t \Leftrightarrow \varepsilon_t) = \theta y_{t-1} \tag{2}$$

$$\text{MA}(\infty) \quad : \quad y_t = \sum_{i=1}^{t+1} \rho^{i-1}(\rho \Leftrightarrow \alpha)y_{t-i} + \varepsilon_t \quad \Leftrightarrow \quad (1 \Leftrightarrow \rho L)(y_t \Leftrightarrow \varepsilon_t) = \theta \varepsilon_t, \tag{3}$$

where $\theta = \rho \Leftrightarrow \alpha$, local nonidentification can easily be recognized. In particular, depending on the specification used, ρ or α are nonidentified when $\theta = 0$, as in this case the model reduces to $y_t = \varepsilon_t$ independently of the value of either ρ or α. As a result, the likelihood function is flat and nonzero in the direction of ρ or α for zero values of θ. Use of a flat prior in a Bayesian analysis of the ARMA$(1,1)$ model, such that the posterior is proportional to the likelihood, therefore results in a flat and nonzero conditional posterior of ρ (or α) at $\theta = 0$. Consequently, the integral over this conditional posterior, and therefore also the marginal posterior of θ, is infinite at $\theta = 0$. So, the use of flat priors leads to an *a posteriori* favor for values of the ARMA parameters at which the local nonidentification problem occurs, which is neither a result of information from the prior nor from the data, but is a result of a model inadequacy. Similar arguments are used in Kleibergen and Van Dijk (1994a) and Kleibergen and Van Dijk (1994b) where similar phenomena are analyzed for cointegration and Simultaneous Equations Models. In section 4.3, the consequences of the use of a diffuse prior on the posterior of the parameters of an ARMA(1,1) model are illustrated. These posteriors are also compared with the posteriors using the priors derived in the following sections.

The parameter ρ or α is locally nonidentified when $\theta = 0$. The parameter θ is however identified for all possible values of either ρ or α. As a consequence, ρ or α should be analyzed conditional on θ and not vice versa. We explicitly focus on this point as it is important in the construction of Markov Chain Monte Carlo (MCMC) procedures in order to calculate the marginal posteriors. For example, the MCMC approach developed in Chib and Greenberg (1994) suffers from the local nonidentification problem. In this algorithm, the conditional posteriors $P(\alpha|\rho, \ldots)$ and $P(\rho|\alpha, \ldots)$ are used in a Gibbs sampling framework. As noted in the concluding remarks of Chib and Greenberg (1994), convergence of sample values fails if common factors or nearly common factors are present. As discussed above, the natural way of conditioning in an ARMA$(1,1)$ model is to analyze ρ or α conditional on θ. Consequently, the Gibbs sampler using the conditional posteriors $P(\alpha|\rho, \ldots)$ and $P(\rho|\alpha, \ldots)$ can lead to a reducible Markov Chain as the points of local nonidentification, $\alpha = \rho$, can form an absorbing state in the Markov Chain. Reducibility of the Markov Chain in Chib and Greenberg (1994) is avoided by the use of independent informative (Normal) priors for the ARMA parameters. Also *a priori* restricting the parameter space, for example to ensure stationarity and invertibility, avoids reducibility of the Markov Chain. However, in both cases convergence is still affected by the local nonidentification problem.

To show the local identification problem in the general ARMA(p, q) model,

$$\begin{aligned} \rho(L)y_t &= \alpha(L)\varepsilon_t \Leftrightarrow \\ (1 \Leftrightarrow \rho_1 L \Leftrightarrow \ldots \Leftrightarrow \rho_p L^p)y_t &= (1 \Leftrightarrow \alpha_1 \Leftrightarrow \ldots \Leftrightarrow \alpha_q L^q)\varepsilon_t, \end{aligned} \tag{4}$$

we again consider the AR(∞) representation[1] of this model

$$y_t = \sum_{i=1}^{t+p+q} c_i y_{t-i} + \varepsilon_t. \tag{5}$$

The coefficients of the AR(∞) representation are given by the following set of relations

$$c_0 = 1 \tag{6}$$

$$c_1 = \rho_1 \Leftrightarrow \alpha_1 \tag{7}$$

$$c_k = \sum_{i=1}^{\min(k,q)} \alpha_i c_{k-i} + \rho_k, \qquad k > 1, \tag{8}$$

where $\rho_k = 0,\ k > p$ and $\alpha_k = 0,\ k > q$, see, *e.g.*, Fuller (1976). If there is no MA component, $\alpha_i = 0, \forall i$, such that $c_k = \rho_k,\ k \leq p$ and $c_k = 0,\ k > p$. As a consequence, we can use the coefficients $c_k,\ k > p$ in order to perform inference on the MA parameters. In particular, it follows from (8) that the parameters $c_k,\ k > p + q$ are functions of the $c_k's,\ k \leq p + q$ only, such that inference on the $p + q$ parameters $\rho_1, \ldots, \rho_k, \alpha_1, \ldots, \alpha_q$ can be based on $c_1, \ldots, c_{p+q}$ solely. The relation between these parameters is given by the following matrix equation, which follows from the set of equations in (7) and (8),

$$C\vartheta = \mathbf{c}, \tag{9}$$

where $\vartheta = (\rho_1, \ldots, \rho_p, \alpha_1, \ldots, \alpha_q)'$, $\mathbf{c} = (c_1, \ldots, c_{p+q})'$,

$$C = \begin{pmatrix} I_p & C_{12} \\ 0_{q\times p} & C_{22} \end{pmatrix} \tag{10}$$

with I_p the identity matrix of dimension p,

$$C_{12} = \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ c_1 & 1 & 0 & \ldots & 0 \\ c_2 & c_1 & 1 & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ c_{p-1} & c_{p-2} & c_{p-3} & \ldots & 1 \end{pmatrix}, \tag{11}$$

[1] Equivalently, the MA(∞) representation can be considered

and

$$C_{22} = \begin{pmatrix} c_p & c_{p-1} & \dots & c_{p-q+1} \\ c_{p+1} & c_p & \dots & c_{p-q+2} \\ \vdots & \vdots & \vdots & \vdots \\ c_{p+q-1} & c_{p+q-2} & \dots & c_p, \end{pmatrix} \tag{12}$$

where $c_0 = 1$ and $c_k = 0,\ k < 0$. From this relation it follows

$$\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_q \end{pmatrix} = C_{22}^{-1} \begin{pmatrix} c_{p+1} \\ c_{p+2} \\ \vdots \\ c_{p+q} \end{pmatrix} \tag{13}$$

and

$$\begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_p \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_p \end{pmatrix} \Leftrightarrow C_{12} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_q \end{pmatrix}. \tag{14}$$

Note that if C_{22} doesnot have full rank, α and consequently ρ cannot be determined uniquely. This is a generalization of the local identification problem in the ARMA$(1,1)$ model. In order to test rank reduction of C_{22}, Galbraith and Zinde-Walsh (1995) propose a Wald test to test the hypothesis $H_0 : |C_{22}| = 0$. In our Bayesian approach, we examine the rank of C_{22} using the following LU decomposition [see also Kleibergen and Van Dijk (1994a) and Kleibergen and Van Dijk (1994b)]

$$C_{22} = \begin{pmatrix} \theta_{11} & 0 & 0 & \dots & 0 \\ \theta_{21} & \theta_{22} & 0 & \dots & 0 \\ \theta_{31} & \theta_{32} & \theta_{33} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \theta_{q1} & \theta_{q2} & \theta_{q3} & \dots & \theta_{qq} \end{pmatrix} \begin{pmatrix} 1 & \psi_{12} & \psi_{13} & \dots & \psi_{1q} \\ 0 & 1 & \psi_{23} & \dots & \psi_{2q} \\ 0 & 0 & 1 & \dots & \psi_{3q} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix}. \tag{15}$$

The rank of C_{22} is now given by the number of nonzero diagonal elements $\theta_{ii},\ i = 1, \dots, q$. Note that the number of zero θ_{ii}s only gives an indication of the number of common roots, and not of the required lag length of the individual AR or MA component. For example, if an ARMA$(1,1)$ is used to estimate an AR(1) model, $\theta_{11} = \rho \neq 0$, although the MA component is redundant.

