Bayesian Analysis of ARMA models using Noninformative Priors

Frank Kleibergen* Henk Hoek†

December 5, 1995

Abstract

Parameters in ARMA models are only locally identified. It is shown that the use of diffuse priors in these models leads to a preference for locally nonidentified parameter values. We therefore suggest to use likelihood based priors like the Jeffreys' priors which overcome these problems. An algorithm involving Importance Sampling for calculating the posteriors of ARMA parameters using Jeffreys' priors is constructed. This algorithm is based on the implied AR specification of ARMA models and shows good performance in our applications. As a byproduct the algorithm allows for the computation of the posteriors of diagnostic parameters which show the identifiability of the MA parameters. As a general to specific modeling approach to ARMA models suffers heavily from the previous mentioned identification problems, we derive posterior odds ratios which are suited for comparing (nonnested) parsimonious (low order) ARMA models. These procedures are applied to two datasets, the (extended) Nelson-Plosser data and monthly observations of US 3-month and 10 year interest rates. For approximately 50% of the series in these two datasets an ARMA model is favored above an AR model which has important consequences for especially the long run parameters.

1 Introduction

Auto Regressive Moving Average (ARMA) models form the basis of time series analysis, see a.o. [7], and are commonly used in applied work. They do possess some well known problems, however, which result from the possibility to explain autocovariances both from an autoregressive as a moving average perspective. It

*Department of Econometrics and Center, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands, email: f.r.kleibergen@kub.nl

†Tinbergen and Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands, email: hoek@tir.few.eur.nl

is important to be able to distinguish between these two though as they lead to different kinds of long run behavior. As the covariances can be explained using both kind of polynomials, ARMA models contain parameters which do not affect the analyzed model for certain specific values of the other parameters. These parameters are said to be locally nonidentified and are common to a lot of models used in econometrics, for example the Simultaneous Equations Model. In the classical statistical literature the behavior of these locally nonidentified parameters has led to a substantial literature covering this subject in detail, see a.o. [15]. The literature on Bayesian analysis of the posteriors of locally nonidentified parameters is still quite small though, see [11]. The consequences of the locally nonidentification on the posteriors can be quite dramatic, however, leading to a strong favor of this local nonidentification when diffuse priors are used. To overcome this implicit favor when using diffuse priors, in [11] the Jeffreys' prior is investigated and it is shown there that this type of prior functionalizes the classical statistical manner of conducting inference in these kind of models, i.e. conduct inference on the locally nonidentified parameters conditional on a test for identification of these parameters. The use of Jeffreys' priors in models containing lagged dependent variables is not entirely clearcut, however, see [16] and [9]. This results from the expectations which have to be constructed to obtain the Jeffreys' prior. Two different manners for constructing these expectations exist. The first one constructs the expectations such that they become a function of the model parameters, for an example see [16]. The second one considers the data as given and takes the expectations as equal to their realizations, see [9]. Since we are also confronted with this problem we briefly discuss it and our interpretation in the third section.

As mentioned before, the Bayesian literature on locally nonidentified parameters is still quite small but this refers also to the literature on Bayesian analysis of ARMA models, see a.o. [13] and [1]. We differ from this literature as we use a different Monte Carlo posterior simulator (Importance Sampling instead of Gibbs as in [1]) and our analysis explicitly focuses on the modeling problems involved when using ARMA models. We are therefore able to construct statistics which show whether these problems are appearing. Since the Gibbs Sampling approach pursued in [1] does not use priors exploiting the a priori knowledge, regarding the identification of the different parameters, its performance will be less than that of the Importance Sampling approach involving priors which do exploit this a priori knowledge.

The paper is organized as follows. Section 2 explains the local identification problem existing in ARMA(1,1) models and shows how it generalizes to ARMA(p, q) models. In section 3, Jeffreys' type priors for ARMA models are discussed and a prior is constructed which is flat in the implicit AR specification. Section 4 contains the Monte Carlo Posterior Simulator for ARMA models. This simulator is straightforward and allows for the construction of diagnostic parameters indicating the identification of the different ARMA parameters. In this section also Posterior Odds Ratios for comparing different (parsimonious) ARMA models, with equal combined AR and MA lengths, are constructed. As a general to specific approach is, because of the identification problems involved, not applicable to ARMA models this procedure allows for a comparison of small parsimonious models. The fifth section contains an application of the developed procedures to two data sets, i.e. the (extended) Nelson-Plosser data and monthly observations of U.S. 3-month and 10-year interest rates. For almost 50% of these series an ARMA model is favored above an AR model and for some series the favor is pronounced. The MA components are shown to strongly influence the long run components of the AR components in some cases leading to different conclusions about the long run behavior of these series. Finally, the sixth section concludes.

