TI 2000-027/4
Tinbergen Institute Discussion Paper

Bayesian Analysis of ARMA Models

Frank R. Kleibergen
Henk Hoek

Tinbergen Institute

The Tinbergen Institute is the institute for economic research of the Erasmus Universiteit Rotterdam, Universiteit van Amsterdam and Vrije Universiteit Amsterdam.

Tinbergen Institute Amsterdam
Keizersgracht 482
1017 EG Amsterdam
The Netherlands
Tel.: +31.(0)20.5513500
Fax: +31.(0)20.5513555

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31.(0)10.4088900
Fax: +31.(0)10.4089031

Bayesian Analysis of ARMA models

Frank Kleibergen* Henk Hoek†

March 21, 2000

Abstract

Root cancellation in Auto Regressive Moving Average (ARMA) models leads to local non-identification of parameters. When we use diffuse or normal priors on the parameters of the ARMA model, posteriors in Bayesian analyzes show an a posteriori favor for this local non-identification. We show that the prior and posterior of the parameters of an ARMA model are the (unique) conditional density of a prior and posterior of the parameters of an encompassing AR model. We can therefore specify priors and posteriors on the parameters of the encompassing AR model and use the prior and posterior that it implies on the parameters of the ARMA model, and vice versa. The posteriors of the ARMA parameters that result from standard priors on the parameters of an encompassing AR model do not lead to an a posteriori favor of root cancellation. We develop simulators to generate parameters from these priors and posteriors. As a byproduct, Bayes factors can be computed to compare (non-nested) parsimonious ARMA models. The procedures are applied to the (extended) Nelson-Plosser data. For approximately 50% of the series 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, *e.g.,* Box *et. al.* (1994) and Harvey (1981), and are commonly used in applied work. They do, however, possess 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 and thus also not the likelihood. These parameters are then said to be locally non-identified. The problem of local non-identification is common to many models in statistics and econometrics, see, for example, the Simultaneous Equation Model which is discussed in, *e.g.,* Phillips (1989).

The Bayesian analysis of ARMA models documented in the literature, see *e.g.,* Monahan (1983), Marriot and Smith (1992), Chib and Greenberg (1994), Marriot *et. al.* (1995) and Zellner (1971), specifies standard priors, like, for example, a normal prior, on the parameters of the ARMA model. These analyzes thus do not explicitly account for the local non-identification. We conduct a different kind of analysis then the one pursued traditionally. We focus on the

*Department of Quantitative Economics, University of Amsterdam, Roetersstraat 11, 1018 WB Amsterdam, the Netherlands, Email:kleiberg@fee.uva.nl.

†Department of Financial Sector Management, Vrije Universiteit Amsterdam, De Boelelaan 1105, 1081 HV Amsterdam, the Netherlands, Email: hhoek@econ.vu.nl.

property of the, troublesome, ARMA model that it is nested within a, relatively easy, encompassing AR model. The restriction that, when imposed on the parameters of the encompassing AR model, leads to the nested ARMA model satisfies the sufficient conditions for a unique expression of the conditional density, see Kleibergen (2000). Hence, the prior and posterior of the parameters of the ARMA model are conditional densities of a prior and posterior of the parameters of the encompassing AR model. We can therefore specify a prior on the parameters of the encompassing AR model and the prior on the parameters of the nested ARMA model then results as a (unique) conditional density. In this respect our approach differs from the earlier Bayesian analyzes of ARMA models since these analyzes directly impose a prior on the parameters of the ARMA model. As identification in AR models is straightforward, the local non-identification is not problematic in our approach while the posteriors easily lead to a favor for local non-identification in the earlier analyzes.

The paper is organized as follows. In section 2, the consequences of the local non-identification of ARMA parameters in Bayesian analyzes using diffuse and normal priors is discussed. We show that these analyzes often result in posteriors that have a favor for local non-identification. In section 3, we show that the prior and posterior of the parameters of an ARMA model are the unique conditional densities of a prior and posterior of the parameters of an encompassing AR model given that the parameters, which unambiguously represent the difference between the encompassing AR and nested ARMA, are equal to zero. In section 4, we discuss Bayes factors for comparing ARMA models. We construct a Bayes factor for comparing ARMA models with equal summed AR and MA lag lengths where the priors on the parameters of the compared models result from the same prior on the parameters of the encompassing AR model. By letting the prior variance converge to infinity, we obtain a Bayes factor for this case that can be considered to incorporate no prior information. In section 5 we construct an Importance Sampler posterior simulator to compute prior and posterior moments and Bayes factors. Next to this simulator, we also construct a Metropolis-Hasting simulator. Section 6 contains an application to the extended Nelson-Plosser data. 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 7 summarizes and concludes.

2 Local Non-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 non-identification of redundant AR and MA parameters.

2.1 ARMA(1,1)

To show the local non-identification, consider the "simplest" ARMA model, the ARMA$(1,1)$ model,

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

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,

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

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

where $\theta = \rho - \alpha$, local non-identification can easily be recognized. In particular, depending on the specification used, ρ or α are non-identified 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 non-zero in the direction of ρ or α for zero values of θ. Use of a flat or diffuse prior in a Bayesian analysis of the ARMA$(1,1)$ model, such that the posterior is proportional to the likelihood, then results in a flat and non-zero 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 the values of the ARMA parameters at which the local non-identification problem occurs. This is neither a result of information from the prior or from the data but of a model property, *i.e.* the local non-identification.

In case of a proper normal prior, as used for example by Chib and Greenberg (1994) and Monahan (1983), the conditional posterior of ρ (or α) given $\theta = 0$ is proper but also proportional to the conditional prior of ρ (or α) given $\theta = 0$. So, at $\theta = 0$, the conditional posterior given θ is proportional to the conditional prior while at the other values of θ it also depends on the likelihood. The importance of the prior for the posterior thus depends on the value of the parameters. We would like to have priors that are such that the importance of the information in the prior for the information in the posterior in an evenly way depends on the value of the parameters. This is one of the motivations for this paper. In other models where local non-identification occurs, like cointegration and simultaneous equations models, the posteriors behave accordingly, see Kleibergen and van Dijk (1994b,1998), but priors with the desired features can be constructed and then lead to posteriors with convenient properties, see Kleibergen and Paap (1998) and Kleibergen and Zivot (1999).

2.1.1 Posterior ARMA(1,1) using Diffuse Prior

To illustrate the consequences of the local non-identification for the posterior of the parameters of the ARMA(1,1) model, we analyze the marginal posteriors of the ARMA parameters for an artificial time series. This series is generated from an ARMA$(1,1)$ model, see (1), with parameters $\rho = 0.6,\ \alpha = 0.4,\ \sigma^2 = 1,\ T = 200$. The identifying parameter $\theta = \rho - \alpha$ thus equals 0.2. We computed the posteriors of the parameters of an ARMA(1,1) model using a diffuse prior on (ρ, α), $p(\rho, \alpha) \propto \sigma^{-2}$. The posteriors are calculated using the analytical expression of the bivariate posterior of (ρ, α), which is proportional to the concentrated conditional likelihood and where we have set $y_{1-(p+q)-i} = \varepsilon_{1-i} = 0,\ i = 1, \ldots, \infty$ $(p = 1,\ q = 1)$.

Figure 1 shows the bivariate marginal posterior of the parameters (α, θ) of an ARMA(1,1) model for the artificially generated time series while figure 2 shows the contour-lines of this bivariate marginal posterior. Table 1 contains the posterior means and maximum likelihood estimates of the different parameters. Also the standard deviations are given. Since the bivariate posterior is proportional to the concentrated likelihood, the location of the posterior mode and the maximum likelihood estimate coincide which is confirmed by the contourlines

prior \ parameter	ρ	α	θ
diffuse on (ρ, α)	0.32 0.49	0.19 0.49	0.12 0.068
ML estimate	0.65 0.22	0.53 0.25	0.12 0.064

Table 1: Posterior moments and ML estimate ARMA(1,1) parameters artificial time-series

in figure 2. Because of the local non-identification, the existence of the posterior moments (and even of the posterior distribution) of α and ρ is doubtful which partly explains the large difference between the posterior means and the maximum likelihood estimate.

The bivariate posterior in figure 1 reveals the local non-identification of α when $\theta = 0$ as it is non-zero and constant in the direction of α at $\theta = 0$. The contour-lines further emphasize the local non-identification of α at $\theta = 0$. The marginal posterior of θ is obtained by integrating the bivariate posterior of (α, θ), shown in figure 1, over α. At $\theta = 0$, the bivariate posterior of (α, θ) is constant in the direction of α. Consequently, the marginal posterior of θ at $\theta = 0$ is proportional to the size of the parameter region of α as the integral of a constant function is proportional to the size of the parameter region. An infinite parameter region for α would therefore imply an infinite value of the marginal posterior of θ at $\theta = 0$. We have chosen a finite parameter region for α, $(-1.3, 1.3)$, such that the marginal posterior is finite at $\theta = 0$. Figure 3 contains the marginal posterior of θ and shows that it indeed has a secondary (local) mode at $\theta = 0$ which solely results from the local non-identification of α. The posterior has therefore more probability mass at $\theta = 0$ and thus has a favor for $\theta = 0$ that solely results from the local non-identification of α at $\theta = 0$.

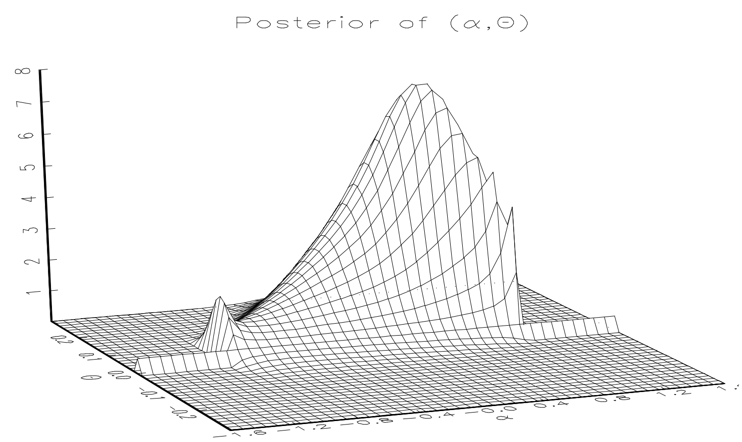

Figure 1: Bivariate posterior (α, θ), artificial time series, diffuse prior on (α, ρ)

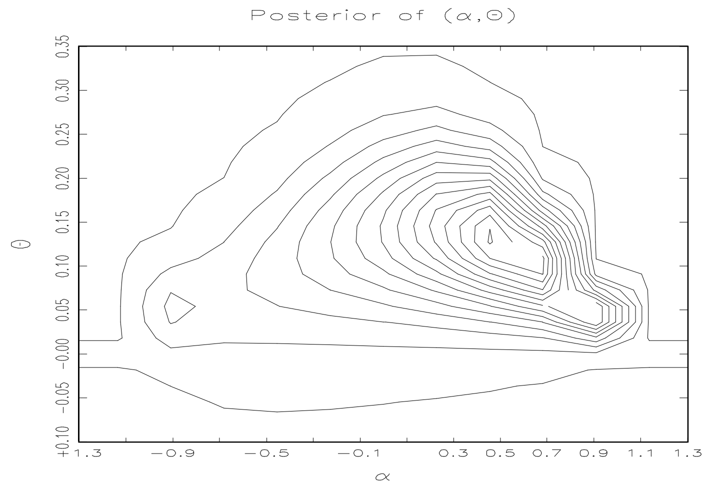

Figure 2: Contourlines bivariate posterior (α, θ), artificial time series, diffuse prior on (α, ρ)

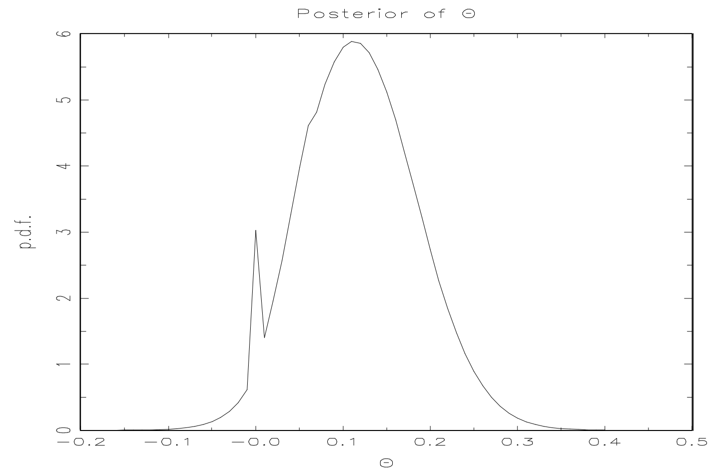

Figure 3: Marginal posterior θ, artificial time series, diffuse prior on (α, ρ).