In a Bayesian analysis of the ARMA(p,q) model, the use of diffuse priors again results in *a posteriori* favor for parameter values at which the local nonidentification occurs. Therefore, in the next sections we propose to change the base of the analysis from the ARMA(p,q) model to the (truncated) AR(∞) representation, in which all parameters are identified. Parameters are simulated from the AR(∞) representation and transformed to ARMA parameters. In this way, we work from the outset with a model with properly identified parameters such that no *a posteriori* favor for 'nonidentified' parameter combinations will result.

Finally, note that the autocorrelations of noninvertible MA models, *i.e.* models with one or more roots of the MA polynomial which lie within the unit circle, cannot be distinguished from the autocorrelations of invertible MA models. Consequently, MA parameters have to be restricted to 'invertible' parameter values, to be identifiable from the autocorrelations. Invertible and noninvertible MA polynomials with identical autocorrelations however lead to different values of the conditional likelihood function (given the first $p+q$ observations). As a result, they can be identified from the likelihood. As we define identification from a likelihood perspective, see also Kadane (1993), we allow for noninvertible MA parameters such that the MA and AR parameters range from $\Leftrightarrow\infty$ to ∞.

3 Priors for ARMA models

In the previous section it has been shown that ARMA parameters are identified conditional on nonzero values of specific (hyper) parameters. This conditionalization rule needs to be reflected in the priors. In particular, in the ARMA(p,q) model a prior for α should be specified conditional on C_{22} (or, equivalently, on the θ_{ii} parameters). This is satisfied by considering diffuse priors for the parameters of the truncated AR(∞) representation. These diffuse priors imply informative priors on the ARMA(p,q) which reflect the local nonidentification problem. In the next paragraphs we discuss prior specification in ARMA models, both without as well as with explanatory variables.

ARMA models without explanatory variables

First, consider the ARMA(p,q) model without explanatory variables. The AR(∞) representation of this model is given by

$$y_t = \sum_{i=1}^{t+p+q} c_i y_{t-i} + \varepsilon_t, \qquad t = 1, \ldots, T, \tag{16}$$

where $c_{p+q+i},\ i > 0$ is a function of $c_1, \ldots, c_{p+q}$,

$$c_{p+q+i} = (c_{p+1}, \ldots, c_{p+q}) C_{22}^{-1} \begin{pmatrix} c_{p+q+i-1} \\ c_{p+q+i-2} \\ \vdots \\ c_{p+i} \end{pmatrix} \tag{17}$$

and C_{22} has been defined in (12). The parameters $c_1, \ldots, c_{p+q}$ are always identified in this model, such that diffuse priors for these parameters can be considered

$$p(c_1, \ldots, c_{p+q}, \sigma^2) \propto \sigma^{-(p+q+2)} \tag{18}$$

The implicit prior on the AR and MA parameters is now given by

$$\pi(\rho_1, \ldots, \rho_p, \alpha_1, \ldots, \alpha_q, \sigma^2) \quad \propto \quad \sigma^{-(p+q+2)} \left| \frac{\partial(c_1, \ldots, c_{p+q})}{\partial(\rho_1, \ldots, \rho_p, \alpha_1, \ldots, \alpha_q)} \right|$$

$$= \quad \sigma^{-(p+q+2)}|C_{22}| = \sigma^{-(p+q+2)} \prod_{i=1}^{q} |\theta_{ii}|, \tag{19}$$

where $\theta_{ii},\ i = 1, \ldots, q$ is defined in (15). This shows that a diffuse prior on the implicit AR(∞) parameters results in a prior for the ARMA parameters $(\rho_1, \ldots, \rho_p, \alpha_1, \ldots, \alpha_q)$ conditional on the identifying parameters $\theta_{ii},\ i = 1, \ldots, q$. Note that this prior can also be derived using likelihood based arguments and therefore belongs to the class of Jeffreys' priors. These class of priors leads to posteriors which are invariant to parameter transformations. Also note that the prior in (19) penalizes parameter values for which the identification problem occurs. A diffuse prior on the ARMA parameters implies a highly informative prior on the parameters of the (truncated) AR(∞) specification, which are properly identified, and therefore leads to pathological posterior behavior.

Under the assumption of normally distributed disturbances ε_t, with mean 0 and variance σ^2, the posterior, which is proportional to the product of the prior and the likelihood function, is given by

$$p(\rho, \alpha, \sigma^2|y) \propto \sigma^{-(T+p+q+2)} \exp\left(\Leftrightarrow \frac{1}{2\sigma^2} \tilde{y}(\rho,\alpha)'\tilde{y}(\rho,\alpha)\right) \prod_{i=1}^{q} |\theta_{ii}|, \tag{20}$$

where $\rho = (\rho_1, \ldots, \rho_p)'$, $\alpha = (\alpha_1, \ldots, \alpha_q)'$, and $\tilde{y}(\rho, \alpha) = (\tilde{y}_1, \ldots, \tilde{y}_T)'$, where $\tilde{y}_t = c(L)y_t$ and $c(L) = \alpha(L)^{-1}\rho(L)$. Integrating with respect to σ^2 this posterior becomes

$$p(\rho, \alpha|y) \propto [\tilde{y}(\rho,\alpha)'\tilde{y}(\rho,\alpha)]^{-\frac{1}{2}(T+p+q)} \prod_{i=1}^{q} |\theta_{ii}|. \tag{21}$$

ARMA models with explanatory variabels

Additional explanatory variables can be incorporated in different ways. In general, the ARMA(p, q) model with explanatory variables becomes

$$\rho(L)y_t = \phi(L)x_t'\beta + \alpha(L)\varepsilon_t, \tag{22}$$

where $\phi(L)$ depends on $\rho(L)$ and $\alpha(L)$ and x_t is a $k \times 1$ vector of explanatory variables. Common choices for $\phi(L)$ are $\phi(L) := \rho(L)$, *i.e.* linear regression with ARMA(p, q) errors, and $\phi(L) := \alpha(L)$, *i.e.* the explanatory variables are not incorporated in the ARMA polynomial. If the MA polynomial is invertible, and the explanatory variables satisfy $x_t = Ax_{t-1},\ A : k \times k$, which holds for deterministic components like constant terms and trends, the marginal posterior of the ARMA parameters is not affected by the choice of $\phi(L)$. The location parameter vector β needs to be analyzed conditional on the ARMA parameters, resulting in the following Jeffreys' type prior

$$p(\beta|\rho, \alpha, \sigma^2) \propto \sigma^{-k} \left|\tilde{X}(\rho,\alpha)'\tilde{X}(\rho,\alpha)\right|^{\frac{1}{2}}, \tag{23}$$

where $\tilde{X}(\rho, \alpha)$ is a $T \times k$ matrix with tth rows given by $\tilde{x_t}' = \alpha(L)^{-1}\phi(L)x_t'$. Note that this prior is diffuse if $\phi(L) = \alpha(L)$. The prior in (23) results in a Normal conditional posterior of β,

$$p(\beta|\rho,\alpha,\sigma^2,y,X) \;\propto\; \sigma^{-k}\left|\tilde{X}(\rho,\alpha)'\tilde{X}(\rho,\alpha)\right| \exp\left(\Leftrightarrow\frac{1}{2\sigma^2}(\beta\Leftrightarrow\hat{\beta})'\left(\tilde{X}(\rho,\alpha)'\tilde{X}(\rho,\alpha)\right)(\beta\Leftrightarrow\hat{\beta})\right), \tag{24}$$