2 Identification consequences for the posteriors in ARMA models

The local nonidentification problems arising in ARMA models are the result of so-called root cancellation, i.e. the AR and MA polynomials have common roots. We show this phenomenon for the simplest ARMA model, an ARMA$(1,1)$ model,

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

where L is the lag-operator, $L^j y_t = y_{t-j}$. The local identification problem becomes obvious when we consider a reparameterization of the model,

$$(1-(\alpha+\delta)L)y_t = (1-\alpha L)\varepsilon_t, \tag{2}$$

where $\delta = \rho - \alpha$. When the parameter $\delta = 0$, which corresponds to $\alpha = \rho$, α disappears from the model as the same kind of operator $(1-\alpha L)$ is used on both sides of equation (2) such that it simplifies to $y_t = \varepsilon_t$, a white noise model. So, when $\delta = 0$ $(\alpha = \rho)$, we cannot learn anything about α and the likelihood of the model will be flat (nonzero) in the direction of α. When we use a diffuse prior such that the joint posterior of the parameters is proportional to the likelihood, the conditional posterior of α given δ is flat and nonzero at $\delta = 0$. Consequently, the integral over this conditional posterior, which is part of the marginal posterior of δ, is infinite. So, the marginal posterior of δ is infinite at $\delta = 0$. These arguments are similar to those in [9] and [11] where similar phenomena are analyzed for cointegration and Simultaneous Equations Models. In section 4.3, the consequences of the use of a diffuse prior on the posterior of the parameters of an ARMA(1,1) model are illustrated. These posteriors are also compared with the posteriors using the priors derived in the following sections.

In general ARMA(p,q) models the local identification problem previously discussed for the ARMA(1,1), is known as the problem of root cancellation. Consider a general ARMA(p, q) model,

$$\rho(L)y_t = \alpha(L)\varepsilon_t, \tag{3}$$

where $\rho(z) = 1 - \sum_{j=1}^{p} \rho_j z^j$ and $\alpha(z) = 1 - \sum_{j=1}^{q} \alpha_j z^j$. Both polynomials can be represented using their roots, $\rho(z) = \prod_{j=1}^{p}(1 - \lambda_j z)$, $\alpha(z) = \prod_{j=1}^{q}(1 - \theta_j z)$. Consider θ_1, every AR root λ_j can be specified as $\lambda_j = \theta_1 + a_j + b_j i$. When both a_j and b_j are 0, the parameter θ_1 drops out of the model and is consequently not identified. A similar kind of reasoning can be performed for the other MA roots and a complicated nonlinear region for the AR parameters results where MA parameters are nonidentified. In section 4.1 a procedure for testing the identification of the MA parameters is constructed.

Note that the autocorrelations of noninvertible MA models, i.e. models with roots of the MA polynomial which lie within the unit circle, cannot be distinguished from the autocorrelations of invertible MA models. Consequently, MA parameters have to be restricted leading to invertible polynomials to be identifiable from the autocorrelations. Invertible and noninvertible MA polynomials with identical autocorrelations lead to different values of the likelihood function, however, such that they can be identified from the likelihood. As we define identification from a likelihood perspective, see also [8], we allow for noninvertible MA parameters and explosive AR parameters, such that the MA and AR parameters range from $-\infty$ to ∞.

3 Priors for ARMA models

As we know a priori that ARMA models contain parameters that are locally nonidentified for specific values of the other parameters, it is possible to construct the prior such that it reflects that certain parameters need to be analyzed conditional on the other parameters. Likelihood based priors like the Jeffreys' prior, which is proportional to the square root of the determinant of the information matrix, possess this property. In the following sections we discuss how Jeffreys' priors for ARMA models can be constructed.