2.1.2 Markov Chain Monte Carlo Posterior Simulators

As mentioned previously, ρ or α is locally non-identified 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 emphasize this point as it is important in the construction of Markov Chain Monte Carlo (MC^2) procedures for computing the marginal posteriors. For example, the MC^2 approach developed in Chib and Greenberg (1994) suffers from the local non-identification problem. In this algorithm, the conditional posteriors $p(\alpha|\rho, \dots)$ and $p(\rho|\alpha, \dots)$ 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 are (approximately) 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 $p(\alpha|\rho,\dots)$ and $p(\rho|\alpha,\dots)$ can lead to a reducible Markov Chain as the points of local non-identification, $\alpha=\rho$, 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 non-identification. Furthermore, the use of independent priors for the different parameters does not correspond with the strong dependence of the parameters within the likelihood. Figure 3 also demonstrates that estimation of the Bayes Factor in favor of the common factor restriction using the Savage-Dickey Density Ratio, see Dickey (1971) and Verdinelli and Wasserman (1995), is problematic. This ratio, which equals the ratio of the marginal posterior and prior of θ evaluated in $\theta=0$, depends on the height of the pike at $\theta=0$, which itself depends on the prior and not on the data. In particular, because the likelihood function is constant when $\theta=0$, using independent normal priors results in proportionality of the prior and posterior, see Kleibergen (2000).

2.2 ARMA(p,q)

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

$$\begin{aligned}\rho(L)y_t &= \alpha(L)\varepsilon_t \Leftrightarrow \\ (1-\rho_1L-\dots-\rho_pL^p)y_t &= (1-\alpha_1-\dots-\alpha_qL^q)\varepsilon_t,\end{aligned} \tag{4}$$

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

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

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

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

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

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

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_i's, with $i\leq p+q$ only, such that inference on the $p+q$ parameters $\rho_1,\dots,\rho_p,\alpha_1,\dots,\alpha_q$ can be based on $c_1,\dots,c_{p+q}$ solely. The relation between these parameters is given by the following matrix equation, which follows from the set of equations in (7) and (8),

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

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

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

with I_p the identity matrix of dimension p,

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

and

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

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

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

If C_{22} does not have a full rank value α, and consequently ρ, can not be determined uniquely. This is a generalization of the local non-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 Golub and van Loan (1989) (see also Gill and Lewbel (1992), Cragg and Donald (1996) and Kleibergen and van Dijk (1994a,b), for other applications of the LU decomposition in econometric and time series models),

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

The rank of C_{22} is now given by the number of non-zero 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 non-identification occurs. As the autocorrelation of a specific order is a function of the parameters of the ARMA model, see, *e.g.*, Box *et. al.* (1994) and Harvey (1981), also an *a posteriori* favor for specific values of the autocorrelations results which is solely a consequence of the local non-identification. When using diffuse or normal priors, the posteriors of the parameters of AR models lead to posteriors of the autocorrelations that do not contain such an *a posteriori* favor as AR models do not contain parameters that are locally non-identified. So, when we use diffuse or normal priors, the posteriors of the autocorrelations of AR and ARMA models are quite different and strongly depend on the model where they result from. This is quite peculiar as the ARMA model can be considered as an AR(∞) model. In the next section we therefore explicitly analyze how an ARMA model is obtained from an AR(∞) model and show what this implies for the prior and posterior of its parameters.

Finally, note that the autocorrelations of non-invertible MA models, *i.e.* models with one or more roots of the MA polynomial which lie within the unit circle, can not 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 non-invertible 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 *e.g.* Kadane (1993), we allow for non-invertible MA parameters such that, in principle, the MA and AR parameters range from $-\infty$ to ∞.

3 AR(MA) as restriction on AR(∞)

The implied AR(∞) specification of an ARMA model, (2) and (5)-(8), shows that the ARMA model can be considered as an AR(∞) model with restrictions on its parameters. Also a finite order AR model can be considered as an AR(∞) model with restrictions on its parameters. We could consider both the prior and posterior of the parameters of the ARMA and AR model to be proportional to the conditional prior and posterior of the parameters of the AR(∞) model given that the parameters of the AR(∞) model satisfy the restrictions of either the ARMA or AR model. In order for such an approach to be feasible, unique expressions for the priors and posteriors of the parameters of the ARMA and AR model should result. In Kleibergen (2000) sufficient conditions for the existence of such unique conditional densities are given. We briefly discuss these kind of conditional densities and apply them to construct priors and posteriors of the parameters of finite order AR and ARMA models. We show that the priors that are typically used for the parameters of AR models result as a conditional density of standard priors on the parameters of the AR(∞). The priors that are typically used for the parameters of ARMA models, however, result as a conditional density of priors on the parameters of the AR(∞) model that favor root cancellation. This explains the *a posteriori* favor for root cancellation/local non-identification that results from the use of standard priors on the parameters of the ARMA model. The priors on the parameters of the ARMA model that result as a conditional density of a standard prior on the parameters of the AR(∞) model lead to posteriors that behave in a convenient way and, for example, lead to posteriors of the autocorrelations that are similar to those that result from AR models.

3.1 Unique Conditional Densities

In Kleibergen (2000), it is shown that when we can uniquely determine the random variable on which we want to impose a restriction, the (conditional) density of the restricted random variable is unique and can be obtained by using two sufficient conditions for a unique expression of the conditional density, which are stated below. The conditional density that results from these sufficient conditions is the only density that just conditions on the desired restriction and it is invariant with respect to the specification of the restriction. Hence, these conditional densities avoid the Borel-Kolmogorov paradox, see *e.g.*, Kolmogorov (1950) and Billingsley (1986). The sufficient conditions read.

Sufficient conditions for the existence of a unique conditional density for the continuous random variable $\varphi : k \times 1$; whose space, on which it is defined, is unrestricted, i.e. the $\Re^k$, and has a continuous and continuous differentiable pdf $p(\varphi)$ which is such that φ is identified everywhere; that only conditions on the restriction $\varphi = g(\delta)$ and nothing else; where $g(\delta)$: $k \times 1$, $\delta : m \times 1$, $m < k$, and $g(\delta)$ is continuous differentiable and is defined on the whole space of δ, i.e. the $\Re^m$; are:

Condition 1. *An invertible relationship between φ and (δ, λ) exists; where $\lambda : (k - m) \times 1$ and $\varphi = f(\delta, \lambda)$ is continuous differentiable; which is such that the set of values of δ for which $\varphi = g(\delta)$ is uniquely defined is equivalent to the set of values of δ for which $\varphi = f(\delta, \lambda)$ is uniquely defined and the latter set does not depend on the value of λ.*

Condition 2. *The restriction $\varphi = g(\delta)$ is equivalent with $(\delta, \lambda) = (\delta, 0)$ and imposes no restrictions on δ.*

The unique expression of the conditional density of φ given that $\varphi = g(\delta)$ is then characterized by the density of δ, see Kleibergen (2000),

$$\begin{aligned} p_r(\delta) &\propto p(\delta, \lambda)|_{\lambda=0} \\ &\propto p(\varphi(\delta, \lambda))|_{\lambda=0} |J(\varphi, (\delta, \lambda))|_{\lambda=0}|, \end{aligned} \tag{16}$$

where r stands for restricted, $|_{\lambda=0}$ stands for evaluated in $\lambda = 0$ and $J(\varphi, (\delta, \lambda))$ is the jacobian of the transformation from φ to (δ, λ). The conditional density (16) is invariant with respect to the specification of (δ, λ) when (δ, λ) satisfies the sufficient conditions. For more details on the unique conditional densities we refer to Kleibergen (2000).

A nested model can be considered as a restriction on the parameters of an encompassing model. Since the parameters are, realizations of, random variables in Bayesian analysis, the prior and posterior of the parameters of a nested model are therefore unique conditional densities of a prior and posterior of the parameters of the encompassing model. We can use these unique conditional densities to construct the priors and posteriors of the parameters of low order AR and ARMA models from priors and posteriors of the parameters of an encompassing high order AR model.

3.2 Posterior High Order AR

We specify the prior and posterior of the parameters of a high order AR model and use it to construct conditional densities given specific restrictions that result in the prior and posterior of the parameters of lower order AR and ARMA models. The high order AR model reads,

$$y_t = \sum_{i=1}^{p_{\max}} c_i y_{t-i} + \varepsilon_t, \tag{17}$$

where ε_t i.i.d. $N(0, \sigma^2)$, $t = 1, \ldots, T$ and $p_{\max}$ is reasonably large but less than T. The (conditional) likelihood of (17) reads,

$$\begin{aligned} l_{\mathrm{AR}(p_{\max})}(\varphi, \sigma^2 | y) &\propto \left|\sigma^2\right|^{-\frac{1}{2}T} \exp\left[-\frac{1}{2\sigma^2}\left(y - Y_{-1}\varphi\right)'\left(y - Y_{-1}\varphi\right)\right] \\ &\propto \left|\sigma^2\right|^{-\frac{1}{2}T} \exp\left[-\frac{1}{2\sigma^2}\left(y' M_{Y_{-1}} y + (\varphi - \hat{\varphi})' Y_{-1}' Y_{-1}\left(\varphi - \hat{\varphi}\right)\right)\right], \end{aligned} \tag{18}$$

where $\varphi = (c_1 \ldots c_{p_{\max}})'$; $\hat{\varphi} = (Y_{-1}' Y_{-1})^{-1} Y_{-1}' y$, $M_{Y_{-1}} = I_T - Y_{-1}(Y_{-1}' Y_{-1})^{-1} Y_{-1}'$, $y = (y_1 \cdots y_T)'$, $Y_{-1} = (x_{-1} \cdots x_{-p_{\max}})$; $x_{-i} = (y_{1-i} \cdots y_{T-i})'$, $i = 1, \ldots, p_{\max}$. For illustrative purposes, we specify a conditional normal prior on φ given σ^2 but essentially any other continuous (differentiable) prior can be used as well. Similarly for σ^2, we specify a diffuse prior, $p(\sigma^2) \propto \left|\sigma^2\right|^{-1}$, such that

$$\begin{aligned} p_{\mathrm{AR}(p_{\max})}(\varphi, \sigma^2) &\propto p_{\mathrm{AR}(p_{\max})}(\varphi | \sigma^2) p_{\mathrm{AR}(p_{\max})}(\sigma^2) \\ &\propto \left|\sigma^2\right|^{-\frac{1}{2}(p_{\max}+2)} \exp\left[-\frac{1}{2\sigma^2}\left(\varphi - \varphi_0\right)' A_0 \left(\varphi - \varphi_0\right)\right], \end{aligned} \tag{19}$$

where $\varphi_0 : p_{\max} \times 1$, is the prior mean and $\sigma^2 A_0^{-1}$, $A_0 : p_{\max} \times p_{\max}$, is the prior covariance matrix, and we obtain the posterior of (φ, σ^2),

$$\begin{aligned} p_{\mathrm{AR}(p_{\max})}(\varphi, \sigma^2 | y) &\propto p_{\mathrm{AR}(p_{\max})}(\varphi, \sigma^2) l_{\mathrm{AR}(p_{\max})}(\varphi, \sigma^2 | y) \\ &\propto \left|\sigma^2\right|^{-\frac{1}{2}(T + p_{\max}+2)} \exp\left[-\frac{1}{2\sigma^2}\left(\tilde{\sigma}^2 + (\varphi - \tilde{\varphi})' V \left(\varphi - \tilde{\varphi}\right)\right)\right] \\ &\propto \left|\sigma^2\right|^{-\frac{1}{2}(T + p_{\max}+2)} \exp\left[-\frac{1}{2\sigma^2}\left((\varphi - \varphi_0)' A_0 \left(\varphi - \varphi_0\right) + (y - Y_{-1}\varphi)'(y - Y_{-1}\varphi)\right)\right], \end{aligned} \tag{20}$$

where $V = A_0 + Y_{-1}' Y_{-1}$, $\tilde{\varphi} = V^{-1}(A_0 \varphi_0 + Y_{-1}' Y_{-1} \hat{\varphi})$, $\tilde{\sigma}^2 = y'y + \varphi_0' A_0 \varphi_0 - (A_0 \varphi_0 + Y_{-1}' Y_{-1} \hat{\varphi})' V^{-1} (A_0 \varphi_0 + Y_{-1}' Y_{-1} \hat{\varphi})$. We use the prior (19) and posterior (20) to construct conditional densities given specific restrictions.