where

$$\hat{\beta} = \left(\tilde{X}(\rho,\alpha)'\tilde{X}(\rho,\alpha)\right)^{-1}\tilde{X}(\rho,\alpha)'\tilde{y}(\rho,\alpha), \tag{25}$$

where $\tilde{y}$ is defined below (21). Marginalizing with respect to β is straightforward and results in

$$p(\rho,\alpha,\sigma^2|y,X) \propto \sigma^{-(T+p+q+2)}\exp\left(\Leftrightarrow\frac{1}{2\sigma^2}\tilde{y}(\rho,\alpha)'M_{\tilde{X}(\rho,\alpha)}\tilde{y}(\rho,\alpha)\right)\prod_{i=1}^{q}|\theta_{ii}|, \tag{26}$$

where $M_Z = I \Leftrightarrow Z(Z'Z)^{-1}Z'$. Integration over σ^2 using proporties of the inverted gamma distribution leads to the marginal posterior of the ARMA parameters

$$p(\rho,\alpha|y,X) \propto \left[\tilde{y}(\rho,\alpha)'M_{\tilde{X}(\rho,\alpha)}\tilde{y}(\rho,\alpha)\right]^{-\frac{1}{2}(T+p+q)}\prod_{i=1}^{q}|\theta_{ii}|. \tag{27}$$

The marginal posteriors in (21) and (27) do not belong to a known class of probability density functions. Therefore, no analytical expressions for the posterior moments exists. In the next section Monte Carlo simulation procedures are constructed for the calculation of posterior moments. Also, posterior odds ratios to compare lag lengths of different ARMA polynomials are constructed.

4 Numerical Analysis of ARMA models

Monte Carlo simulation methods can be used to compute posterior moments of the marginal posteriors defined in (21) and (27). In Chib and Greenberg (1994) the marginal posteriors of the ARMA parameters are calculated using Gibbs Sampling. As discussed in section 2, the use of diffuse priors can lead to a reducible Markov Chain in the Gibbs Sampling algoritm because the local nonidentification problem is ignored. In this paper, we propose both an Importance Sampling [see, *e.g.*, Kloek and Van Dijk (1978) and Geweke (1989a)] and a Metropolis-Hastings [see, *e.g.*, Metropolis, Rosenbluth, Rosenbluth, Teller and Teller (1953) and Hastings (1970)] sampling procedure.

4.1 Importance Sampling

In order to construct an Importance sampling scheme we again exploit the AR(∞) representation of the ARMA(p,q) model. In particular, the importance function is chosen as a $(p+q)$ dimensional multivariate t density based on the (least squares) estimate of an AR($p+q$) model. The sampling scheme is given by

Importance Sampling Scheme for ARMA parameters

1. Choose the degrees of freedom of the Importance function, λ, the number of drawings, N and set $i = 1$.

2. Consider two cases, depending on the way the explanatory variables enter the analysis. First, assume $\phi(L) = \alpha(L)$. Next, estimate the model

$$y_t = c_1 y_{t-1} + \ldots + c_{p+q} y_{t-p-q} + x_t'\beta + u_t, \tag{28}$$

   to obtain $\hat{c}$ and $\mathrm{cov}(\hat{c})$. Second, assume $\phi(L) = \rho(L)$. An estimate of c is now obtained in two steps. First, we estimate

$$y_t = x_t'\beta + u_t, \tag{29}$$

   construct the LS residuals $\hat{u}_t$, and estimate c from

$$\hat{u}_t = c_1 \hat{u}_{t-1} + \ldots + c_{p+q} \hat{u}_{t-p-q} + \epsilon_t. \tag{30}$$

3. Generate $c_i,\ i = 1, \ldots, p+q$, from $q(c) = [\lambda + (c \Leftrightarrow \hat{c})'(\mathrm{cov}(\hat{c}))^{-1}(c \Leftrightarrow \hat{c})]^{-\frac{1}{2}(\lambda+p+q)}$.

4. Solve for ρ and α using (9).

5. Contruct weight: $w_i(\rho, \alpha) = \frac{\left[\tilde{y}(\rho,\alpha)' M_{\tilde{X}(\rho,\alpha)} \tilde{y}(\rho,\alpha)\right]^{-\frac{1}{2}(T+p+q)}}{[\lambda + (c-\hat{c})'(\mathrm{cov}(\hat{c}))^{-1}(c-\hat{c})]^{-\frac{1}{2}(\lambda+p+q)}}$.

6. Set $i = i + 1$, and if $i < N$ go to step 3.

7. Compute $E(g(\rho, \alpha)) = \frac{\sum_{i=1}^{N} w_i(\rho,\alpha) g(\rho,\alpha)}{\sum_{i=1}^{N} w_i(\rho,\alpha)}$.

8. To improve numerical accuracy, update $\hat{c}$ and $\mathrm{cov}(\hat{c})$ by considering $g(\rho, \alpha) = c$, set $i = 1$ and go to step 3.

Note that the prior on ρ, α, given in (19), doesnot explicitly appear in the importance weight as this prior is implicitly taken into account by the Jacobian of the transformation from the AR to the ARMA parameters. Alternatively, the importance weight may be interpreted as the ratio of the posterior of the AR(∞) model and the posterior of the AR($p+q$) model. A similar prior has been assumed in both models, *i.e.* a flat prior on the parameters $c_1, \ldots, c_{p+q}$. Note that sampling of β and σ^2 is straightforward using the conditional densities in (24) and (26). Given values of ρ and α we sample σ^2 and β from these conditional densities, and attach the same importance weight to these drawings as to ρ and α.

In step 4 of the algorithm the matrix C_{22} is needed. As a byproduct, this enables us to compute the diagonal elements of the lower diagonal matrix in (15), $\theta_{ii},\ i = 1, \ldots, q$.

These parameters show the identifiability of the MA parameters. In particular, if one of the θ_{ii}s is close to zero the matrix C_{22} is nearly singular and the constructed MA parameters, $\alpha = C_{22}^{-1}(c_{p+1}, \ldots, c_{p+q})'$ may be very large. In this case we expect that the posterior densities of the θ parameters are fat-tailed. Note that if the model is overspecified, *i.e.* p and/or q are chosen too large, this is likely to be the case. It is therefore difficult to perform a general to specific approach in the analysis of ARMA models. In the next paragraph we propose a posterior odds approach to compare (parsimonious) ARMA models with the same total number of parameters, *i.e.* $p+q$ is constant. Since these models are not nested in each other, a comparison using classical statistical analysis is difficult.

Posterior Odds

The θ_{ii} parameters combined with the AR and MA parameters, ρ and α, enable inference on the order of the AR and MA polynomials. However, it must be noted that an AR(k) model is able to explain the first k autocorrelations of any ARMA(p, q) model with $p+q = k$. It is therefore difficult to distinguish between, for example, an ARMA$(2,1)$ and an ARMA$(1,2)$ model, also because these models are not nested. Below we develop a posterior odds ratio in order to be able to compare these models. We assume that the exogenous variables enter both models in the same way. Also, we assume identical priors, prior odds and parameter regions for the implicit AR parameters, $c_1, \ldots, c_{p+q}$, such that the posterior odds ratio is not dependent on the choice of the prior. The posterior odds ratio is given by

$$\mathrm{POR}(\mathrm{H}_1|\mathrm{H}_2) = \frac{\int p(c|y, \mathrm{H}_1)dc}{\int p(c|y, \mathrm{H}_2)dc}, \tag{31}$$

where $c = c_1, \ldots, c_{p+q}$ and H_1 and H_2 represent two different ARMA models with the same number of parameters, *i.e.* $p_{\mathrm{H}_1} + q_{\mathrm{H}_1} = p_{\mathrm{H}_2} + q_{\mathrm{H}_2}$. By comparing parsimonious models, for which all parameters are expected to be different from 0, the difficulties of the general to specific modelling approach can be avoided. For example, we compare ARMA$(2,1)$, ARMA$(1,2)$, AR(3) and MA(3) models with each order, as these models are equally capable in explaining the first three autocovariances but differ for higher order autocovariances. A general to specific approach would start with an ARMA$(3,3)$ model, in which it is particularly hard to identify the MA parameters as the AR parameters can explain part of the (short run) behavior resulting from a MA polynomial.