3.1 Information matrix priors for the AR(1) model

In time series models, the curvature of the likelihood strongly depends on the values of the parameters of the model which generated the data. In AR(1) models for example, we know that the curvature of the likelihood is much more pronounced when the analyzed data is generated by a unit root value of the AR(1) parameter than for a stationary value of the AR(1) parameter. As this curvature of the likelihood strongly depends on the value of the parameters generating the data also the information matrix measuring this curvature strongly depends on these parameter values. When we use priors depending on this information matrix, like the Jeffreys' prior, these priors will strongly vary over the model parameters, see a.o. [16], [9], [19] and [20]. Consequently, when using Jeffreys' priors, certain specific parameters values get more weight a priori. It is arguable whether this leads to sensible inference as it can bias the posteriors quite strongly since large a priori weight is attached to parameter values leading to a strong curvature of the likelihood (unit root and explosive parameter values). It can be shown, see [9], that the value of these kind of priors crucially depends on how certain expectations are taken in the construction of the prior. If in the construction of the prior the data is considered as given, the resulting Jeffreys' prior is flat in AR(1) models and does not depend on the value of the AR(1) parameter. To show this, consider the AR(1) model,

$$y_t = \rho y_{t-1} + \varepsilon_t. \tag{4}$$

Assuming $(T+1)$ observations, $t = 0, ..., T$, and i.i.d. disturbances with a normal distribution with mean 0 and variance σ^2, the Jeffreys' prior for this model reads,

$$p(\rho, \sigma^2) \propto \left| \sigma^{-3} E(\sum_{t=1}^{T} y_{t-1}^2) \right|^{\frac{1}{2}}. \tag{5}$$

When the expectation is taken the prior reads, see [16],

$$p(\rho, \sigma^2) \propto \left| \sigma^{-3} \left(\frac{T\sigma^2}{1-\rho^2} + \frac{1-\rho^{2T}}{1-\rho^2} \left(E(y_0^2) - \frac{\sigma^2}{1-\rho^2} \right) \right) \right|^{\frac{1}{2}}. \tag{6}$$

As shown by equation (6), the value of the prior crucially depends on the value of ρ and gives a lot of a priori weight to large values of ρ. It can be argued that taking the expectation over the y_{t-1}'s in equation (5) is not necessary as the posterior is conditional on the data and the realized values of the y_{t-1}'s can therefore be used instead of their expectations. Use of a prior which shows that the data contain more information for larger values of ρ is therefore misleading as the amount of information in the data does not change for a different value of ρ as the data does not change. The resulting prior then reads,

$$p(\rho, \sigma^2) \propto \left| \sigma^{-3} \sum_{t=1}^{T} y_{t-1}^2 \right|^{\frac{1}{2}}. \tag{7}$$

The prior in equation (7) is uniform in the parameter ρ and does therefore give different values of ρ equal a priori weight. The reason why we stress this phenomenon is that similar difficulties arise in the construction of the Jeffreys' prior in ARMA models and we show this in the following section for the ARMA(1,1) model.

3.2 Information matrix priors for the ARMA(1,1) model

As specification for the ARMA$(1,1)$ model, we use, see equation (1),

$$y_t = \rho y_{t-1} - \alpha\varepsilon_{t-1} + \varepsilon_t, \qquad t = 1, ..., T, \tag{8}$$

where we assume $(T+2)$ observations y_t, $t = -1, 0, ..., T$, and the disturbances, ε_t, $t = 1, ..., T$, are i.i.d. with a normal distribution with mean 0 and variance σ^2. The resulting AR and MA polynomials are $\rho(z) = 1 - \rho z$ and $\alpha(z) = 1 - \alpha z$ such that $\rho(L)y_t = \alpha(L)\varepsilon_t$ and L is the usual lag operator. By inverting the MA polynomial, the specification of the disturbances results,

$$\varepsilon_t = y_t - \sum_{i=1}^{t+1} \alpha^{i-1}(\rho - \alpha)y_{t-i} = y_t - \sum_{i=1}^{t+1} c_i y_{t-i}, \tag{9}$$

where $c_1 = \rho - \alpha$, $c_2 = \alpha(\rho - \alpha)$, $c_i = c_2(c_2/c_1)^{i-2}$, $i = 3, ..., T+1$. Note that the construction of the disturbances as in equation (9) allows for noninvertible MA parameters as we explicitly assume that T is finite. This manner of constructing the disturbances will be used throughout the whole paper. Using the disturbances from equation (9), the information matrix can be derived,

$$p(\theta, \sigma^2) \propto |I(\theta, \sigma^2)|^{\frac{1}{2}} \propto \sigma^{-(2+k)} \left| E\left(\sum_{t=1}^{T} \left(\frac{\partial\varepsilon_t}{\partial\theta}\right)\left(\frac{\partial\varepsilon_t}{\partial\theta}\right)'\right)\right|^{\frac{1}{2}}, \tag{10}$$

where $\theta : k \times 1$. So, for the construction of the information matrix we need to construct the derivative $\frac{\partial\varepsilon_t}{\partial\theta}$, where $\theta = \begin{pmatrix} \rho \\ \alpha \end{pmatrix}$, $k = 2$,

$$\frac{\partial\varepsilon_t}{\partial\theta} = -\begin{pmatrix} \sum_{i=0}^{t} \alpha^i y_{t-i-1} \\ \sum_{i=0}^{t} \alpha^{i-1}(i\rho - (i+1)\alpha)y_{t-i-1} \end{pmatrix}. \tag{11}$$