3.2.1 Posterior Low Order AR=Conditional Posterior given Linear Restriction

The prior and posterior of a lower order AR model are proportional to a conditional prior and posterior of the parameters of the high order AR given that the parameters of the high order AR equal the parameters of the low order AR. We show this by using the AR(p) model,

$$y_t = \sum_{i=1}^{p} \rho_i y_{t-i} + \varepsilon_t, \tag{21}$$

where ε_t i.i.d. $N(0, \sigma^2)$, $t = 1, \ldots, T$, $\rho = (\rho_1 \cdots \rho_p)'$ and we assume that $p < p_{\max}$.

We can specify the parameter φ in (18)-(20) as $\varphi = (\delta' \;\; \lambda')'$, where $\delta : p \times 1$ and $\lambda : (p_{\max} - p) \times 1$. The AR($p$) model (21) is identical to the AR($p_{\max}$) model evaluated at $\varphi = (\delta'$

$0\ldots0)'$, with $\delta=\rho$. The restriction that is imposed by the AR(p) model on the parameters of the AR($p_{\max}$) model is therefore $\varphi=(\delta'\ 0\ldots0)'$. We analyze whether the specification (δ,λ) satisfies the sufficient conditions for a unique density of φ given that $\varphi=(\delta'\ 0\ldots0)'$. The restriction $\varphi=(\delta'\ 0\ldots0)'$ is such that φ is uniquely determined by δ for all values of δ. The specification $\varphi=(\delta'\ \lambda')'$ is such that φ is uniquely determined by (δ,λ) for all values of (δ,λ). Hence the first sufficient condition is satisfied. The second sufficient condition is also satisfied as $\varphi=(\delta'\ 0\ldots0)'\Leftrightarrow(\delta'\ \lambda')=(\delta'\ 0\ldots0)$. The specification (δ,λ) thus satisfies the sufficient conditions for the existence of a unique conditional density. The prior and posterior of the parameters of (21) are then a conditional prior and posterior of (δ,σ^2) in (17) given that $\lambda=0$. To further illustrate this, we construct the resulting conditional prior and posterior.

As the joint density of (δ,λ) is normal, the conditional prior of (δ,σ^2) given λ results from the well-known result that the conditional densities of normal distributed random variables are also normal. This conditional prior then reads

$$\begin{aligned}p_{\mathrm{AR}(p_{\max})}(\delta,\sigma^2|\lambda)\propto\left|\sigma^2\right|^{-\frac{1}{2}(p_{\max}+2)}\exp\Big[-\frac{1}{2\sigma^2}\Big(\left((\delta-\delta_0)-A_{0,11}^{-1}A_{0,12}(\lambda-\lambda_0)\right)'\\ A_{0,11}\left((\delta-\delta_0)-A_{0,11}^{-1}A_{0,12}(\lambda-\lambda_0)\right)+(\lambda-\lambda_0)'A_{0,22.1}(\lambda-\lambda_0)\Big)\Big],\end{aligned}\tag{22}$$

where $\varphi_0=(\delta_0'\ \lambda_0')'$, $\delta_0:p\times1$, $\lambda_0:(p_{\max}-p)\times1$, and $A_0=\begin{pmatrix}A_{0,11}&A_{0,12}\\A_{0,21}&A_{0,22}\end{pmatrix}$, $A_{0,11}:p\times p$, $A_{0,21}'$, $A_{0,12}:p\times(p_{\max}-p)$, $A_{0,22}:(p_{\max}-p)\times(p_{\max}-p)$, $A_{0,22.1}=A_{0,22}-A_{0,21}A_{0,11}^{-1}A_{0,12}$. When we only analyze the nested AR(p) model (21), the prior mean of λ equals zero, $\lambda_0=0$. The conditional prior of (δ,σ^2) given $\lambda=0$,

$$\begin{aligned}p_{\mathrm{AR}(p_{\max})-r}(\delta,\sigma^2)&\propto p_{\mathrm{AR}(p_{\max})}(\delta,\sigma^2|\lambda=0)\\ &\propto\left|\sigma^2\right|^{-\frac{1}{2}(p+(p_{\max}-p)+2)}\exp\left[-\frac{1}{2\sigma^2}(\delta-\delta_0)'A_{0,11}(\delta-\delta_0)\right],\end{aligned}\tag{23}$$

where the subscript $-r$ denotes “restricted”, then exactly corresponds with a normal-diffuse prior on the parameters of (21),

$$\begin{aligned}p_{\mathrm{AR}(p)}(\rho,\sigma^2)&=p_{\mathrm{AR}(p_{\max})-r}(\delta(\rho),\sigma^2)\\ &\propto p_{\mathrm{AR}(p_{\max})}(\delta(\rho),\lambda,\sigma^2)|_{\lambda=0},\end{aligned}\tag{24}$$

and which shows that the prior for (ρ,σ^2) in (21) is a conditional prior for (δ,σ^2) given $\lambda=0$ in (17).

In a similar way, we can construct the conditional posterior of (δ,σ^2) given λ which reads

$$\begin{aligned}p_{\mathrm{AR}(p_{\max})}(\delta,\sigma^2|\lambda,y)\propto\left|\sigma^2\right|^{-\frac{1}{2}(T+p_{\max}+2)}\exp\Big[-\frac{1}{2\sigma^2}\Big(\tilde{\sigma}^2+\left(\lambda-\tilde{\lambda}\right)'V_{22.1}\left(\lambda-\tilde{\lambda}\right)\\ +\left(\left(\delta-\tilde{\delta}\right)-V_{11}^{-1}V_{12}\left(\lambda-\tilde{\lambda}\right)\right)'V_{11}\left(\left(\delta-\tilde{\delta}\right)-V_{11}^{-1}V_{12}\left(\lambda-\tilde{\lambda}\right)\right)\Big)\Big]\\ \propto\left|\sigma^2\right|^{-\frac{1}{2}(T+p_{\max}+2)}\exp\left[-\frac{1}{2\sigma^2}\left((\varphi-\varphi_0)'A_0(\varphi-\varphi_0)+(y-Y_{-1}\varphi)'(y-Y_{-1}\varphi)\right)\right],\end{aligned}\tag{25}$$

where $\tilde{\varphi}=(\tilde{\delta}'\ \tilde{\lambda}')'$, $\tilde{\delta}:p\times1$, $\tilde{\lambda}:(p_{\max}-p)\times1$, $V=\begin{pmatrix}V_{11}&V_{12}\\V_{21}&V_{22}\end{pmatrix}$, $V_{11}:p\times p$, V_{21}', $V_{12}:p\times(p_{\max}-p)$, $V_{22}:(p_{\max}-p)\times(p_{\max}-p)$, $V_{22.1}=V_{22}-V_{21}V_{11}^{-1}V_{12}$. When we analyze the

AR(p) (21), the prior mean of λ, λ_0, is again equal to zero such that the posterior of (ρ, σ^2) using the prior (24) equals the conditional posterior of (δ, σ^2) given $\lambda = 0$ when using the prior (19),

$$\begin{aligned} p_{\mathrm{AR}(p)}(\rho, \sigma^2|y) &= p_{\mathrm{AR}(p_{\max})-r}(\delta(\rho), \sigma^2|y) \\ &= p_{\mathrm{AR}(p_{\max})}(\delta(\rho), \sigma^2|\lambda = 0, y) \\ &\propto p_{\mathrm{AR}(p_{\max})}(\delta(\rho), \lambda, \sigma^2|y)|_{\lambda=0}. \end{aligned} \tag{26}$$

This shows that, as the likelihood is continuous in the parameters, the relationship (24) for the prior extends to the posterior.

It is common practice to specify priors directly on the parameters of the analyzed model and not to derive them as the conditional prior of the parameters of an encompassing model. The above results show that it does not matter for the prior and posterior of the parameters of the AR(p) (21) whether we directly impose a normal prior on its parameters or construct the prior as the conditional prior that results from a normal prior on the parameters of an encompassing AR model. This results as the restriction imposed by the AR(p) (21) on the AR($p_{\max}$) (17) is linear and linear combinations of normal random variables and the conditional densities of normal random variables are normal. This does, however, not extend to non-linear restrictions that satisfy sufficient conditions 1 and 2. For the nested models that result from these kind of restrictions on the parameters of an encompassing linear model, normal priors that are specified directly on the parameters of the nested model do not coincide with the conditional prior that results from a normal prior on the parameters of an encompassing linear model. An example of this is the ARMA(1,1) model which we observe next.

3.2.2 Posterior ARMA(1,1)=Conditional Posterior given Exponential Restriction

When the order $p_{\max}$ of the AR($p_{\max}$) model (17) is large, the ARMA(1,1) model (1) can be considered to result from a set of exponential restrictions on the parameters of the AR($p_{\max}$) model (17). These restrictions can be specified as, see Kleibergen (2000),

$$c_i = \alpha^{i-1}\theta, \qquad i = 1, \ldots, p_{\max}, \tag{27}$$

where $\theta = \rho - \alpha$. We introduce a set of additional parameters $\lambda = (\lambda_1 \cdots \lambda_{p_{\max}-2})' : (p_{\max} - 2) \times 1$, to span these restrictions,

$$\lambda_i = c_{i+2} - \alpha^{i+1}c_2, \qquad i = 1, \ldots, p_{\max} - 2, \tag{28}$$

such that φ can be specified as

$$\varphi = \begin{pmatrix} c_1 \\ c_2 \\ c_3 \\ \vdots \\ c_{p_{\max}} \end{pmatrix} = \begin{pmatrix} \theta \\ \alpha\theta \\ \lambda_1 + \alpha^2\theta \\ \vdots \\ \lambda_{p_{\max}-2} + \alpha^{p_{\max}-1}\theta \end{pmatrix}, \tag{29}$$

The specification of the restrictions, imposed by the ARMA(1,1) model on the AR($p_{\max}$), (28)-(29) satisfies the sufficient conditions for the existence of a unique conditional density of φ given (27). Condition 1 is satisfied as we can uniquely solve for φ from $(\alpha, \theta, \lambda)$ when $\theta \neq 0$ for all values of λ. Similarly, (27) also shows that the value of c_i is uniquely determined by (α, θ) when $\theta \neq 0$. Because the restriction is not imposed on θ, condition 1 is thus satisfied. Condition 2 is satisfied because (29) is equivalent with (27) when $\lambda = 0$ in (29). To further verify these conditions we check the invariance property that the conditional density is invariant to the specification of the parameters that result from the sufficient conditions. To analyze this invariance property, it is important to note that although α is present in all the elements of φ in (29), it can only be obtained as the ratio of $\frac{c_{i+1}}{c_i}$, $i = 1, \ldots, p_{\max} - 1$, since square roots of negative values are not properly defined. Another specification of φ in terms of parameters representing restriction (27) is then, for example,

$$\varphi = \begin{pmatrix} c_1 \\ \vdots \\ c_{i-1} \\ c_i \\ c_{i+1} \\ c_{i+2} \\ \vdots \\ c_{p_{\max}} \end{pmatrix} = \begin{pmatrix} \mu_1 + \psi^{-i}\gamma \\ \vdots \\ \mu_{i-1} + \psi^{-1}\gamma \\ \gamma \\ \psi\gamma \\ \mu_i + \psi^2\gamma \\ \vdots \\ \mu_{p_{\max}-2} + \psi^{p_{\max}-i-1}\gamma \end{pmatrix}, \tag{30}$$

where $1 < i < p_{\max}$, γ, $\psi : 1 \times 1$, $\mu = (\mu_1 \cdots \mu_{p_{\max}-2})' : (p_{\max} - 2) \times 1$, and when $\mu = 0$, (30) is equivalent to (27). The jacobian of the transformation from φ to $(\theta, \alpha, \lambda)$, $|J(\varphi, (\theta, \alpha, \lambda))|$, is equal to $|\theta|$. The jacobian of the transformation from φ to (γ, ψ, μ), $|J(\varphi, (\gamma, \psi, \mu))|$, is equal to $|\gamma|$. Under the restriction (27), $\lambda = \mu = 0$, such that $\theta = \psi^{-i}\gamma$ and $\alpha = \psi$. The jacobian of the transformation from (γ, ψ) to (θ, α) is then, under the restriction, equal to $|J((\gamma, \psi), (\theta, \alpha))| = |\psi^{-i}| = |\alpha^{-i}|$. Combining these jacobians we obtain that

$$|J(\varphi, (\theta, \alpha, \lambda))|_{\lambda=0}| = |J(\varphi, (\gamma, \psi, \mu))|_{\mu=0}||J((\gamma, \psi), (\theta, \alpha))| = |\gamma||\psi^{-i}| = |\alpha^i\theta||\alpha^{-i}| = |\theta|, \tag{31}$$

which is the property that is needed to have a conditional density that is invariant with respect to the specification of the parameters that result from the sufficient conditions.