The posterior odds ratio in (31) can be calculated using Importance Sampling, see Geweke (1989b). The posterior odds ratio equals the ratio of marginal likelihoods under both models. In Geweke (1989a) it is shown that

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} w_i(\rho, \alpha) \Leftrightarrow \frac{\int p(c|y)dc}{\int q(c)dc}\right) \Rightarrow N(0, \omega), \tag{32}$$

where $p(\cdot)$ is the (unnormalized) posterior, $q(\cdot)$ is the importance density, $\Rightarrow$ indicates weak convergence, and $\omega = E((w(\vartheta) \Leftrightarrow E(w(\vartheta)))^2)$, which can be estimated by $\omega \approx$

$\frac{1}{N}\sum_{i=1}^{N} w_i(\rho,\alpha)^2 \Leftrightarrow (\frac{1}{n}\sum_{i=1}^{N} w_i(\rho,\alpha))^2$. Equation (32) can be used to estimate the marginal likelihood

$$\int p(c|y)dc \approx \left(\int q(c)dc\right) \times \left(\frac{1}{N}\sum_{i=1}^{N} w_i(\rho,\alpha)\right) \tag{33}$$

Note that sofar we represented the posterior by its kernel, without the normalizing constants. In the construction of the posterior odds ratio however we need to include these normalizing constants. Doing this, the posterior odds ratio is approximated by

$$\mathrm{POR}(\mathrm{H}_1|\mathrm{H}_2) \approx \frac{\frac{1}{N_1}\sum_{i=1}^{N_1} w_i(\rho,\alpha,\mathrm{H}_1),\ (\frac{1}{2}(\lambda_2+p+q)),\ (\frac{1}{2}\lambda_1)\lambda_2^{\frac{1}{2}\lambda_2}}{\frac{1}{N_2}\sum_{i=1}^{N_2} w_i(\rho,\alpha,\mathrm{H}_1),\ (\frac{1}{2}(\lambda_1+p+q)),\ (\frac{1}{2}\lambda_2)\lambda_1^{\frac{1}{2}\lambda_1}} \left(\frac{|\mathrm{cov}(\hat{c}_1)|}{|\mathrm{cov}(\hat{c}_2)|}\right)^{\frac{1}{2}} \tag{34}$$

where $w_i(\rho,\alpha,\mathrm{H}_j)$ are the weights for model j, N_j is the number of Importance Sampling drawings from model j, λ_j is the degrees of freedom of the Importance function used for model j and $\mathrm{cov}(\hat{c}_j)$ is the covariance matrix of the Importance functions used for model j. If $\lambda_1 = \lambda_2$ the weight ratio approximating the posterior odds ratio simplifies to,

$$\mathrm{POR}(\mathrm{H}_1|\mathrm{H}_2) \approx \frac{\frac{1}{N_1}\sum_{i=1}^{N_1} w_i(\rho,\alpha,\mathrm{H}_1)}{\frac{1}{N_2}\sum_{i=1}^{N_2} w_i(\rho,\alpha,\mathrm{H}_2)} \left(\frac{|\mathrm{cov}(\hat{c}_1)|}{|\mathrm{cov}(\hat{c}_2)|}\right)^{\frac{1}{2}} \tag{35}$$

Further simplifications are possible if one of the models is an AR model, in which case the corresponding integral can be evaluated analytically. In section 5 we apply the posterior odds ratio approach to compare different ARMA models for the extended Nelson-Plosser data, see Schotman and Van Dijk (1993), and for monthly observations of 3-month and 10 year US interest rates.

4.2 Metropolis-Hastings Sampling

Instead of Importance Sampling, we could also use the Gibbs sampler in combination with the Metropolis-Hastings (MH) algorithm, see, *e.g.* Chib and Greenberg (1995). We consider the general model

$$\rho(L)y_t = \phi(L)x_t'\beta + \alpha(L)\varepsilon_t \tag{36}$$

and define $\tilde{y}(\rho,\alpha) = \alpha(L)^{-1}\rho(L)y$ and $\tilde{X}(\rho,\alpha) = \alpha(L)^{-1}\phi(L)$. The Gibbs/Metropolis sampling algorithm can now be set up as follows

Gibbs/Metropolis Sampling Scheme for ARMA parameters

1. Choose starting values $\sigma^0, \beta^0, \phi^0, \theta^0$, the number of iteration, N, and set $i = 1$. Note that also $c^0 := c_1^0, \ldots, c_{p+q}^0$ is implicitly chosen.

2. Given ρ and α, the model is linear in β,

$$\tilde{y}(\rho, \alpha) = \tilde{X}(\rho, \alpha)\beta + \varepsilon, \tag{37}$$

   where $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_T)'$, ε_t i.i.d. $N(0, \sigma^2)$, such that β^i can be generated from a Normal distribution with mean $\hat{\beta}$ and variance matrix $\mathrm{cov}(\hat{\beta})$, which are computed by Least Squares regression of (37).

3. Again, consider two cases, depending on the way the explanatory variables enter the analysis. First, assume $\phi(L) = \alpha(L)$. Next, estimate the model

$$y_t \Leftrightarrow x_t'\beta^i = c_1 y_{t-1} + \ldots + c_{p+q} y_{t-p-q} + u_t, \tag{38}$$

   to obtain $\hat{c}$ and $\mathrm{cov}(\hat{c})$. Second, assume $\phi(L) = \rho(L)$. Construct $u_t = y_t \Leftrightarrow x_t'\beta$ and estimate c and $\mathrm{cov}(\hat{c})$ from

$$u_t = c_1 u_{t-1} + \ldots + c_{p+q} u_{t-p-q} + \epsilon_t. \tag{39}$$

4. The probing density in the MH step is given by $N(\hat{c}, \mathrm{cov}(\hat{c}))$. Generate a candidate c^{new} from this density, transform c^{new} to ρ^{new} and α^{new}, and apply the following acceptance probability

$$\psi = \frac{w(\rho^{\text{new}}, \alpha^{\text{new}}, c^{\text{new}})}{w(\rho^{i-1}, \alpha^{i-1}, c^{i-1})} \tag{40}$$

   where

$$w(\rho, \alpha, c) = \frac{L(\rho, \alpha | y, \beta, \sigma^2)}{N(c | \hat{c}, \hat{V}_{\hat{c}})} \tag{41}$$

   where $L(\cdot)$ is the likelihood of the ARMA model,[2]

$$L(\rho, \alpha | y, \beta, \sigma^2) \propto \prod_{t=1}^{T} \exp\left(\Leftrightarrow \frac{\varepsilon_t^2}{2\sigma^2}\right), \tag{42}$$

   with $\varepsilon_t := \alpha(L)^{-1}(\rho(L)y_t \Leftrightarrow \phi(L)x_t'\beta)$. Note that the MH acceptance probability can be interpreted as the ratio of the importance weight in the model with given β and σ. Next, with probability ψ we set $(\rho^i, \alpha^i, c^i) = (\rho^{\text{new}}, \alpha^{\text{new}}, c^{\text{new}})$ and with probability $(1 \Leftrightarrow \psi)$ $(\rho^i, \alpha^i, c^i) = (\rho^{i-1}, \alpha^{i-1}, c^{i-1})$.

5. Conditional on ρ, α and β, σ^2 has an inverted Gamma distribution. Generate σ^2 from this distribution.

6. If $i < N$ set $i = i + 1$ and go to step 2.

Note again that the identifying parameters θ_{ii} are obtained as a byproduct in step 3, such that also the posterior for these parameters can be obtained from the Gibbs sampler.