Conditioning on the observed values of the observations y_t, $t = -1, 0, ..., T$, such that $E(y_{t-i}) = y_{t-i}$, the specification of the Jeffreys' prior reads,

$$p(\rho, \alpha, \sigma^2) \propto \sigma^{-4}$$
$$\left|\sum_{t=1}^{T}\begin{pmatrix} \sum_{i=0}^{t} \alpha^i y_{t-i-1} \\ \sum_{i=0}^{t} \alpha^{i-1}(i\rho - (i+1)\alpha)y_{t-i-1} \end{pmatrix}\begin{pmatrix} \sum_{i=0}^{t} \alpha^i y_{t-i-1} \\ \sum_{i=0}^{t} \alpha^{i-1}(i\rho - (i+1)\alpha)y_{t-i-1} \end{pmatrix}'\right|^{\frac{1}{2}}. \tag{12}$$

The Jeffreys' prior in equation (12) is increasing in the value of the MA parameter α. So, parameter points with relatively large values of α get more a priori weight than parameter points with a relatively small value of α. This is comparable to the Jeffreys' prior in equation (6) where large values of the AR parameter get more a weight than smaller values. For both cases it holds that the resulting posteriors have a (unplausibly) large amount of probability mass at large values of the MA resp. AR parameter. For both cases we find this unplausible and we therefore choose a prior which only takes into account the first $(p+q)$ lags of the implicit AR polynomial in equation (9). This prior is flat in the implicit AR parameters (which are properly identified) which explains why we choose for this specification.

The derivative $\frac{\partial \varepsilon_t}{\partial \theta}$ then becomes,

$$\begin{aligned}
\frac{\partial \varepsilon_t}{\partial \theta} &= -\begin{pmatrix} \sum_{i=0}^{1} \alpha^i y_{t-i-1} \\ \sum_{i=0}^{1} \alpha^{i-1}(i\rho-(i+1)\alpha)y_{t-i-1} \end{pmatrix} \\
&= -\begin{pmatrix} y_{t-1}+\alpha y_{t-2} \\ y_{t-1}+(\rho-2\alpha)y_{t-2} \end{pmatrix} = -\begin{pmatrix} 1 & \alpha \\ -1 & \rho-2\alpha \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}.
\end{aligned} \tag{13}$$

Using this specification of the derivative $\frac{\partial \varepsilon_t}{\partial \theta}$, the resulting "Jeffreys' prior" becomes,

$$\begin{aligned}
p(\rho,\alpha,\sigma^2) &\propto \sigma^{-4}\left|\begin{pmatrix} 1 & \alpha \\ -1 & \rho-2\alpha \end{pmatrix}\sum_{t=1}^{T}(\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}')\begin{pmatrix} 1 & \alpha \\ -1 & \rho-2\alpha \end{pmatrix}'\right|^{\frac{1}{2}} \\
&\propto \sigma^{-4}\left|\begin{pmatrix} 1 & \alpha \\ -1 & \rho-2\alpha \end{pmatrix}\right|\left|\sum_{t=1}^{T}\left(\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}'\right)'\right|^{\frac{1}{2}} \\
&\propto \sigma^{-4}|\rho-\alpha|\left|\sum_{t=1}^{T}\left(\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}'\right)'\right|^{\frac{1}{2}}.
\end{aligned} \tag{14}$$

Although the Jeffreys' prior, when using the derivative in equation (13), is still increasing in the MA parameter α, it is flat for the parameters of the implicit AR model shown in equation (9) when this model is specified in terms of the parameters c_1 and c_2. The matrix $\begin{pmatrix} 1 & \alpha \\ -1 & \rho-2\alpha \end{pmatrix}$ is namely the jacobian of the transformation of c $(=(c_1\ c_2)')$ to θ $(=(\rho\ \alpha)')$ such that $dc = \left|\begin{pmatrix} 1 & \alpha \\ -1 & \rho-2\alpha \end{pmatrix}\right| d\theta$. So, in the specification of the parameters (c,σ^2), the prior reads,

$$p(c_1,c_2,\sigma^2) \propto \sigma^{-4}\left|\sum_{t=1}^{T}\left(\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}\begin{pmatrix} y_{t-1} \\ y_{t-2} \end{pmatrix}'\right)'\right|^{\frac{1}{2}}. \tag{15}$$