As the restrictions (27)-(29), that lead from the AR($p_{\max}$) model to the ARMA(1,1) model, satisfy the sufficient conditions for the existence of a unique conditional density, the prior and posterior of the parameters of the ARMA(1,1) model are a conditional prior and posterior of the parameters of the AR($p_{\max}$) model given that these restrictions hold. Hence, we can specify a normal prior on the parameters of the AR($p_{\max}$) like (19), which is a natural conjugate prior for that model as the prior has the same specification as the likelihood, see Poirier (1995), and construct the (conditional) prior that it implies on the parameters of the ARMA(1,1) model. To illustrate this, we construct such a prior and the resulting posterior.

Normal Prior on AR($p_{\max}$) parameters The prior on $(\alpha, \theta, \lambda, \sigma^2)$ implied by (19) reads,

$$\begin{aligned} p_{\mathrm{AR}(p_{\max})}(\alpha, \theta, \lambda, \sigma^2) &= p_{\mathrm{AR}(p_{\max})}(\varphi(\alpha, \theta, \lambda), \sigma^2)|J(\varphi, (\alpha, \theta, \lambda))| \qquad (32) \\ &\propto |\theta| \left|\sigma^2\right|^{-\frac{1}{2}(p_{\max}+2)} \exp\left[-\frac{1}{2\sigma^2} \left(\varphi(\alpha, \theta, \lambda) - \varphi_0\right)' A_0 \left(\varphi(\alpha, \theta, \lambda) - \varphi_0\right)\right], \end{aligned}$$

where $\varphi(\alpha, \theta, \lambda)$ results from (29). The proper prior for the parameters of the ARMA(1,1) model that results from (19) is equal to the conditional prior that results from (32) given that

$\lambda = 0$,

$$\begin{aligned} p_{\text{ARMA}(1,1)}(\alpha,\theta,\sigma^2) &\propto p_{\text{AR}(p_{\max})}(\alpha,\theta,\lambda,\sigma^2)|_{\lambda=0} \\ &\propto p_{\text{AR}(p_{\max})}(\varphi(\alpha,\theta,\lambda),\sigma^2)|_{\lambda=0}|J(\varphi,(\alpha,\theta,\lambda))|_{\lambda=0}| \\ &\propto |\theta|\left|\sigma^2\right|^{-\frac{1}{2}(p_{\max}+2)}\exp\left[-\frac{1}{2\sigma^2}\left(\varphi(\alpha,\theta,\lambda)|_{\lambda=0}-\varphi_0\right)' A_0\left(\varphi(\alpha,\theta,\lambda)|_{\lambda=0}-\varphi_0\right)\right], \end{aligned} \tag{33}$$

which is a proper prior with a normalizing constant equal to $\iiint p_{\text{AR}(p_{\max})}(\alpha,\theta,\lambda,\sigma^2)|_{\lambda=0}d\alpha d\theta d\sigma^2$. The prior (33) results from a natural conjugate prior that is specified on the parameters of the AR($p_{\max}$) model. It therefore shows the specification of a natural conjugate prior on the parameters of the ARMA(1,1) model. When we use the prior (33) for the parameters of the ARMA(1,1) model, the posterior is proportional to the product of the prior and the likelihood,

$$\begin{aligned} p_{\text{ARMA}(1,1)}(\alpha,\theta,\sigma^2|y) &\propto p_{\text{ARMA}(1,1)}(\alpha,\theta,\sigma^2)l_{\text{ARMA}(1,1)}(\alpha,\theta,\sigma^2|y) \\ &\propto p_{\text{AR}(p_{\max})}(\alpha,\theta,\lambda,\sigma^2)|_{\lambda=0}l_{\text{AR}(p_{\max})}(\varphi(\alpha,\theta,\lambda),\sigma^2|y)|_{\lambda=0} \\ &\propto p_{\text{AR}(p_{\max})}(\varphi(\alpha,\theta,\lambda),\sigma^2)|_{\lambda=0}|J(\varphi,(\alpha,\theta,\lambda))|_{\lambda=0}|l_{\text{AR}(p_{\max})}(\varphi(\alpha,\theta,\lambda),\sigma^2|y)|_{\lambda=0} \\ &\propto p_{\text{AR}(p_{\max})}(\varphi(\alpha,\theta,\lambda),\sigma^2|y)|_{\lambda=0}|J(\varphi,(\alpha,\theta,\lambda))|_{\lambda=0} \\ &\propto p_{\text{AR}(p_{\max})}(\alpha,\theta,\lambda,\sigma^2|y)|_{\lambda=0}, \end{aligned} \tag{34}$$

and this posterior of the ARMA(1,1) parameters also equals the conditional posterior of the parameters of the AR($p_{\max}$) using the prior (19) given that the ARMA(1,1) restriction holds.

While $p_{\max}$ should be less than T for the analysis of the AR($p_{\max}$) model, there is essentially no restriction on $p_{\max}$ for the analysis of the ARMA(1,1) model such that we set $p_{\max}$ equal to the number of observations, T, in which case the likelihood of the ARMA(1,1) model is equal to the conditional likelihood given the first $p+q$ $(=2)$ observations. Hence, we set $p_{\max}=T$ in the sequel of the paper.

Diffuse Prior on AR($p_{\max}$) parameters Instead of a normal prior on the parameters of the AR($p_{\max}$) model any other kind of continuous (differentiable) prior can be specified as well. So, we can also specify an improper diffuse prior on the parameters of the AR($p_{\max}$) model,

$$p_{\text{AR}(p_{\max})}(\varphi,\sigma^2) \propto \left|\sigma^2\right|^{-1}, \tag{35}$$

which specification is considered to be a non-informative one in linear models, see Berger (1985), and construct the implied improper prior on the parameters of the ARMA(1,1) using (33),

$$\begin{aligned} p_{\text{ARMA}(1,1)}(\alpha,\theta,\sigma^2) &\propto p_{\text{AR}(p_{\max})}(\varphi(\alpha,\theta,\lambda),\sigma^2)|_{\lambda=0}|J(\varphi,(\alpha,\theta,\lambda))|_{\lambda=0}, \\ &\propto \left|\sigma^2\right|^{-1}|\theta|. \end{aligned} \tag{36}$$

The specification of this prior is identical to the prior that is advocated by Box *et. al.* (1994, p.274-275) as a Jeffreys' prior for the parameters of an ARMA(1,1) model. The important difference with the diffuse prior that was used to construct the posterior in section 2.1.2 is the presence of $|\theta|$ in the prior. This factor results from the conditional identification of α on θ and accounts for the local non-identification of α at $\theta=0$. Note that this factor offsets the asymptote in the marginal posterior of θ at $\theta=0$, depicted in Figure 3. The posterior using

prior \ parameter	ρ	α	θ
diffuse on (ρ, α)	$\underset{0.49}{0.32}$	$\underset{0.49}{0.19}$	$\underset{0.068}{0.12}$
diffuse on φ	$\underset{0.37}{0.38}$	$\underset{0.36}{0.22}$	$\underset{0.062}{0.16}$
ML estimate	$\underset{0.22}{0.65}$	$\underset{0.25}{0.53}$	$\underset{0.064}{0.12}$

Table 2: Posterior moments and ML estimate ARMA(1,1) parameters artificial time-series

the prior (36) results directly from (34) and is a conditional density of the posterior using a diffuse prior of the AR($p_{\max}$) parameters.

Diffuse priors in linear models lead to posteriors that primarily reveal the information in the data. This then also holds for the posterior of the ARMA(1,1) parameters that results from prior (36). The resulting posteriors of, for example, the autocorrelations and the parameters are therefore more in line with the posteriors in AR models than the posteriors that result from diffuse or normal priors on the parameters of the ARMA(1,1) model. To illustrate this, we computed the marginal posteriors of the parameters of the ARMA(1,1) model using the prior (36) for the artificial data-set analyzed in section 2.1.2. For comparison, we also show some of the posteriors and posterior means that result when we use a diffuse prior on the ARMA(1,1) parameters as in section 2.1.2. The posteriors are computed using the Importance Sampler posterior simulator that is constructed in section 5.1.

Figures 4 and 5 show the bivariate posterior of (α, θ) and its contour-lines when we use prior (36). These figures show that the local non-identification of α at $\theta = 0$ no longer leads to a ridge in the bivariate posterior of (α, θ) at $\theta = 0$. Figures 6-8 contain the marginal posteriors of θ, α and ρ and, for comparison, these figures also show the posterior in case of a diffuse prior on the ARMA(1,1) parameters. Table 2 shows the maximum likelihood estimates and the posterior moments both in case of the prior (36) and in case of a diffuse prior on the ARMA(1,1) parameters. Note that the maximum likelihood estimate no longer coincides with the posterior mode when we specify a diffuse prior on φ, which can also be concluded from the contour-lines in figure 5, and that the existence of the posterior means of ρ and α is doubtful for both specifications of the prior. The figures show that the *a posteriori* favor for $\theta = 0$ compared to the diffuse prior on the ARMA(1,1) parameters has disappeared. The results using prior (36) are therefore more in line with what we expect when we use a diffuse (non-informative) prior as it should lead to posteriors that only show the information in the data and nothing else. A diffuse prior directly specified on the ARMA(1,1) parameters is clearly informative, given its *a posteriori* favor for $\theta = 0$, and thus not only shows the information in the data.

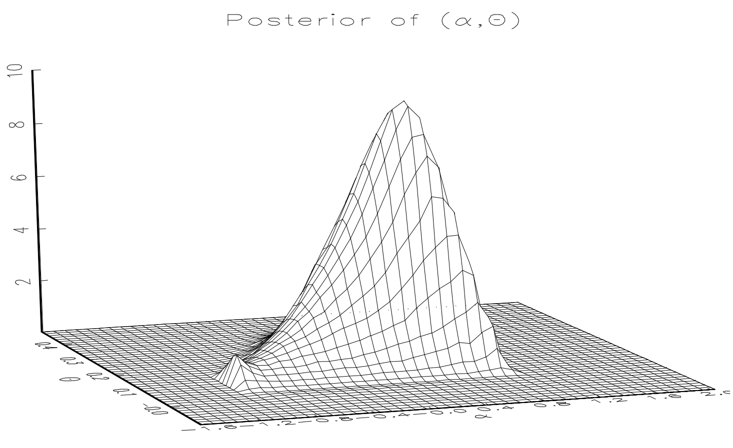

Figure 4: Bivariate posterior (α, θ), artificial time series, diffuse prior on φ.

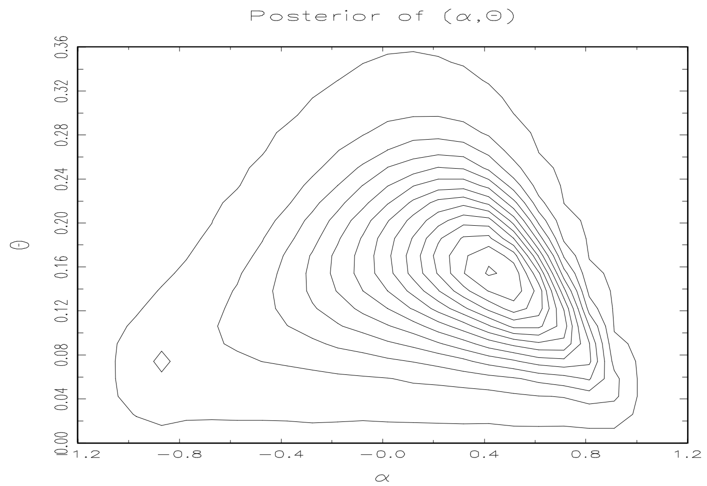

Figure 5: Contourlines bivariate posterior (α, θ), artificial time series, diffuse prior on φ.