[2] Note that again the prior is implicitly taken acount of by the Jacobian of the transformation.

prior \ parameter	ρ	α	θ
diffuse on (ρ, α)	0.32 0.49	0.19 0.49	0.12 0.068
diffuse on (c_1, c_2)	0.38 0.37	0.22 0.36	0.16 0.062

Table 1: Posterior moments ARMA$(1,1)$ parameters artificial time series

4.3 An Example

To illustrate the consequences of specific priors on either the ARMA parameters or the implied AR parameters, we compare the posteriors of the ARMA parameters for an artificial time series. This series is generated from an ARMA$(1,1)$ model, see (1), with parameters $\rho = 0.6$, $\alpha = 0.4$, $\sigma^2 = 0.005$, $T = 200$. Note that the identifying parameter $\theta = \rho \Leftrightarrow \alpha$ equals 0.2. We calculated the posteriors of the parameters of an ARMA(1,1) model both using a diffuse prior on (ρ, α) and a diffuse prior on (c_1, c_2). For the diffuse prior on (ρ, α) $(p(\rho, \alpha) \propto \sigma^{-4})$, the posteriors are calculated using the analytical expression of the bivariate posterior of (ρ, α), which is proportional to the conditional likelihood. For the diffuse prior on (c_1, c_2) $(p(c_1, c_2) \propto \sigma^{-4} \Rightarrow p(\rho, \alpha) \propto \sigma^{-4}|\rho \Leftrightarrow \alpha|)$, the Importance Sampling Algorithm from section 4.1 is used.

Figures 1 to 7 contain the marginal posteriors of the parameters of an ARMA(1,1) model for the artificially generated time series. Table 1 contains the posterior means and standard deviations of the different parameters. The bivariate posterior of θ and α and its contourlines are shown in figures 1 and 2 (diffuse prior on (ρ, α)) and 3 and 4 (diffuse prior on (c_1, c_2)). The bivariate posterior and its contourlines show that the bivariate posterior using the diffuse prior on (ρ, α) is constant in the direction of α around $\theta = 0$. The posterior using the diffuse prior has much more probability mass at $\theta = 0$ compared to the posterior using a diffuse prior on (c_1, c_2). The marginal posteriors of θ shown in figure 5 confirm this as the marginal posterior using the diffuse prior on (ρ, α) has a secondary mode at $\theta = 0$ such that this posterior has more probability mass at $\theta = 0$. Theoretically the value at $\theta = 0$ of this posterior is infinite as we have integrated over a parameter, α, which does not influence the nonzero joint posterior of (θ, α) at $\theta = 0$. We have chosen a finite parameter region of α, $(\Leftrightarrow 1.3, 1.3)$, however, such that the posterior in figure 5 is finite at $\theta = 0$ as the integral of a constant function over a finite region is finite. Note the direct linkage between the size of the parameter region and the posterior value at $\theta = 0$. As a consequence, the use of a diffuse prior on (ρ, α) results in an implicit favor for $\theta = 0$. The larger probability mass at $\theta = 0$ is also reflected in the marginal posterior of α and ρ, shown in figures 6 and 7. For both figures, it holds that the marginal posterior using the diffuse prior on (ρ, α) has much fatter tails and also shows some irregularities at the boundary of the stationary (invertible) parameter region, see also DeJong and Whiteman (1993) and Sargan and Bhargava (1983). The posteriors using the diffuse prior on (c_1, c_2) have a more regular behavior.

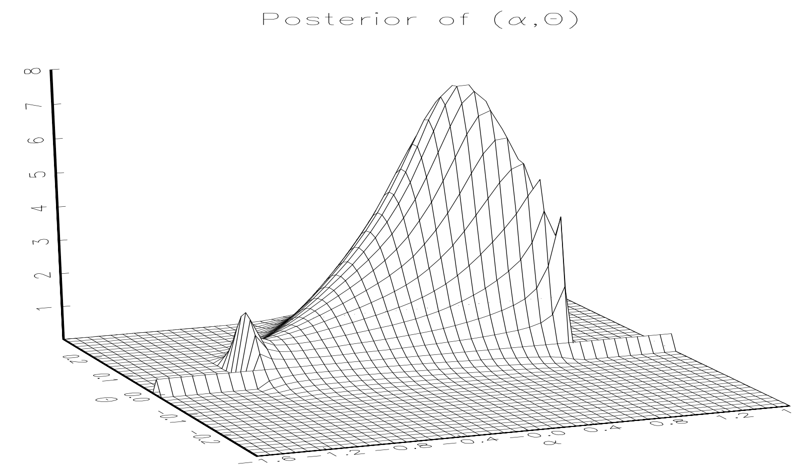

Figure 1: Bivariate posterior (α, θ), artificial time series, diffuse prior on (ρ, α). sample

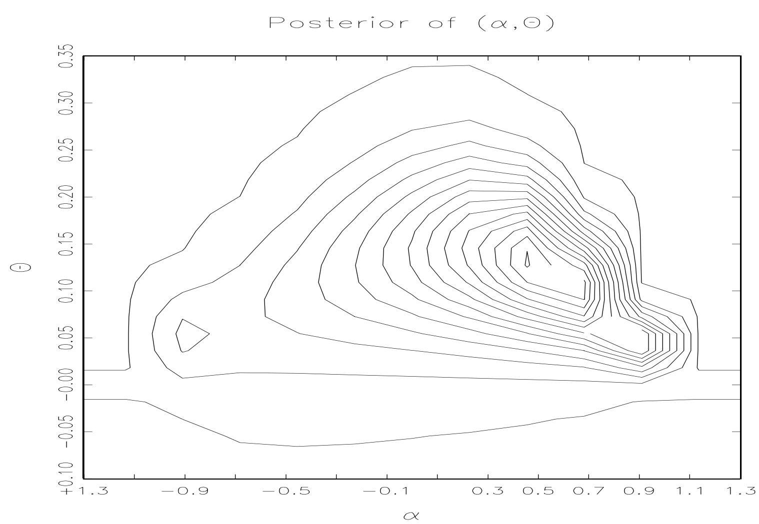

Figure 2: Contourlines bivariate posterior (α, θ), artificial time series, diffuse prior on (ρ, α). sample

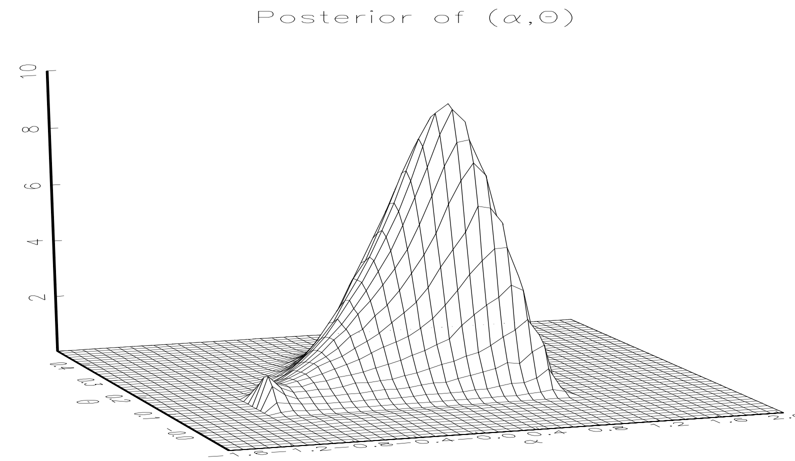

Figure 3: Bivariate posterior (α, θ), artificial time series, diffuse prior on (c_1, c_2). sample