This prior is clearly flat in the parameters (c_1,c_2) as it does not depend on these parameters. In the implicit AR model from equation (9) all parameters are properly identified and there are no local identification issues as in the original ARMA model from equation (8). The implicit AR model can therefore be considered to be the implicit reduced form model of the (structural) ARMA model. This is comparable to the Simultaneous Equations Model (SEM) where also local identification issues arise in the structural form specification while the parameters of the reduced form are properly identified, see a.o. [11] and [15]. By using a flat prior in the specification where the parameters are properly identified, we penalize the points were the identification issues documented in the previous section appear. The implicit posterior favor for the points of local nonidentification therefore disappears, see also [9] and [11] for a discussion of Bayesian analyses of other kind of models with locally nonidentified parameters.

The use of the prior from equation (14) can be generalized in a natural manner to ARMA models of any specific order and also allows for a straightforward importance sampling scheme. These issues are discussed in the following sections.

3.3 Noninformative Priors for general ARMA(p,q) models

To derive "Jeffreys' type priors" for general univariate ARMA(p, q) model including exogenous variables, we consider three different specifications of the exogenous variables within an ARMA model: a specification incorporating the exogenous variables within the ARMA polynomial, a specification incorporating the exogenous variables within the MA polynomial and a specification where the exogenous variables directly influence the disturbances without the intervenience of a specific lag polynomial. In the following, we discus each of the three different specifications and the "Jeffreys' priors" and posteriors resulting from them.

The first specification incorporates the exogenous variables within the ARMA polynomial,

$$\begin{aligned} y_t &= x_t'\beta + \varepsilon_t, \\ \rho(L)\varepsilon_t &= \alpha(L)u_t, \end{aligned} \tag{16}$$

where $x_t : k \times 1$, is generated independently from the variables y_t (weakly exogenous), $t = -(p + q - 1), ..., 0, ..., T$; and $\rho(z) = 1 - \sum_{i=1}^{p} \rho_i z^i,\ \alpha(z) = 1 - \sum_{i=1}^{q} \alpha_i z^i$. The model in equation (16) is identical to the model,

$$\rho(L)(y_t - x_t'\beta) = \alpha(L)u_t. \tag{17}$$

The infinite order AR specification of the model in equation (17) reads,

$$\begin{aligned} \alpha(L)^{-1}\rho(L)y_t = \alpha(L)^{-1}\rho(L)x_t'\beta + u_t \quad &\Leftrightarrow \\ c(L)y_t = c(L)x_t'\beta + u_t, & \end{aligned} \tag{18}$$

where $c(z) = 1 - \sum_{i=1}^{\infty} c_i z^i,\ c_{p+q+j} = f_j(c_1, ..., c_{p+q}),\ j = 1, ..., \infty$. Note that as in the previous section, we construct the inverses of the MA polynomials such that we allow for "noninvertible MA polynomials". When we multiply the inverted MA polynomial, $\alpha(L)^{-1}$, by a certain observation, say y_t, resulting in $\alpha(L)^{-1}y_t$, we assume that $\alpha(L)^{-1}y_t = \sum_{i=0}^{t+p+q} c_i y_{t-i}$ is such that even in case of a noninvertible MA polynomial we can still construct $\alpha(L)^{-1}y_t$ as we only use elements of c_i with a finite i, implying finite c_i's in the construction of $\alpha(L)^{-1}y_t$.

Equation (18) allows us to decompose the "Jeffreys' prior" when assuming i.i.d. normal disturbances with mean 0 and variance σ^2 into "conditional and marginal priors", see also [10]. This results from the dependence of β on the ARMA parameters $\alpha_i,\ i = 1,...,q;\ \rho_j,\ j = 1,...,p$. The conditional Jeffreys' prior of β given $\alpha_i,\ i = 1,...,q;\ \rho_j,\ j = 1,...,p$, therefore becomes,

$$\begin{aligned} p(\beta|\alpha,\rho,\sigma^2) &\propto \sigma^{-k}\left|\sum_{t=1}^{T}(\alpha(L)^{-1}\rho(L)x_t x_t'\rho(L)\alpha(L)^{-1})\right|^{\frac{1}{2}} \\ &\propto \sigma^{-k}|(Xc(L))'(Xc(L))|^{\frac{1}{2}}, \end{aligned} \tag{19}$$