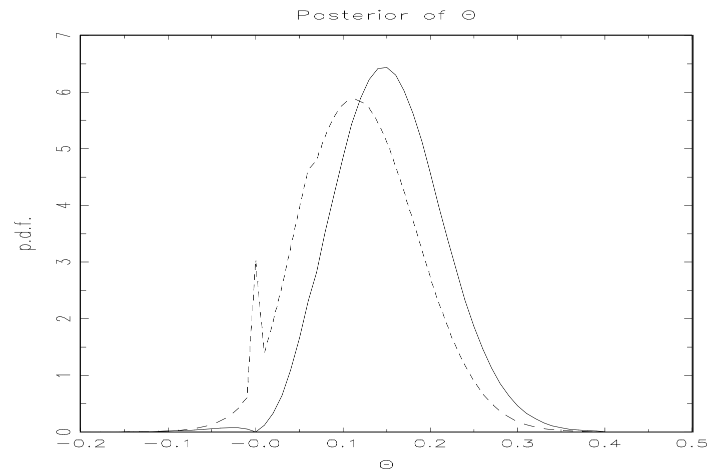

Figure 6: Marginal posterior θ, artificial time series, diffuse prior on (α, ρ) (- -), on φ (—).

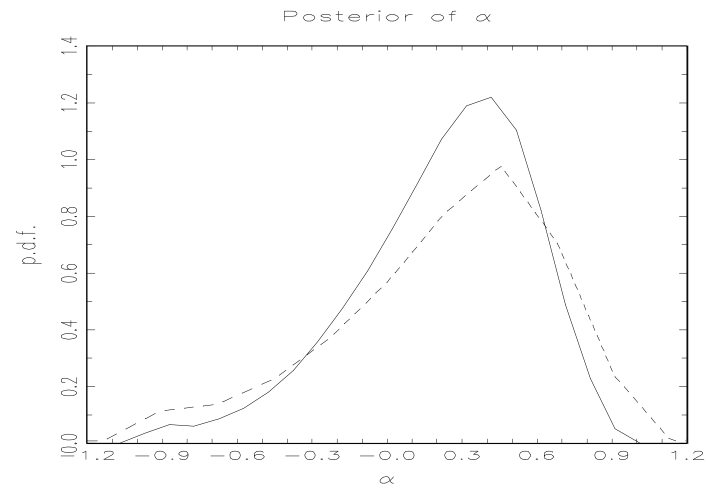

Figure 7: Marginal posterior α, artificial time series, diffuse prior on (α, ρ) (- -), on φ (—).

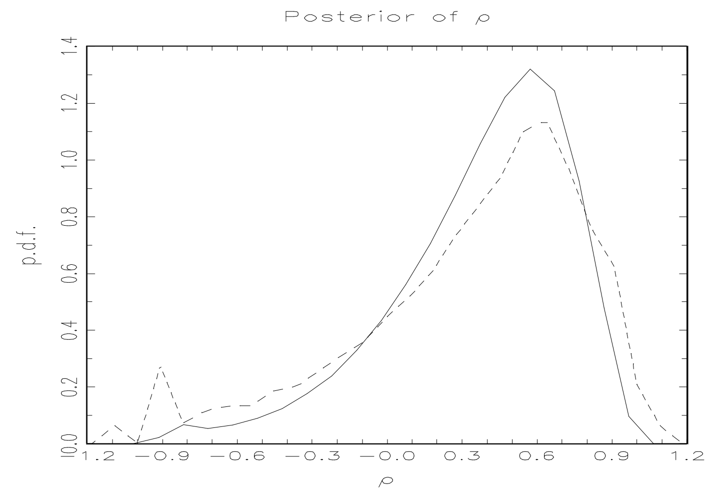

Figure 8: Marginal posterior ρ, artificial time series, diffuse prior on (α, ρ) (- -), on φ (—).

Implied Prior on parameters AR($p_{\max}$) Previously, we constructed the prior for the parameters of the ARMA(1,1) model from a prior that is specified on the parameters of the AR($p_{\max}$) model. It is also possible to construct the class of priors on the parameters of the AR($p_{\max}$) model that is implied by an already specified prior on the parameters of the ARMA(1,1) model, see Kleibergen (2000). This is convenient because the ARMA(1,1) model is non-linear in its parameters while the AR($p_{\max}$) model is linear. It is therefore not directly obvious how the information in the prior of the ARMA(1,1) parameters is reflected in the marginal posteriors. In linear models, it is clear how prior information is updated with the likelihood to posterior information since all information in the prior is in the same way reflected in the marginal posteriors.

To show the implications for the ARMA(1,1) model, consider a normal prior on (α, ρ) given σ^2, which is frequently used in practice, see, *e.g.,* Chib and Greenberg (1994) and Monahan (1983),

$$p_{\mathrm{ARMA}(1,1)}(\alpha, \rho|\sigma^2) \propto \left|\sigma^2\right|^{-1} \exp\left[-\frac{1}{2\sigma^2}\begin{pmatrix} \rho-\rho_0 \\ \alpha-\alpha_0 \end{pmatrix}' W_0 \begin{pmatrix} \rho-\rho_0 \\ \alpha-\alpha_0 \end{pmatrix}\right], \tag{37}$$

where ρ_0 and α_0 are the prior means and $\sigma^2 W_0^{-1}$, $W_0 : 2 \times 2$, is the prior covariance matrix. This implies the prior

$$p_{\mathrm{ARMA}(1,1)}(\alpha, \theta|\sigma^2) \propto \left|\sigma^2\right|^{-1} \exp\left[-\frac{1}{2\sigma^2}\begin{pmatrix} \theta-\theta_0 \\ \alpha-\alpha_0 \end{pmatrix}' V_0 \begin{pmatrix} \theta-\theta_0 \\ \alpha-\alpha_0 \end{pmatrix}\right], \tag{38}$$

where $\theta = \rho - \alpha$, $\theta_0 = \rho_0 - \alpha_0$, $V_0 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}' W_0 \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$, on θ and α. Just like (33), the prior (38) is a conditional prior of (θ, α) given $\lambda = 0$. We can then construct the class of priors on $(\alpha, \theta, \lambda)$ that imply (38),

$$\begin{aligned} & p_{\mathrm{ARMA}(1,1)}(\alpha, \theta|\sigma^2) \propto p_{\mathrm{AR}(p_{\max})}(\alpha, \theta, \lambda|\sigma^2)|_{\lambda=0} \\ \propto & \left|\sigma^2\right|^{-1} \exp\left[-\frac{1}{2\sigma^2}\left(\varphi(\alpha,\theta,\lambda)|_{\lambda=0} - \varphi_0\right)' V(\theta) \left(\varphi(\alpha,\theta,\lambda)|_{\lambda=0} - \varphi_0\right)\right], \end{aligned} \tag{39}$$

where $V(\theta) = \begin{pmatrix} 1 & 0 \\ 0 & \theta^{-1} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} V_0 \begin{pmatrix} 1 & 0 \\ 0 & \theta^{-1} \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix}' : p_{\max} \times p_{\max}$, $\varphi_0 = \begin{pmatrix} 1 & 0 \\ 0 & \theta \\ 0 & 0 \\ \vdots & \vdots \\ 0 & 0 \end{pmatrix} \begin{pmatrix} \theta_0 \\ \alpha_0 \end{pmatrix}$, and also the class of priors on the parameters of the AR($p_{\max}$) that imply (38),

$$\begin{aligned} p_{\mathrm{AR}(p_{\max})}(\varphi|\sigma^2)|_{c_{i+1}=\alpha^i\theta} &\propto p_{\mathrm{AR}(p_{\max})}(\alpha,\theta,\lambda|\sigma^2)|_{\lambda=0}|J((\alpha,\theta,\lambda),\varphi)|_{\lambda=0}| \\ &\propto |c_1|^{-1} p_{\mathrm{AR}(p_{\max})}(\alpha(\varphi),\theta(\varphi),\lambda(\varphi)|\sigma^2)|_{\lambda=0}. \end{aligned} \tag{40}$$

Although the prior (40) is only specified on values of φ for which $c_{i+1} = \alpha^i\theta$, $i = 1,\ldots,p_{\max}-1$, as it can not be determined for the other values of φ, it reveals the properties of the marginal posteriors of α, ρ and θ when we use the prior (37). This results as the AR($p_{\max}$) model is linear in φ such that all properties that are present in the prior are also present in the marginal posteriors. Furthermore, since we analyze the ARMA(1,1) model, we are also not interested in the behavior of the prior at other values of φ. The implicit prior (40) shows that prior (37) does not account for the local non-identification of α at $\theta = 0$ as (40) is infinite at $\theta = c_1 = 0$. As the AR($p_{\max}$) model is linear in φ, also the marginal posterior of θ with prior (37) is infinite at $\theta = 0$ which corresponds with section 2.1.2 where the marginal posterior of θ is shown to be infinite at $\theta = 0$. This shows a convenient feature of the class of priors on the parameters of the AR($p_{\max}$), that implies the already specified prior on the parameters of the ARMA(1,1) model, as it enables us to verify the plausibility of the prior that is specified on the parameters of the ARMA(1,1) model without the need to compute the marginal posteriors.

3.2.3 Posterior ARMA(p,q)=Conditional Posterior

The ARMA(p,q) model (4) also results from a set of restrictions on the AR($p_{\max}$) model,

$$c_{p+i} = \alpha' \begin{pmatrix} c_{p+i-1} \\ \vdots \\ c_{p+i-q} \end{pmatrix}, \qquad i = 1,\ldots,p_{\max}-p, \tag{41}$$

with $c_0 = 1$, $c_{-i} = 0$, $i = 1,\ldots,\infty$. We can span these restrictions using the parameters,

$$\lambda_i = c_{p+q+i} - \alpha' \begin{pmatrix} \tilde{c}_{p+q+i-1} \\ \vdots \\ \tilde{c}_{p+i} \end{pmatrix} = \begin{pmatrix} 1 & -\alpha' \end{pmatrix} \begin{pmatrix} c_{p+q+i} \\ \tilde{c}_{p+q+i-1} \\ \vdots \\ \tilde{c}_{p+i} \end{pmatrix}, \qquad i = 1,\ldots,p_{\max}-p-q, \tag{42}$$

where

$$\begin{aligned} \tilde{c}_i &= c_i, & i \le p+q \\ &= \alpha' \begin{pmatrix} \tilde{c}_{i-1} \\ \vdots \\ \tilde{c}_{i-q} \end{pmatrix} & i > p+q, \end{aligned} \tag{43}$$

such that we can specify φ as $\varphi = (\varphi_1' \ \varphi_2' \ \varphi_3')'$, $\varphi_1 = (c_1 \cdots c_p)'$, $\varphi_2 = (c_{p+1} \cdots c_{p+q})'$, $\varphi_3 = (c_{p+q+1} \dots c_{p_{\max}})'$, with

$$
\begin{aligned}
\varphi_1 &= \begin{pmatrix} c_1 \\ \vdots \\ c_p \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \vdots \\ \rho_p \end{pmatrix} + C_{12} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_q \end{pmatrix}, \\
\varphi_2 &= \begin{pmatrix} c_{p+1} \\ \vdots \\ c_{p+q} \end{pmatrix} = C_{22} \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_q \end{pmatrix}, \\
c_{p+q+i} &= \lambda_i + \alpha' \begin{pmatrix} \tilde{c}_{p+q+i-1} \\ \vdots \\ \tilde{c}_{p+i} \end{pmatrix}, \qquad i = 1, \dots, p_{\max} - p - q,
\end{aligned} \tag{44}
$$

and where C_{12} and C_{22} are defined in (11)-(12). Since the restriction, spanned by λ (42), does not involve any elements of C_{22}, the specification of (α, ρ, λ) (42)-(44) satisfies the sufficient conditions for a unique expression of the conditional density. The prior and posterior of the parameters of the ARMA(p, q) model are thus conditional densities of a prior and posterior of the parameters of the AR$(p_{\max})$ model.