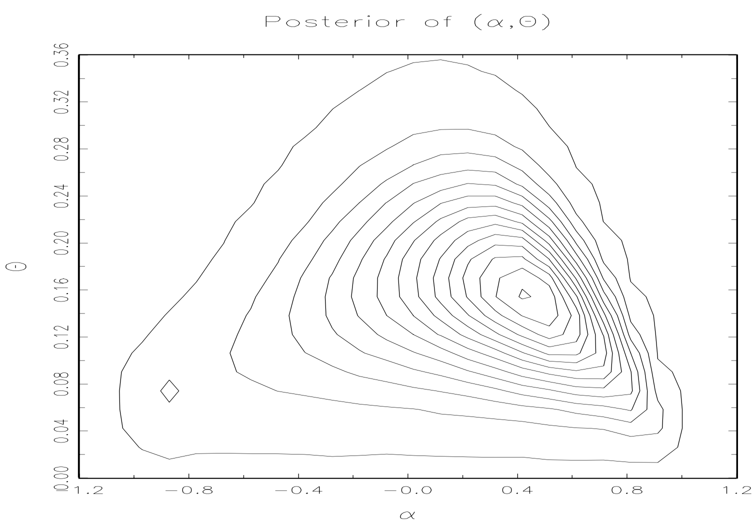

Figure 4: Contourlines bivariate posterior (α, θ), artificial time series, diffuse prior on (c_1, c_2). sample

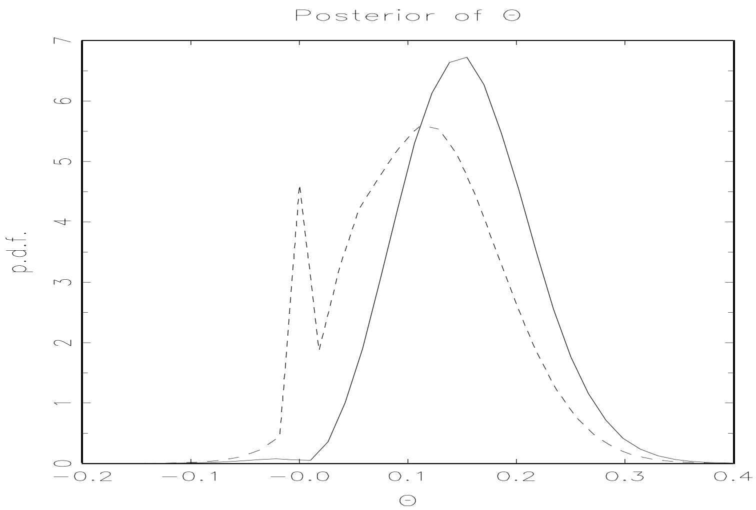

Figure 5: Marginal posterior of θ, artificial time series, $p(\alpha, \rho) \propto 1$: - -, $p(c_1, c_2) \propto 1$: —.
sample

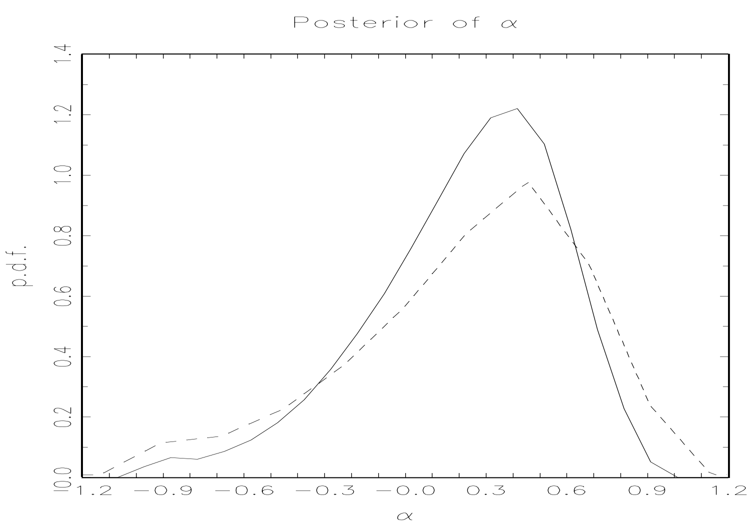

Figure 6: Marginal posterior of ρ, artificial time series, $p(\alpha, \rho) \propto 1$: - -, $p(c_1, c_2) \propto 1$: —.
sample

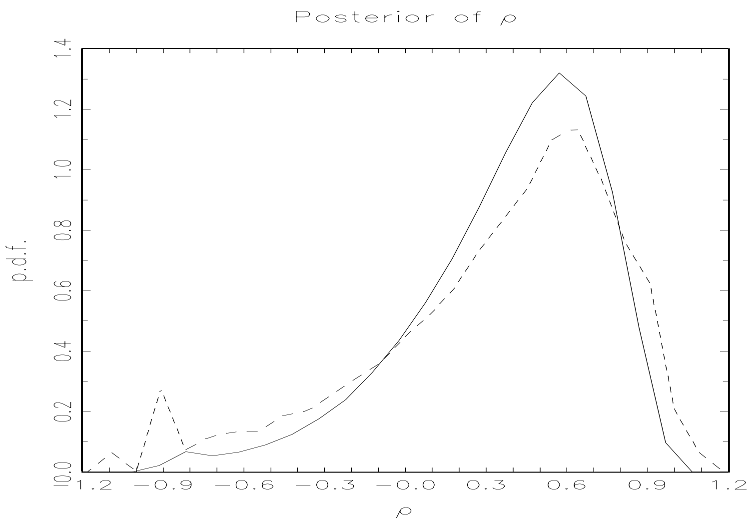

Figure 7: Marginal posterior of α, artificial time series, $p(\alpha, \rho) \propto 1$: - -, $p(c_1, c_2) \propto 1$: —. sample

5 Empirical Application

To show the applicability of the derived theory and simulation procedures, we applied them to two data sets. First, we consider the extended Nelson-Plosser data. This data set consists of yearly observations of 14 macroeconomic variables. The original sample period ended in 1970 (see Nelson and Plosser (1982)), but the sample period has been extended to 1988 (see Schotman and Van Dijk (1993)). The second data set consists of monthly observations from January 1957 to April 1989 of the U.S. three month treasury bill rate and of interest rates having a maturity of ten years. We start by analysing the first data set.

We model the (extended) Nelson-Plosser series using ARMA models with three ARMA parameters ($p + q = 3$). Following previous analysis of these series a constant term and a trend variable are included in the model,

$$\rho(L)(y_t \Leftrightarrow \mu \Leftrightarrow \gamma t) = \alpha(L)\varepsilon_t, \tag{43}$$

where $\rho(L) = 1 \Leftrightarrow \rho_1 L \Leftrightarrow \ldots \Leftrightarrow \rho_p L^p$, $\alpha(L) = 1 \Leftrightarrow \alpha_1 L \Leftrightarrow \ldots \Leftrightarrow \alpha_q L^q$, $\varepsilon_t \sim N(0, \sigma^2)$. Note that the marginal posteriors of the ARMA parameters are not affected by the specification of the dynamic structure for the exogenous variables, as discussed in section 3. By considering a diffuse prior on the first three parameters on the AR(∞) representation of the model, we constructed the posterior odds ratios using the average weights resulting from the Importance Sampling Algorithm, as discussed in the previous section. The Impor-

Series \ ARMA order	3,0/2,1	3,0/1,2	0,3/3,0	2,1/1,2	0,3/2,1	0,3/1,2
Real GNP	0.969	1.082	0.003	1.117	0.003	0.003
Nominal GNP	1.019	1.422	0.000	1.395	0.000	0.000
GNP Capita	0.975	1.091	0.005	1.119	0.005	0.005
Indus. Prod.	0.638	0.842	0.000	1.320	0.000	0.000
Employment	0.549	0.844	0.000	1.537	0.000	0.000
Unemploy.	0.069	0.166	0.420	2.418	0.029	0.070
GNP Def.	1.682	6.821	0.000	4.055	0.000	0.000
Cons. Price Ind.	0.219	0.638	0.000	2.915	0.000	0.000
Wages	0.852	1.338	0.000	1.570	0.000	0.000
Real Wages	0.795	0.951	0.000	1.197	0.000	0.000
Money	0.923	14.73	0.000	15.96	0.000	0.000
Velocity	1.020	1.005	0.000	0.985	0.000	0.000
Interest	0.301	0.340	0.000	1.127	0.000	0.000
S&P 500	0.694	0.846	0.000	1.220	0.000	0.000