where $Xc(L) = (x_1c(L)...x_Tc(L))'$, while the marginal prior for (α,ρ,σ^2) reads,

$$p(\alpha,\rho,\sigma^2) \propto \sigma^{-(p+q+2)}\left|\frac{\partial(c_1,...,c_{p+q})}{\partial(\alpha_1,...,\alpha_q,\rho_1,...,\rho_p)}\right||Y'M_{Xc(L)}Y|^{\frac{1}{2}}, \tag{20}$$

where $Y = (\begin{pmatrix} y_0 \\ . \\ y_{-p-q+1}\end{pmatrix}' ... \begin{pmatrix} y_{T-1} \\ . \\ y_{T-p-q}\end{pmatrix}'),\ M_Z = I_T - Z(Z'Z)^{-1}Z',\ Z = Xc(L).$
The prior in equation (20) shows that a prior is used for the parameters of the implicit AR($p+q$) model which is proportional to $|Y'M_{Xc(L)}Y|^{\frac{1}{2}}$. The conditional prior for β given (α,ρ,σ^2) contains part of the scaling factor of the multivariate t conditional posterior of β given (α,ρ). The conditional posterior of β given (α,ρ) therefore reads,

$$\begin{aligned} p(\beta|\alpha,\rho,data) \propto \left|\sum_{t=1}^{T} c(L)x_t x_t' c(L)\right|^{\frac{1}{2}} \\ \left|(yc(L))'M_{Xc(L)}(yc(L)) + (\beta-\hat{\beta})'\left(\sum_{t=1}^{T} c(L)x_t x_t' c(L)\right)(\beta-\hat{\beta})\right|^{-\frac{1}{2}(T+k)} \end{aligned} \tag{21}$$

and the marginal posterior of (α,ρ) reads,

$$\begin{aligned} p(\alpha,\rho|data) \propto \left|\tfrac{\partial(c_1,...,c_{p+q})}{\partial(\alpha_1,...,\alpha_q,\rho_1,...,\rho_p)}\right||Y'M_{Xc(L)}Y|^{\frac{1}{2}} \\ |(yc(L))'M_{Xc(L)}yc(L)|^{-\frac{1}{2}(T+p+q)} \end{aligned} \tag{22}$$

such that the marginal posterior of c becomes,

$$p(c|data) \propto |Y'M_{Xc(L)}Y|^{\frac{1}{2}}|(yc(L))'M_{Xc(L)}yc(L)|^{-\frac{1}{2}(T+p+q)}, \tag{23}$$

where $yc(L) = (y_1 c(L)...y_T c(L))'$, $\hat{\beta} = (\sum_{t=1}^{T} c(L)x_t x_t' c(L))^{-1}(\sum_{t=1}^{T} c(L)x_t y_t c(L)) = ((Xc(L))'Xc(L))^{-1}(Xc(L))'yc(L)$, $c(L) = \alpha(L)^{-1}\rho(L)$. The functional expressions in equations (16) to (23) are all based on the specification of the exogenous variables from equation (16).

The second specification of the exogenous variables we consider allows for lagged influence of the exogenous variables on the disturbances through the MA polynomial,

$$\rho(L)y_t = x_t'\beta + \alpha(L)\varepsilon_t. \tag{24}$$

This property becomes clearer when we construct the infinite AR specification,

$$\alpha(L)^{-1}\rho(L)y_t - \alpha(L)^{-1}x_t'\beta = \varepsilon_t. \tag{25}$$

The infinite AR specification shows that there is lagged influence of the exogenous variables on the disturbances through the MA polynomial. The "Jeffreys' prior" for the parameters of the model in equation (24) can again be decomposed in a conditional prior for β given (α, ρ, σ^2) and a marginal prior of (α, ρ, σ^2). The conditional prior of β given (α, ρ, σ^2) then reads,

$$\begin{aligned} p(\beta|\alpha, \rho, \sigma^2) &\propto \sigma^{-k}\left|\sum_{t=1}^{T}(\alpha(L)^{-1}x_t x_t'\alpha(L)^{-1})\right|^{\frac{1}{2}} \\ &\propto \sigma^{-k}|(X\alpha(L)^{-1})'(X\alpha(L)^{-1})|^{\frac{1}{2}}, \end{aligned} \tag{26}$$

where $X\alpha(L)^{-1} = (x_1\alpha(L)^{-1}...x_T\alpha(L)^{-1})'$, while the marginal prior for (α, ρ, σ^2) reads,

$$p(\alpha, \rho, \sigma^2) \propto \sigma^{-(p+q+2)}\left|\frac{\partial(c_1, ..., c_{p+q})}{\partial(\alpha_1, ..., \alpha_q, \rho_1, ..., \rho_p)}\right||Y'M_{X\alpha(L)^{-1}}Y|^{\frac{1}{2}}. \tag{27}$$