Normal Prior on AR($p_{\max}$) parameters Analogue to the ARMA(1,1) model, we can specify a normal prior on the parameters of the AR$(p_{\max})$ model (19) and construct the prior that it implies on the parameters of the ARMA(p, q) model,

$$
\begin{aligned}
p_{\mathrm{ARMA}(p,q)}(\alpha, \rho, \sigma^2) &\propto p_{\mathrm{AR}(p_{\max})}(\alpha, \rho, \lambda, \sigma^2)|_{\lambda=0} \\
&\propto p_{\mathrm{AR}(p_{\max})}(\varphi(\alpha, \rho, \lambda), \sigma^2)|_{\lambda=0} |J(\varphi, (\alpha, \rho, \lambda))|_{\lambda=0}| \\
&\propto |C_{22}| \left|\sigma^2\right|^{-\frac{1}{2}(p_{\max}+2)} \exp\left[-\frac{1}{2\sigma^2} \left(\varphi(\alpha, \rho, \lambda)|_{\lambda=0} - \varphi_0\right)' A_0 \left(\varphi(\alpha, \rho, \lambda)|_{\lambda=0} - \varphi_0\right) \right],
\end{aligned} \tag{45}
$$

where $\varphi(\alpha, \rho, \lambda)$ is defined in (44). Since the prior (45) is a conditional density that results from a proper normal prior, it is also proper itself. Similar as for the ARMA(1,1) model, the posterior that results from the prior (45) is also the conditional posterior using the prior (19) of (α, ρ) given that $\lambda = 0$ in the AR$(p_{\max})$ model. Equation (45) shows the functional form of a natural conjugate prior for the parameters of an ARMA(p, q) model as it is the unique conditional density that results from the natural conjugate prior on the parameters of the AR$(p_{\max})$ model.

Diffuse Prior on AR($p_{\max}$) parameters The diffuse prior on the parameters of the AR$(p_{\max})$ model (36) leads to the prior,

$$
p_{\mathrm{ARMA}(p,q)}(\alpha, \rho, \sigma^2) \propto \left|\sigma^2\right|^{-1} |C_{22}|, \tag{46}
$$

on the parameters of the ARMA(p, q) model. For the ARMA(1,1) model, this prior is identical to prior (36) and (46) thus also leads to posteriors that primarily reveal the information in the data. The priors on the parameters of the ARMA(p, q) model that result from other continuous priors on the parameters of the AR$(p_{\max})$ model can be constructed analogously.

4 Bayes Factors

The priors on the parameters of several different ARMA models can be constructed such that they are conditional densities that result from one prior that is specified on the parameters of an encompassing AR($p_{\max}$) model. When we compare these models using the Bayes factor, the priors on the parameters of the different models then accord with one another as they all result from the same prior on the parameters of the encompassing AR($p_{\max}$) model. Another convenience for the Bayes factor is that we only have to specify one prior, the prior on the parameters of the AR($p_{\max}$) model.

The Bayes factor for comparing model the ARMA(p_0, q_0) model H_0 with the ARMA(p_1, q_1) model H_1 is defined as, see *e.g.* Berger (1985),

$$\mathrm{BF}(H_0|H_1) = \frac{\iiint \left[p_{H_0}(\alpha_0, \rho_0, \sigma^2) l_{H_0}(\alpha_0, \rho_0, \sigma^2|y)\right] d\alpha_0 d\rho_0 d\sigma^2}{\iiint \left[p_{H_1}(\alpha_1, \rho_1, \sigma^2) l_{H_1}(\alpha_1, \rho_1, \sigma^2|y)\right] d\alpha_1 d\rho_1 d\sigma^2}, \tag{47}$$

where $p_{H_0}(\alpha_0, \rho_0, \sigma^2)$ and $p_{H_1}(\alpha_1, \rho_1, \sigma^2)$ are the proper priors for the parameters under H_0 and H_1. Both $p_{H_0}(\alpha_0, \rho_0, \sigma^2)$ and $p_{H_1}(\alpha_1, \rho_1, \sigma^2)$ are proper conditional densities that result from the same prior on the parameters of the AR($p_{\max}$) model,

$$\begin{aligned} p_{H_0}(\alpha_0, \rho_0, \sigma^2) &= \frac{p_{\mathrm{AR}(p_{\max})}(\varphi(\alpha_0,\rho_0,\lambda_0),\sigma^2)|_{\lambda_0=0}|J(\varphi,(\alpha_0,\rho_0,\lambda_0))|_{\lambda_0=0}|}{\iiint\left[p_{\mathrm{AR}(p_{\max})}(\varphi(\alpha_0,\rho_0,\lambda_0),\sigma^2)|_{\lambda_0=0}|J(\varphi,(\alpha_0,\rho_0,\lambda_0))|_{\lambda_0=0}|\right]d\alpha_0 d\rho_0 d\sigma^2}, \\ p_{H_1}(\alpha_1, \rho_1, \sigma^2) &= \frac{p_{\mathrm{AR}(p_{\max})}(\varphi(\alpha_1,\rho_1,\lambda_1),\sigma^2)|_{\lambda_1=0}|J(\varphi,(\alpha_1,\rho_1,\lambda_1))|_{\lambda_1=0}|}{\iiint\left[p_{\mathrm{AR}(p_{\max})}(\varphi(\alpha_1,\rho_1,\lambda_1),\sigma^2)|_{\lambda_1=0}|J(\varphi,(\alpha_1,\rho_1,\lambda_1))|_{\lambda_1=0}|\right]d\alpha_1 d\rho_1 d\sigma^2}, \end{aligned} \tag{48}$$

where $(\alpha_0, \rho_0, \lambda_0)$ and $(\alpha_1, \rho_1, \lambda_1)$ satisfy the sufficient conditions for a unique conditional density such that the conditional densities of (α_0, ρ_0) given $\lambda_0 = 0$ and (α_1, ρ_1) given $\lambda_1 = 0$ are invariant with respect to the specification of the parameters. Note that the use of standard normal priors on the parameters of the ARMA(p, q) models implies that the prior on the parameters of the encompassing AR($p_{\max}$), of which the prior of the nested ARMA model is a conditional density, is different for every ARMA model. Hence, the priors of the ARMA models then do not accord with one another.

4.1 Bayes factor for comparing ARMA models with identical p+q

When $p_0 + q_0 = p_1 + q_1$ and we specify a normal prior on φ given σ^2 with an infinite (prior) variance for $(c_{p_0+q_0+1}, \cdots, c_{p_{\max}})$, the Bayes factor simplifies considerably. Not only the normalizing constant of the likelihoods under H_0 and H_1 cancel, also the normalizing constants of the priors are identical. Consider, for example, the case of a normal-gamma prior on (φ, σ^2),

$$p_{\mathrm{AR}(p_{\max})}(\varphi, \sigma^2) = p_{\mathrm{AR}(p_{\max})}(\varphi|\sigma^2) p_{\mathrm{AR}(p_{\max})}(\sigma^2), \tag{49}$$

with

$$\begin{aligned} p_{\mathrm{AR}(p_{\max})}(\sigma^2) &= \left(2^{\frac{1}{2}n_0}\Gamma(\tfrac{1}{2}n_0)\right)^{-1} |\sigma^2|^{-\frac{1}{2}(n_0+2)} \exp\left[-\tfrac{s_0}{2\sigma^2}\right], \\ p_{\mathrm{AR}(p_{\max})}(\varphi|\sigma^2) &= (2\pi)^{-\frac{1}{2}p_{\max}} |A_0| \, |\sigma^2|^{-\frac{1}{2}p_{\max}} \exp\left[-\tfrac{1}{2\sigma^2}(\varphi - \varphi_0)' A_0 (\varphi - \varphi_0)\right], \end{aligned} \tag{50}$$

and where n_0, s_0 are the prior parameters for the prior for σ^2 and φ_0, A_0 for the prior for φ given σ^2, see (19). We specify $\varphi = (\delta' \ \lambda')$, $\varphi_0 = (\delta_0' \ \lambda_0')'$; $\delta, \delta_0 : (p_0+q_0)\times 1$; $\lambda, \lambda_0 : (p_{\max}-(p_0+q_0))\times 1$, and A_0 as

$$A_0 = \begin{pmatrix} A_{0,11} & A_{0,12} \\ A_{0,21} & A_{0,22} \end{pmatrix}, \tag{51}$$

where $A_{0,11} : (p_0 + q_0) \times (p_0 + q_0)$, $A'_{0,21}$, $A_{0,12} : (p_0 + q_0) \times (p_{\max} - (p_0 + q_0))$, $A_{0,22} : (p_{\max} - (p_0 + q_0)) \times (p_{\max} - (p_0 + q_0))$, with $A'_{0,21} = A_{0,12} = 0$, $A_{0,22} = 0$, which leads to a diffuse prior on λ. The proper priors implied by (50) on the parameters of the ARMA models H_0 and H_1 now both correspond with the same $(p_0 + q_0)$-variate normal density,

$$\begin{aligned} p_{H_0}(\alpha_0, \rho_0|\sigma^2) &= (2\pi)^{-\frac{1}{2}(p_0+q_0)} |A_{0,11}| \left|\sigma^2\right|^{-\frac{1}{2}(p_0+q_0)} |J(\delta, (\alpha_0, \rho_0))| \\ &\quad \exp\left[-\frac{1}{2\sigma^2} (\delta(\alpha_0, \rho_0) - \delta_0)' A_{0,11} (\delta(\alpha_0, \rho_0) - \delta_0)\right], \\ p_{H_1}(\alpha_1, \rho_1|\sigma^2) &= (2\pi)^{-\frac{1}{2}(p_0+q_0)} |A_{0,11}| \left|\sigma^2\right|^{-\frac{1}{2}(p_0+q_0)} |J(\delta, (\alpha_1, \rho_1))| \\ &\quad \exp\left[-\frac{1}{2\sigma^2} (\delta(\alpha_1, \rho_1) - \delta_0)' A_{0,11} (\delta(\alpha_1, \rho_1) - \delta_0)\right], \end{aligned} \tag{52}$$

such that it results from (44) that

$$\delta(\alpha_0, \rho_0) = \begin{pmatrix} \rho_0 + C_{12,0}\alpha_0 \\ C_{22,0}\alpha_0 \end{pmatrix}, \ \delta(\alpha_1, \rho_1) = \begin{pmatrix} \rho_1 + C_{12,1}\alpha_1 \\ C_{22,1}\alpha_1 \end{pmatrix}, \tag{53}$$

with $C_{12,0}$, $C_{12,1}$ and $C_{22,0}$ and $C_{22,1}$ are the specifications C_{12} and C_{22} under H_0 and H_1 resp., see (11) and (12), such that $J(\delta, (\alpha_0, \rho_0)) = C_{22,0}$ and $J(\delta, (\alpha_1, \rho_1)) = C_{22,1}$. Note that $|C_{22}| = 1$ for pure AR and MA models. The normalizing constants of the priors under H_0 and H_1 (52) are identical such that they cancel out in the Bayes factor. We can therefore let the prior variances converge to infinity, $A_{0,11} \to 0$, $s_0 \to 0$, $n_0 = 0$, and still maintain a properly defined Bayes factor which reads,

$$\begin{aligned} \mathrm{BF}(H_0|H_1) &= \frac{\iiint \left[|\sigma^2|^{-\frac{1}{2}(p_0+q_0+2)} |J(\varphi_1, (\alpha_0, \rho_0))| l_{H_0}(\alpha_0, \rho_0, \sigma^2|y)\right] d\alpha_0 d\rho_0 d\sigma^2}{\iiint \left[|\sigma^2|^{-\frac{1}{2}(p_0+q_0+2)} |J(\varphi_1, (\alpha_1, \rho_1))| l_{H_1}(\alpha_1, \rho_1, \sigma^2|y)\right] d\alpha_1 d\rho_1 d\sigma^2} \\ &= \frac{\iiint \left[|\sigma^2|^{-\frac{1}{2}(p_0+q_0+2)} |C_{22,0}| l_{H_0}(\alpha_0, \rho_0, \sigma^2|y)\right] d\alpha_0 d\rho_0 d\sigma^2}{\iiint \left[|\sigma^2|^{-\frac{1}{2}(p_0+q_0+2)} |C_{22,1}| l_{H_1}(\alpha_1, \rho_1, \sigma^2|y)\right] d\alpha_1 d\rho_1 d\sigma^2}. \end{aligned} \tag{54}$$

The Bayes factor (54) can be considered as a Bayes factor for comparing ARMA models with equal summed AR and MA lag lengths that uses non-informative priors for the parameters of the compared ARMA models. Normally, the use of improper non-informative priors leads to Bayes factors that suffer from the Lindley paradox, see Poirier (1995). The Bayes factor (54) avoids the Lindley paradox because of the manner in which the priors are constructed and the fact that the compared models have the same number of parameters. In case of normal distributed disturbances, the Bayes factor (54) can be simplified further by analytically integrating out σ^2,

$$\mathrm{BF}(H_0|H_1) = \frac{\iint \left[|\varepsilon(\alpha_0, \rho_0)'\varepsilon(\alpha_0, \rho_0)|^{-\frac{1}{2}(T+p_0+q_0)} |C_{22,0}|\right] d\alpha_0 d\rho_0}{\iint \left[|\varepsilon(\alpha_1, \rho_1)'\varepsilon(\alpha_1, \rho_1)|^{-\frac{1}{2}(T+p_0+q_0)} |C_{22,1}|\right] d\alpha_1 d\rho_1}, \tag{55}$$

where $\varepsilon = (\varepsilon_1 \cdots \varepsilon_T)' : T \times 1$ is the vector of disturbances. The Bayes factor (55) can also be approximated using the Schwarz (Bayesian) Information Criterium (BIC), see Schwarz (1978), $BF(H_0|H_1) \approx \exp\left[\frac{1}{2}(BIC(H_1) - BIC(H_0))\right]$, but this approximation does not involve the jacobian factors $|C_{22,0}|$ and $|C_{22,1}|$. In a later section, we use the Bayes factor (55) to compare different ARMA(p, q) models with identical $p + q$. Note that these models are non-nested and can not be compared using classical test statistics.