Table 2: Posterior Odds Ratios Extended Nelson-Plosser series

tance Sampling Algorithm converges very fast and because of the good approximation of the posterior by the Importance function, the Importance function could even be used for direct acceptance-rejection sampling from the posterior. We performed this exercise for all ARMA models containing three ARMA parameters. Posterior odds ratios are calculated for ARMA(3,0) [=AR(3)], ARMA(2,1), ARMA(1,2) and ARMA(0,3) [=MA(3)] models. The resulting posterior odds ratios are listed in table 2. We also approximated the posterior odds ratios using the Schwarz (Bayesian) Information Criterium (BIC), see Schwarz (1978), $POR(H_1,H_2) \approx \exp[\frac{1}{2}(BIC(H_2) \Leftrightarrow BIC(H_1))]$, of which we obtained estimates from MICROTSP. For the series for which MICROTSP was capable to give reasonably precise parameter estimates, the posterior odds ratios from both procedures are close to one another. For the nonprecise estimates, the posterior odds ratios were rather different as the posterior odds ratios resulting from the BIC's are inprecise. The numerical errors for the posterior odds ratios resulting from the Importance Sampling are also in these cases very small such that we prefer this latter procedure for calculating the posterior odds ratios.

The Posterior Odds Ratios from table 2 are quite surprising as for most of the series, an ARMA(2,1) model is preferred above an AR(3) model. A possible explanation for this phenomenon could be that many series consist of time averages which introduces MA errors in the series. For some series, the ARMA(2,1) model is clearly preferred above an AR(3) model given the value of the posterior odds ratios. This holds for example for Industrial Production, Employment, Unemployment, Consumer Price Index, Interest and the Standard and Poor 500. For other series the posterior odds ratios indicate that both models are more or less equally likely. The ARMA(2,1) model can also be approximated by a high order AR model but an important difference between AR and MA components lies in their consequences for the long run behavior of the series. In particular, MA components

series \ ARMA par.	ρ_1	ρ_2	ρ_3	α_1	θ_{11}	$\rho = \sum_{i=1}^{p} \rho_i$
Real GNP	1.18 0.23	⇔0.37 0.21		⇔0.07 0.22	0.46 0.15	0.81 0.062
Nominal GNP	1.45 0.12	⇔0.57 0.20	0.063 0.12			0.94 0.032
GNP Capita	1.17 0.24	⇔0.37 0.21		⇔0.062 0.23	0.45 0.14	0.80 0.06
Ind. Prod.	0.69 0.32	0.075 0.27		⇔0.29 0.30	0.21 0.10	0.77 0.08
Employment	0.97 0.22	⇔0.14 0.21		⇔0.33 0.21	0.57 0.16	0.82 0.061
Unemploy.	0.41 0.18	0.15 0.16		⇔0.66 0.16	0.55 0.14	0.56 0.10
GNP Def.	1.43 0.11	⇔0.38 0.18	⇔0.09 0.11			0.97 0.02
Cons. Price Ind.	1.36 0.12	⇔0.38 0.12		⇔0.47 0.12	1.24 0.18	0.99 0.015
Wages	1.27 0.20	⇔0.35 0.19		⇔0.23 0.19	0.70 0.18	0.93 0.035
Real Wages	0.93 0.34	⇔0.018 0.33		⇔0.30 0.30	0.38 0.14	0.91 0.056
Money	1.50 0.14	⇔0.56 0.14		⇔0.19 0.16	0.89 0.20	0.93 0.027
Velocity	1.09 0.094	⇔0.15 0.14	0.026 0.093			0.97 0.025
Interest	0.72 0.22	0.20 0.21		⇔0.54 0.19	0.47 0.16	0.92 0.052
S&P 500	0.80 0.22	0.094 0.21		⇔0.42 0.20	0.42 0.13	0.89 0.05

Table 3: Posterior Moments ARMA parameters Nelson-Plosser series

have autocorrelations which abruptly die out while the autocorrelations of AR components decrease exponentially. So, it is interesting to investigate the influence of the MA parameters on the parameters reflecting the long run behavior of the analyzed series, like the unit root parameter, $\sum_{i=1}^{p} \rho_i$. We perform such an analysis and the results are listed in table 3, which contains the posterior means and standard deviations (given below the means) of the ARMA model that is preferred by the posterior odds ratios from table 2. Note that a MA(3) model is implausible for all series since this model leads to a very restricted type of long run behavior of the analyzed series.

For all series, except the Consumer Price Index (CPI), the MA parameter, α_1, has a positive correlation with the unit root parameter. The posterior mean of the unit root parameter of the ARMA(2,1) is, therefore, for all series, except CPI, smaller than the posterior mean of the unit root parameter of the AR(3) model. Depending on the size of the MA parameter, this decrease of the MA parameter can be quite large and it is most pronounced for the unemployment series. For this series, the unit root parameter decreases from 0.74 to 0.56. For the other series, which contain significant MA components, the decrease is also relatively large: Industrial Production (0.06), Employment (0.05), Interest (0.03), S&P 500 (0.04). Also, for all series the posterior standard deviations increase slightly from AR(3) to ARMA(2,1). It is typical that the series which vary a lot, like CPI and Interest, contain large MA components. When combined with an AR component, these MA components can explain the long run memory in the first differences of these series, like inflation.

The parameter θ_{11}, see equation (15) for an interpretation of this parameter, shows that for the series for which an ARMA(2,1) model is preferred, the MA parameter, α_1, is identified as the posterior mean of θ_{11} does not lie relatively close to 0. Exceptions are the series of Industrial Production and Velocity. For the velocity series, an AR(3) model is preferred. For Industrial Production, there is some posterior probability for zero values of θ_{11} leading to fat tailed behavior of the posteriors. This behavior disappears when we consider an ARMA$(1,1)$ model, which is sensible since the posterior mean of ρ_2 lies close to 0. In the resulting ARMA(1,1), α_1 is properly identified, see table 5. If the posteriors of an ARMA$(2,1)$ model for velocity are calculated, the posterior of θ_{11} has a considerable amount of probability mass close to zero leading to fat tailed posteriors for the other parameters. This also indicates that an ARMA(2,1) is not the appropriate model for velocity, which can also be concluded from the posterior odds ratios from table 2.

Since for many series contained in table 2, the posterior means indicate that either ρ_2 or/and α_1 lies close to zero, we calculated the posterior odds ratios of an AR(2) model compared to an ARMA$(1,1)$ model for these series. The resulting posterior odds ratios are listed in table 4.

series \ odds	2,0/1,1
Real GNP	5.212
Nominal GNP	3.105
Indus. Prod.	0.770
Employ.	0.741
Wages	3.819
Real Wages	0.942
Money	671.3
S&P 500	0.306

Table 4: Posterior Odds for AR(2) vs. ARMA(1,1) Nelson-Plosser series

Table 4 shows that Industrial Production, Employment, Real Wages and S&P 500 are better characterized by an ARMA(1,1) than a AR(2) model according to the Posterior Odds Ratios. The opposite holds for Real GNP, Nominal GNP, Wages and Money. This accords with the results in tables 2 and 3 which show that these series are either preferred to be AR(3) or the MA parameter α_1 lies relatively closer to 0 than the AR parameter ρ_2. Table 5 shows the posterior moments of the parameters of the resulting ARMA(1,1) models.