For the implicit AR parameters, the prior in equation (27) shows that we have used an implicit prior for these parameters which is proportional to $|Y'M_{X\alpha(L)^{-1}}Y|^{\frac{1}{2}}$. The conditional prior for β given (α, ρ, σ^2) contains part of the scaling factor of the multivariate t conditional posterior of β given (α, ρ). The conditional posterior of β given (α, ρ) therefore reads,

$$\begin{gathered} p(\beta|\alpha, \rho, data) \propto \left|\sum_{t=1}^{T}\alpha(L)^{-1}x_t x_t'\alpha(L)^{-1}\right|^{\frac{1}{2}} \\ \left|(yc(L))'M_{X\alpha(L)^{-1}}(yc(L)) + (\beta - \hat{\beta})'\left(\sum_{t=1}^{T}\alpha(L)^{-1}x_t x_t'\alpha(L)^{-1}\right)(\beta - \hat{\beta})\right|^{-\frac{1}{2}(T+k)} \end{gathered} \tag{28}$$

and the marginal posterior of (α, ρ) reads,

$$\begin{gathered} p(\alpha, \rho|data) \propto \left|\frac{\partial(c_1, ..., c_{p+q})}{\partial(\alpha_1, ..., \alpha_q, \rho_1, ..., \rho_p)}\right||Y'M_{X\alpha(L)^{-1}}Y|^{\frac{1}{2}} \\ |(yc(L))'M_{X\alpha(L)^{-1}}yc(L)|^{-\frac{1}{2}(T+p+q)} \end{gathered} \tag{29}$$

such that the marginal posterior of c becomes,

$$p(c|data) \propto |Y'M_{X\alpha(L)^{-1}}Y|^{\frac{1}{2}}|(yc(L))'M_{X\alpha(L)^{-1}}yc(L)|^{-\frac{1}{2}(T+p+q)}. \tag{30}$$

For our third specification of the exogenous variables, it holds that the exogenous variables directly influence the disturbances without the intervenience of a lag polynomial. The specification of this model reads,

$$\rho(L)y_t = \alpha(L)(x_t'\beta + \varepsilon_t). \tag{31}$$

The direct influence becomes again clearer when the implicit AR model is constructed,

$$\alpha(L)^{-1}\rho(L)y_t - x_t'\beta = \varepsilon_t, \tag{32}$$

which clearly shows the direct influence. As the exogenous variables influence the disturbances instantaneously, the parameters of the exogenous variables can be estimated by means of regression. Essentially, the specification of the exogenous variables in equation (31) is the most general one since by adding lagged exogenous variables, the other specifications can also be obtained. This specification also allows for other kind of dynamic influence of the exogenous variables on the dependent variables compared to the dynamic influence of the dependent variables on itselves. The specification in equation (31) is therefore also more general in this respect. The prior can again be decomposed into a conditional prior of β given (α, ρ, σ^2) and a marginal prior of (α, ρ, σ^2) but there is an important difference between the priors for this specification of the exogenous variables and the priors in the previous specifications, as the conditional prior of β given (α, ρ, σ^2) is not influenced by α and ρ. The conditional prior of β given (α, ρ, σ^2) reads,

$$\begin{aligned} p(\beta|\alpha, \rho, \sigma^2) &\propto \sigma^{-k}\left|\sum_{t=1}^{T} x_t x_t'\right|^{\frac{1}{2}} \\ &\propto \sigma^{-k}|X'X|^{\frac{1}{2}}, \end{aligned} \tag{33}$$

where $X = (x_1 ... x_T)'$, while the marginal prior for (α, ρ, σ^2) reads,

$$p(\alpha, \rho, \sigma^2) \propto \sigma^{-(p+q+2)}\left|\frac{\partial(c_1, ..., c_{p+q})}{\partial(\alpha_1, ..., \alpha_q, \rho_1, ..., \rho_p)}\right||Y'M_XY|^{\frac{1}{2}}, \tag{34}$$