5 Posterior Simulators

The developed priors and posteriors of the parameters of ARMA models do not belong to a known class of densities. This implies that we have to use numerical techniques to evaluate the priors and posteriors of the parameters of these kind of models. We therefore construct a simulator that generates drawings from the prior and/or posterior. We note that the priors and posteriors of the parameters of ARMA models, that result as conditional densities of normal and diffuse priors on the parameters of encompassing AR models, are such that the conditional densities of the AR parameters given the MA parameters, and vice versa, do not belong to a known class of densities. This property also results when we specify other kind of priors on the parameters of the ARMA models that result from plausible priors on the parameters of the encompassing AR model. As a consequence, we can not sequentially simulate the AR parameters given the MA parameters and vice versa and use the generated drawings in a Markov Chain Monte Carlo (MC^2) algorithm as, for example, in Chib and Greenberg (1994). The simulator thus has to generate the AR and MA parameters jointly and we can use these jointly generated drawings in, for example, a MC^2 or Importance Sampling algorithm.

5.1 Importance Sampling

We construct an Importance Sampling scheme to compute Bayes factors and prior and posterior moments of the ARMA parameters that result from normal and diffuse priors on the parameters of the encompassing AR model. Also the disturbances are assumed to be normally distributed with mean zero and variance σ^2, which assumption can, however, be relaxed such that we can also, for example, allow for independent student t disturbances. We thus assume that a prior on φ given σ^2 is specified like (19),

$$p_{\mathrm{AR}(p_{\max})}(\varphi|\sigma^2) \propto \left|\sigma^2\right|^{-\frac{1}{2}p_{\max}} \exp\left[-\frac{1}{2\sigma^2}(\varphi-\varphi_0)' A_0 (\varphi-\varphi_0)\right], \tag{56}$$

with $\varphi_0 = (c_{0,1} \cdots c_{0,p_{\max}})'$ and some $j < p_{\max}$ exists such that $c_{0,i} = 0$ for $i > j$, and on σ^2 like (50),

$$p_{\mathrm{AR}(p_{\max})}(\sigma^2) \propto \left|\sigma^2\right|^{-\frac{1}{2}(n_0+2)} \exp\left[-\frac{s_0}{2\sigma^2}\right]. \tag{57}$$

The Importance Function only needs to approximate the prior or posterior where we want to sample from. As initial parameters of our Importance Function, we therefore use the mean and variance of the marginal posterior of the first $p+q$ parameters of an AR($p+q+h$) model, where $p+q+h \geq j$ and $h \geq 0$, where h is set a priori. Using the prior (56)-(57), this posterior reads,

$$p_{\mathrm{AR}(p+q+h)}(\varphi_h, \sigma^2|y) = p_{\mathrm{AR}(p_{\max})}(\varphi_h, \sigma^2|\varphi_{-h} = 0, y), \tag{58}$$

where $\varphi = (\varphi_h' \; \varphi_{-h}')'$, $\varphi_h : (p+q+h) \times 1$, $\varphi_{-h} : (p_{\max} - (p+q+h)) \times 1$, and corresponds with a conditional normal posterior on φ_h given σ^2 and an inverted-gamma posterior for σ^2. We then specify φ_h as $\varphi_h = (\delta' \; \varphi_{h2}')$, $\delta : (p+q) \times 1$, $\varphi_{h2} : h \times 1$, and construct the marginal posterior of δ,

$$p_{\mathrm{AR}(p+q+h)}(\delta|y) = \iint p_{\mathrm{AR}(p+q+h)}(\varphi_h(\delta, \varphi_{h2}), \sigma^2|y) d\varphi_{h2} d\sigma^2, \tag{59}$$

which is a $(p+q)$-variate t density. We denote the posterior mean of δ by $\hat{\delta}$ and its posterior covariance matrix by $\text{cov}(\delta)$. So, in case of, for example, the diffuse prior (46) and we have set h equal to 0, $\hat{\delta} = (Y'_{-(p+q)}Y_{-(p+q)})^{-1}Y'_{-(p+q)}y$, $\hat{\sigma}^2 = \frac{1}{T-(p+q)}(y - Y_{-(p+q)}\hat{\delta})'(y - Y_{-(p+q)}\hat{\delta})$, and $\text{cov}(\delta) = \hat{\sigma}^2(Y'_{-(p+q)}Y_{-(p+q)})^{-1}$, where $x_{-i} = (y_{1-i} \cdots y_{T-i})'$, $i = 0, \ldots, p+q$, $y = x_0$, $Y_{-(p+q)} = (x_{-1} \cdots x_{-(p+q)})$. In case of computing the moments of the prior we proceed according to the above but we use the prior instead of the posterior of the parameters of the AR($p_{\max}$) model.

We use the posterior mean and variances of the parameters of the AR($p+q+h$) as the initial parameters of the Importance Function, which we can update in later rounds of the Importance Sampler. As σ^2 can be integrated out analytically, the Importance Sampler involves the marginal posterior of (α, ρ),

$$\begin{aligned} p_{\text{ARMA}(p,q)}(\alpha, \rho|y) &\propto \int p_{\text{ARMA}(p,q)}(\alpha, \rho, \sigma^2|y) d\sigma^2 \\ &\propto \int \left[p_{\text{AR}(p_{\max})}(\varphi(\alpha,\rho,\lambda), \sigma^2|y)|_{\lambda=0} |J(\varphi, (\alpha,\rho,\lambda)|_{\lambda=0}| \right] d\sigma^2 \\ &\propto |J(\varphi, (\alpha,\rho,\lambda)|_{\lambda=0}| \int \left[p_{\text{AR}(p_{\max})}(\varphi(\alpha,\rho,\lambda), \sigma^2|y)|_{\lambda=0} \right] d\sigma^2 \\ &\propto p_{\text{AR}(p_{\max})}(\varphi(\alpha,\rho,\lambda))|y)|_{\lambda=0} |J(\varphi, (\alpha,\rho,\lambda)|_{\lambda=0}|, \end{aligned} \tag{60}$$

where, for example, in case of the diffuse prior (46), $p_{\text{AR}(p_{\max})}(\varphi(\alpha,\rho,\lambda))|y)|_{\lambda=0}$ reads,

$$p_{\text{AR}(p_{\max})}(\varphi(\alpha,\rho,\lambda))|y)|_{\lambda=0} \propto |\varepsilon(\alpha,\rho)'\varepsilon(\alpha,\rho)|^{-\frac{1}{2}T}. \tag{61}$$

The resulting Importance Sampling scheme is then given by

Importance Sampling Scheme for ARMA parameters

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

2. Generate δ^i from $q(\delta) \propto \left[\upsilon + (\delta - \hat{\delta})'(\text{cov}(\delta))^{-1}(\delta - \hat{\delta})\right]^{-\frac{1}{2}(\upsilon+p+q)}$.

3. Solve for ρ^i and α^i from δ^i using (11)-(14).

4. Construct weight:

$$\begin{aligned} w(\rho^i, \alpha^i) &= \frac{p_{\text{ARMA}(p,q)}(\alpha^i, \rho^i|y)}{q_{\text{ARMA}(p,q)}(\alpha^i, \rho^i|y)} = \frac{p_{\text{ARMA}(p,q)}(\alpha^i, \rho^i|y)}{|J(\delta^i, (\alpha^i, \rho^i))| q(\delta^i|y)} \\ &= \frac{p_{\text{AR}(p_{\max})}(\varphi(\alpha^i, \rho^i, \lambda)|y)|_{\lambda=0} |J(\varphi^i, (\alpha^i, \rho^i, \lambda)|_{\lambda=0}|,}{|J(\delta^i, (\alpha^i, \rho^i))| \left[\upsilon + (\delta^i - \hat{\delta})'(\text{cov}(\delta))^{-1}(\delta^i - \hat{\delta})\right]^{-\frac{1}{2}(\upsilon+p+q)}} \\ &= \frac{p_{\text{AR}(p_{\max})}(\varphi(\alpha^i, \rho^i, \lambda)|y)|_{\lambda=0}}{\left[\upsilon + (\delta^i - \hat{\delta})'(\text{cov}(\delta))^{-1}(\delta^i - \hat{\delta})\right]^{-\frac{1}{2}(\upsilon+p+q)}}, \end{aligned} \tag{62}$$

since $|J(\varphi, (\alpha,\rho,\lambda)|_{\lambda=0}| = |C_{22}| = |J(\delta, (\alpha,\rho))|$.

5. Set $i = i+1$, and if $i < N$ go to step 2.

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

7. To improve numerical accuracy, update $\hat{\delta}$ and $\text{cov}(\hat{\delta})$ by considering $g(\rho,\alpha) = \delta$, set $i = 1$ and go to step 2.

In step 3 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 (AR)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. The posterior densities of the ARMA parameters can then be 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.

Computation of Bayes Factor for Models with Equal Summed AR and MA lag lengths In section 4.1 we constructed a Bayes factor for comparing two ARMA models H_0 and H_1 that have equal summed AR and MA lag lengths. The priors on the parameters of both models resulted from the same normal prior on the parameters of the encompassing AR($p_{\max}$) model such that the Bayes factor can be made independent of the prior by letting the prior variance converge to infinity. It results in the Bayes factor (55). This Bayes factor can be computed using Importance Sampling, see Geweke (1989b). The Bayes factor equals the ratio of the marginal likelihoods under both models. In Geweke (1989a), it is shown that

$$\sqrt{N}\left(\frac{1}{N}\sum_{i=1}^{N} w(\rho^i,\alpha^i) - \frac{\iint p_{\text{ARMA}(p,q)}(\alpha,\rho|y)d\alpha d\rho}{\iint q_{\text{ARMA}(p,q)}(\alpha,\rho)d\alpha d\rho}\right) \Rightarrow N(0,\omega), \tag{63}$$

where $p_{\text{ARMA}(p,q)}(\cdot)$ is the (unnormalized) posterior, $q_{\text{ARMA}(p,q)}(\cdot)$ is the importance density, $\Rightarrow$ indicates weak convergence, and $\omega = E((w(\vartheta) - E(w(\vartheta)))^2)$, which can be estimated by $\omega \approx \frac{1}{N}\sum_{i=1}^{N} w(\rho^i,\alpha^i)^2 - (\frac{1}{n}\sum_{i=1}^{N} w(\rho^i,\alpha^i))^2$. Equation (63) can be used to estimate the marginal likelihood

$$\iint p_{\text{ARMA}(p,q)}(\alpha,\rho|y)d\alpha d\rho \approx \left(\iint q_{\text{ARMA}(p,q)}(\alpha,\rho)d\alpha d\rho\right) \times \left(\frac{1}{N}\sum_{i=1}^{N} w(\rho^i,\alpha^i)\right). \tag{64}$$

Note that sofar we represented the Importance Function by its kernel, without the normalizing constants. In the construction of the Bayes factor however we need to include these normalizing constants. Doing this, the Bayes factor is approximated by

$$\text{BF}(\text{H}_0|\text{H}_1) \approx \left[\frac{\Gamma(\frac{1}{2}(\lambda_1+p_1+q_1))\Gamma(\frac{1}{2}\lambda_0)\lambda_1^{\frac{1}{2}\upsilon_1}}{\Gamma(\frac{1}{2}(\lambda_0+p_0+q_0))\Gamma(\frac{1}{2}\lambda_1)\lambda_0^{\frac{1}{2}\upsilon_0}}\left(\frac{|\text{cov}(\delta_0)|}{|\text{cov}(\delta_1)|}\right)^{\frac{1}{2}}\right]\left\{\frac{\frac{1}{N_0}\sum_{i=1}^{N_0} w(\rho^i,\alpha^i,\text{H}_0)}{\frac{1}{N_1}\sum_{i=1}^{N_1} w(\rho^i,\alpha^i,\text{H}_1)}\right\}, \tag{65}$$

where $w(\rho,\alpha,\text{H}_j)$ are the weights for model j, N_j is the number of Importance Sampling drawings from model j, υ_j is the degrees of freedom of the Importance function used for model j and $\text{cov}(\delta_j)$ is the covariance matrix of the Importance functions used for model j. If

$\upsilon_1 = \upsilon_2$ the weight ratio approximating the Bayes factor simplifies to,

$$\text{BF}(\text{H}_0|\text{H}_1) \approx \left[\left(\frac{|\text{COV}(\delta_0)|}{|\text{COV}(\delta_1)|}\right)^{\frac{1}{2}} \right] \left\{ \frac{\frac{1}{N0}\sum_{i=1}^{N_0} w(\rho^i,\alpha^i,\text{H}_0)}{\frac{1}{N_1}\sum_{i=1}^{N_1} w(\rho^i,\alpha^i,\text{H}_1)} \right\}. \tag{66}$$

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 6 we apply the Bayes factor to compare different ARMA models for the extended Nelson-Plosser data, see Schotman and van Dijk (1993).