Table 5 shows that the summed posterior mean changes of ρ_1 and α_1 of the ARMA(1,1) model compared to ARMA(2,1) model approximately equal the posterior mean of ρ_2 in the ARMA(2,1) model. Since the identifying parameter θ_{11} differs much more from 0 than

series \ parameter	ρ_1	α_1	θ_{11}
Ind. Prod.	0.79 0.06	⇔0.18 0.11	⇔0.97 0.09
Employ.	0.82 0.06	⇔0.43 0.09	⇔1.25 0.09
Real Wages	0.92 0.05	⇔0.28 0.12	⇔1.18 0.11
S&P 500	0.89 0.05	⇔0.31 0.14	⇔1.21 0.10

Table 5: Posterior moments of ARMA(1,1) model for Nelson-Plosser series

in the ARMA(2,1) model, the posterior standard deviations of the parameters are much smaller than in the ARMA(2,1) model. It is typical that the posterior standard deviation of the unit root parameter is however similar in both models, indicating that the information regarding the long run behavior is not by affected by deleting ρ_2.

We also calculated the posteriors of the parameters of ARMA models for U.S. short and long term interest rates. Again the orders of the ARMA models, $p+q$, are supposed to equal 3. In contrast to the Nelson-Plosser data we only consider a constant term in the regression equation. To determine the favoured univariate ARMA model for both interest rates we calculated the posterior odds ratios for all models with ARMA order, $p+q$, equal to 3. These posterior odds ratios are listed in table 6.

series \ ARMA order	3,0/2,1	3,0/1,2	0,3/3,0	2,1/1,2	0,3/1,2	0,3/2,1
short (3 month)	5.1023	0.9976	0.0000	0.1943	0.0000	0.0000
long (10 year)	0.6637	0.3606	0.0000	0.5434	0.0000	0.0000

Table 6: Posterior Odds Ratios Interest Rate Series

The posterior odds ratios show that an ARMA(1,2) model is equally likely for the short term interest rates as an AR(3) model. This is rather typical as the ARMA(2,1) model is less likely than these other two models. For the long term interest rate an ARMA(1,2) model is favored. Table 7 lists the posterior means and standard deviations of the parameters of the models which are preferred by the posterior odds ratios.

The posterior moments in table 7 show that the ARMA(1,2) model for long term interest rate has properly identified MA parameters as both identifying parameters θ_{11} and θ_{22} have almost no probability mass at 0 as indicated by the posterior means and standard deviations of these parameters. As the MA parameters of the ARMA model for the long term interest rates are not close to 0, the long run behavior of the long term interest rate will significantly differ from an standard random walk model. Furthermore since the AR parameter of the long term interest rate, ρ_1, lies close to 1, the long term interest rate can be characterized by an IMA(2) model (i.e. random walk plus noise model).

series \ ARMA par.	ρ_1	ρ_2	ρ_3	α_1	α_2	θ_{11}	θ_{22}	$\rho = \sum_{i=1}^{p} \rho_i$
short (3 month)	0.99 0.051	⇔0.16 0.072	0.13 0.051					0.976 0.012
short (3 month)	0.978 0.011			⇔0.032 0.052	0.13 0.052	⇔0.85 0.051	⇔1.01 0.051	0.978 0.011
long (10 year)	0.99 0.006			⇔0.41 0.05	0.14 0.05	⇔0.89 0.04	⇔1.4 0.05	0.99 0.006

Table 7: Posterior Moments ARMA models interest rates

6 Conclusions

It has been shown that parameters in ARMA models are identified conditional on the value of other (hyper) parameters. This implies that parameters have to be analyzed conditional on these identifying parameters. Priors have to incorporate this feature in order to lead to regular posteriors of the parameters. A class of priors, which accomplishes this, implies diffuse (or natural conjugate) priors for the first $p+q$ parameters of the implicit AR(∞) representation of the ARMA(p,q) model. As an approximation of the AR(∞) model we consider an AR($p+q$) model. The posterior of the parameters in the AR($p+q$) model is used as an importance function in an Importance Sampling framework. Also, a Metropolis-Hastings sampling algorithm is constructed. For the conducted applications, the Importance Sampling Algorithm converged rapidly. Quite surprisingly, in the applications we found that many series, which are traditionally modelled using AR models, contain strong MA components. These MA components can influence the long run parameters such that the use of MA components can be important for forecasting purposes, see also Franses and Kleibergen (1995).

In future work, we extend the analysis to ARMA models containing seasonal lags and Vector ARMA models. Also, by considering the Metropolis-Hastings algorithm, extensions of the model by, *e.g.*, structural changes, can be analyzed in a Gibbs Sampling framework.

References

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

Chib, S. and E. Greenberg (1995), 'Understanding the Metropolis Algorithm', *The American Statistician*, **49**, 327–335.

DeJong, D. N. and C. H. Whiteman (1993), 'Estimating Moving Average Parameters: Classical Pileups and Bayesian Posteriors', *Journal of Business and Economic Statistics*, **64**, 183–206.

Franses, P.H. and F. Kleibergen (1995), 'Unit Roots in the Nelson-Plosser Data: Do They Matter for Forecasting?', *International Journal of Forecasting*, **12**, 283–288.

Fuller, W.A. (1976), *Introduction to Statistical Time Series*, Wiley (New York).

Galbraith, J. W. and V. Zinde-Walsh (1995), 'Simple Estimation and Identification Techniques for General ARMA Models', Working Paper.

Geweke, J. (1989a), 'Bayesian Inference in Econometric Models using Monte Carlo Integration', *Econometrica*, **57**, 1317–1339.

Geweke, J. (1989b), 'Exact Predictive Densities for Linear Models with ARCH Disturbances', *Journal of Econometrics*, **40**, 63–86.

Harvey, A. C. (1981), *Time Series Models*, Philip Allan (London).

Hastings, W.K. (1970), 'Monte Carlo Sampling using Markov Chains and their Applications', *Biometrika*, **57**, 97–10.

Kadane, J.B. (1993), The Role of Identification in Bayesian Theory, *in* Fienberg, S.E. and A. Zellner, eds, 'Studies in Bayesian Econometrics and Statistics', North-Holland (Amsterdam), pp. 341–361.

Kleibergen, F. and H.K. Van Dijk (1994a), 'Direct Cointegration Testing in Error Correction Models', *Journal of Econometrics*, **63**, 61–103.

Kleibergen, F. and H.K. Van Dijk (1994b), 'On the Shape of the Likelihood/Posterior in Cointegration Models', *Econometric Theory*, **10**, 514–551.

Kloek, T. and H.K. Van Dijk (1978), 'Bayesian Estimates of Equation System Parameters: An Application of Iintegration by Monte-Carlo', *Econometrica*, **44**, 345–351.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller and E. Teller (1953), 'Equations of State Calculations by fast Computing Machines', *Journal of Chemical Physics*, **21**, 1087–1092.

Monahan, J. F. (1983), 'Fully Bayesian Analysis of ARMA Time Series Models', *Journal of Econometrics*, **21**, 307–331.

Nelson, C. R. and C. I. Plosser (1982), 'Trends and Random Walks in Macroeconomic Time Series', *Journal of Monetary Economics*, **10**, 139–162.

Phillips, P.C.B. (1989), 'Partially Identified Econometric Models', *Econometric Theory*, **5**, 181–240.

Sargan, J.D. and A. Bhargava (1983), 'Maximum Likelihood Estimation of Regression Models with First Order Moving Average Errors when the Root lies on the Unit Circle', *Econometrica*, **51**, 799–820.

Schotman, P. C. and H.K. Van Dijk (1993), Posterior Analysis of Possibly Integrated Time Series with an Application to Real GNP, *in* Caines, P., J. Geweke and M. Taqqu, eds, 'New Directions in Time Series Analysis, Part II', Springer Verlag (Berlin), pp. 341–361.

Schwarz, G. (1978), 'Estimating the Dimension of a Model', *Annals of Statistics*, **6**, 461–464.

Zellner, A. (1971), *An Introduction to Bayesian Inference in Econometrics*, Wiley (New York).