The Bayes factor for comparing ARMA models with different summed AR and MA lag lengths and/or informative proper priors can also be computed using the Importance Sampler. The Importance Sampler is constructed such that it can easily accomodate informative priors that result from a normal prior on the parameters of the encompassing AR($p_{\max}$) model. For these kind of priors, the Importance Sampler also has to be used to compute the normalizing constant of the prior on the ARMA parameters, see (48). The Bayes factor then follows from (47) where the marginal likelihoods can again be computed according to (64)-(65).

5.2 Metropolis-Hastings Sampling

Instead of Importance Sampling, we can also use the Metropolis-Hastings (MH) algorithm, see, *e.g.*, Chib and Greenberg (1995). The Metropolis sampling algorithm can be set up as follows

Metropolis Sampling Scheme for ARMA parameters

1. Choose starting values $(\sigma^2)^0$, α^0, ρ^0, the number of iterations, N, and set $i = 1$. Note that also $\delta^0 := (c_1^0, \ldots, c_{p+q}^0)$ is implicitly chosen.

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

$$\psi = \frac{w(\rho^{\text{new}}, \alpha^{\text{new}}, \delta^{\text{new}})}{w(\rho^{i-1}, \alpha^{i-1}, \delta^{i-1})}, \tag{67}$$

where

$$w(\rho, \alpha, \delta) = \frac{p_{\text{AR}(p_{\max})}(\varphi(\rho, \alpha, \lambda)|y, \sigma^2)|_{\lambda=0}}{N(\delta|\hat{\delta}, \text{cov}(\delta))}. \tag{68}$$

Note that the jacobians cancel out in the weight function and that the MH acceptance probability can be interpreted as the ratio of the importance weight in the model with given σ^2. Next, with probability ψ we set $(\rho^i, \alpha^i, \delta^i) = (\rho^{\text{new}}, \alpha^{\text{new}}, \delta^{\text{new}})$ and with probability $(1 - \psi)$ we set $(\rho^i, \alpha^i, \delta^i) = (\rho^{i-1}, \alpha^{i-1}, \delta^{i-1})$.

3. Conditional on ρ and α, σ^2 has an inverted-Wishart distribution. Generate σ^2 from this distribution.

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

Note again that the identifying parameter matrix C_{22} is obtained as a byproduct in step 2, such that also the posterior of the diagonal elements of C_{22} can be obtained, and that the above MC^2 sampler generates α and ρ jointly.

6 Empirical Application

We conduct a Bayesian analysis of ARMA models for the extended Nelson-Plosser data. This data set consists of yearly observations of 14 macro-economic 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).

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

$$\rho(L)y_t = \alpha(L)(\varepsilon_t + c + dt), \tag{69}$$

where $\rho(L) = 1 - \rho_1 L - \ldots - \rho_p L^p$, $\alpha(L) = 1 - \alpha_1 L - \ldots - \alpha_q L^q$, $\varepsilon_t \sim N(0, \sigma^2)$. We specify the deterministic components such that,

$$c(L)y_t = \varepsilon_t + c + dt, \tag{70}$$

where $c(L) = \alpha(L)^{-1}\rho(L)$, and we can integrate out c and d analytically from the joint posterior of $(\alpha, \rho, \sigma^2, c, d)$, when diffuse or normal priors are specified on c and d. For example, in case of the prior (46) on (α, ρ, σ^2) and a flat prior on c and d, the analysis corresponds with the ARMA(p, q) model,

$$\rho(L)\tilde{y}_t = \alpha(L)\varepsilon_t, \tag{71}$$

where $\tilde{y}_t$ are the residuals that result after regressing a constant and linear trend on y_t.

We computed the marginal posteriors of the parameters of the ARMA(p, q) model (71) using the prior (46) on (α, ρ, σ^2). We also calculated the Bayes factor (55) for comparing models with equal summed AR and MA lag lengths using the average weights that result from the Importance Sampling Algorithm (65). The Importance 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. Bayes factors are calculated for ARMA(3,0) [=AR(3)], ARMA(2,1), ARMA(1,2) and ARMA(0,3) [=MA(3)] models. The resulting Bayes factors are listed in table 3. We also approximated the Bayes factors using the Schwarz (Bayesian) Information Criterium (BIC), see Schwarz (1978), $BF(H_0, H_1) \approx \exp[\frac{1}{2}(BIC(H_1) - BIC(H_0))]$, of which we obtained estimates from EVIEWS. For the series for which EVIEWS was capable to give reasonably precise parameter estimates, the Bayes factors from both procedures are close to one another. For the non-precise estimates, the Bayes factors were rather different as the Bayes factors resulting from the BIC's are inprecise. The numerical errors for the Bayes factors resulting from the Importance Sampler are also in these cases very small such that we prefer this latter procedure for calculating the Bayes factor.

The Bayes factors from table 3 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 Bayes factor. This holds, for example, for Industrial Production, Employment, Unemployment, Consumer Price Index, Interest and the Standard and Poor 500. For other series the Bayes factors indicate that both models are more or less equally likely. The ARMA(2,1)

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 3: Bayes Factors Extended Nelson-Plosser Series

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 4: Posterior Moments ARMA parameters Nelson-Plosser Series

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 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 4, which contains the posterior means and standard deviations (given below the means) of the ARMA model that is preferred by the Bayes factor from table 3. 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 (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 6. 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 Bayes factors from table 3.

Since for many series contained in table 3, the posterior means indicate that either ρ_2 or/and α_1 lies close to zero, we calculated the Bayes factors of an AR(2) model compared to an ARMA$(1,1)$ model for these series. The resulting Bayes factors are listed in table 5.

Table 5 shows that Industrial Production, Employment, Real Wages and S&P 500 are better characterized by an ARMA(1,1) than an AR(2) model according to the Bayes factors. The opposite holds for Real GNP, Nominal GNP, Wages and Money. This accords with the results in tables 3 and 4 which show that these series are either preferred to be AR(3) or that the MA parameter α_1 lies relatively closer to 0 than the AR parameter ρ_2. Table 6 shows the posterior moments of the parameters of the resulting ARMA(1,1) models.

Table 6 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 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

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 5: Bayes Factors for AR(2) vs. ARMA(1,1) Nelson-Plosser Series

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

Table 6: Posterior Moments of ARMA(1,1) model for Nelson-Plosser Series

parameter is, however, similar in both models, indicating that the information regarding the long run behavior is not affected by deleting ρ_2.

7 Conclusions

An ARMA model is nested within an encompassing AR model. Since the restriction that is imposed by the ARMA model on the parameters of the encompassing AR model satisfies the sufficient conditions for a unique conditional density developed in Kleibergen (2000), the prior and posterior of the parameters of the ARMA model are conditional densities of a prior and posterior of the parameters of the encompassing AR model. We can thus specify a prior and posterior on the parameters of the encompassing AR model and construct the prior and posterior that it implies on the parameters of the ARMA model. Priors on the parameters of ARMA models that result from standard priors on the parameters of the encompassing AR model lead to posteriors that are similar to those that result in finite order AR models. Because of the local non-identification of parameters, standard priors that are directly specified on the parameters of the ARMA model do not lead to such kind of posteriors. We construct Importance and Metropolis-Hasting simulators to generate parameters from the priors and posteriors of ARMA models. Also Bayes factors for model comparison are developed.

For the conducted applications, the Importance Sampling Algorithm converged rapidly and, quite surprisingly, we discovered 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.

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 MC^2 framework.

References

[1] Berger, J.O. *Statistical Decision Theory and Bayesian Inference.* Springer-Verlag (New York), (1985).

[2] Billingsley, P. *Probability and Measure.* Wiley (New York), (1986).

[3] Box G.E.P., G.M. Jenkins and G.C. Reinsel. *Time Series Analysis: Forecasting and Control.* Prentice Hall, 1994.

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

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

[6] Cragg, J.C. and S.G. Donald. On the asymptotic properties of LDU-based tests of the rank of a matrix. *Journal of the American Statistical Association*, 91:1301–1309, 1996.

[7] Dickey, J.M. The Weighted Likelihood Ratio, Linear Hypotheses on Normal Location Parameters. *The Annals of Mathematical Statistics*, **42**:204–223, 1971.

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

[9] Galbraith, J. W. and V. Zinde-Walsh. Simple Estimation and Identification Techniques for General ARMA Models. *Biometrika*, (1995).

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

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

[12] Gill, L. and A. Lewbel. Testing the rank and definiteness of estimated matrices with applications to factor, state-space and ARMA models. *Journal of the American Statistical Association*, 87:766–776, 1992.

[13] Golub, G.H. and C.F. van Loan. *Matrix Computations.* The John Hopkins University Press (Baltimore), (1989).

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

[15] Kadane, J.B. The Role of Identification in Bayesian Theory. In S.E. Fienberg and A. Zellner, editors, *Studies in Bayesian Econometrics and Statistics*, pages 341–361. North-Holland (Amsterdam), (1993).

[16] Kleibergen, F. Conditional Densities in Statistical Modelling. Working paper, University of Amsterdam, 2000.

[17] Kleibergen F. and E. Zivot. Bayesian and Classical Approaches to Instrumental Variable Regression. Econometric Institute Report 9835/A, 1998.

[18] Kleibergen, F. and H.K. van Dijk. Direct Cointegration Testing in Error Correction Models. *Journal of Econometrics*, **63**:61–103, (1994a).

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

[20] Kleibergen, F. and H.K. van Dijk. Bayesian Simultaneous Equation Analysis using Reduced Rank Structures. *Econometric Theory*, **14**:701–743, 1998.

[21] Kleibergen, F. and R. Paap. Priors, Posteriors and Bayes Factors for a Bayesian Analysis of Cointegration. Econometric Institute Report 9821/A, 1998.

[22] Kolmogorov, A.N. *Foundations of the Theory of Probability.* Chelsea (New York), (1950).

[23] Marriot, J.M. and A.F.M. Smith. Reparameterization Aspects of Numerical Bayesian Methodology for Auto-Regressive Moving Average Models. *Journal of Time Series Analysis*, **13**:327–343, 1992.

[24] Marriott, J.M., N. Ravishanker, A. Gelfand and J. Pai. Bayesian Analysis of ARMA processes: Complete Sampling Based Inference under Exact Likelihoods. In Berry, D.A., K.M. Chaloner and J. Geweke, editor, *Bayesian Analysis in Statistics and Econometrics: Essays in Honor of Arnold Zellner.* Wiley, New York, 1995.

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

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

[27] Phillips, P.C.B. Partially Identified Econometric Models. *Econometric Theory*, **5**:181–240, (1989).

[28] Poirier, D.J. *Intermediate Statistics and Econometrics: A Comparative Approach.* MIT Press, (Cambridge, MA), (1995).

[29] Schotman, P. C. and H.K. Van Dijk. *Posterior Analysis of Possibly Integrated Time Series with an Application to Real GNP*, pages 341–361. Springer Verlag (Berlin), (1993).

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

[31] Verdinelli, I. and L. Wasserman. Computing Bayes Factors Using a Generalization of the Savage-Dickey Density Ratio. *Journal of the American Statistical Association*, **90**:614–618, 1995.

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