The prior in equation (34) shows that a flat prior is used for the parameters of the implicit AR($p+q$) model which does not depend on the value of α or ρ. The conditional prior for β given (α, ρ, σ^2) contains part of the scaling factor of the multivariate t conditional posterior of β given (α, ρ). The conditional posterior of β given (α, ρ) therefore reads,

$$\begin{gathered} p(\beta|\alpha, \rho, data) \propto \left|\sum_{t=1}^{T} x_t x_t'\right|^{\frac{1}{2}} \\ \left|(yc(L))'M_X(yc(L)) + (\beta - \hat{\beta})'\left(\sum_{t=1}^{T} x_t x_t'\right)(\beta - \hat{\beta})\right|^{-\frac{1}{2}(T+k)} \end{gathered} \tag{35}$$

and the marginal posterior of (α, ρ) reads,

$$p(\alpha, \rho|data) \propto \left|\frac{\partial(c_1,\ldots,c_{p+q})}{\partial(\alpha_1,\ldots,\alpha_q,\rho_1,\ldots,\rho_p)}\right| |Y'M_XY|^{\frac{1}{2}} \\ |(yc(L))'M_Xyc(L)|^{-\frac{1}{2}(T+p+q)}, \tag{36}$$

such that the marginal posterior of c becomes,

$$p(c|data) \propto |Y'M_XY|^{\frac{1}{2}}|(yc(L))'M_Xyc(L)|^{-\frac{1}{2}(T+p+q)}. \tag{37}$$

The marginal posteriors of α and ρ differ over the specifications of the exogenous variables depending on whether the disturbances of the model are influenced by the exogenous variables through a certain lag polynomial or directly. For certain specific choices of the exogenous variables, the marginal posterior of the ARMA parameters is identical regardless of what specification is chosen. A necessary and sufficient condition to achieve this irrelevance of specification of the exogenous variables is that $x_t = Ax_{t-1}$, where A is a $k \times k$ matrix, and invertibility of the MA polynomial. The marginal posteriors of the ARMA parameters of ARMA Models containing only constant terms, linear trends (quadratic trends, etc.) are therefore not affected by the choice of the specification of these exogenous variables. In the fifth section, we calculate the marginal posteriors of ARMA parameters of ARMA models only containing a constant term and a linear trend and these posteriors are as a consequence not affected by the way in which these deterministic components are incorporated in the model.

As the marginal posteriors in the different equations in this section, like (22), (29) and (36) do not belong to a known class of probability density functions, in the next sections Monte Carlo simulation procedures for the calculation of moments of these posteriors are constructed. Furthermore, also the posterior odds ratios to test for the lag length of the AR and MA polynomials are constructed.

4 Numerical analysis of the posteriors of parameters of ARMA models

4.1 Efficient Generation of Model Parameters

As the posteriors in equations (22), (29) and (36) do not belong to a standard class of probability density functions, the moments of these probability density functions are unknown. We therefore calculate these moments using Monte Carlo Integration. Direct simulation from the posterior is, however, not feasible and we use other simulation techniques. In [1], the marginal posteriors of the parameters of ARMA models are calculated using Gibbs sampling. The priors involved in that analysis are of the natural conjugate type while the posteriors in equations (22), (29) and (36) involve "Jeffreys' type priors". Consequently, we cannot apply the Gibbs Sampling algorithm developed in [1]. We use Importance Sampling, see [12] and [5], to calculate the marginal posteriors of parameters of ARMA models. As the Importance Function involved is often a good approximation of the posterior, it can also be used to simulate directly from the posterior using acceptance-rejection sampling.

The Importance function used to approximate the posterior in equation (23) is a $(p+q)$ dimensional multivariate t density initially based on the (least squares) estimates of an AR$(p+q)$ model. The following simulation scheme for generating the parameters of the ARMA(p,q) model is used. In this scheme, n shows the a priori chosen number of random drawings from the Importance Function, λ shows the degrees of freedom of the t Importance Function and $f_t(c|\hat{c}, cov(\hat{c}), \lambda)$ is the probability density function of a $(p+q)-$ variate t density with kernel, $|\lambda + (c-\hat{c})'(cov(\hat{c}))^{-1}(c-\hat{c})|^{-\frac{1}{2}(\lambda+p+q)}$.

Importance Sampling Scheme for ARMA parameters

0. Choose degrees of freedom Importance function, λ, and number of drawings, n.

1. Estimate c, β: $y_t = \sum_{i=1}^{p+q} c_i y_{t-i} + x_t'\beta + u_t$

2. Construct $\hat{c}$, $cov(\hat{c})$

3. Generate c_i, $i = 1, ..., n$, from $f_t(c|\hat{c}, cov(\hat{c}), \lambda)$

4. Solve for α, ρ using :

$$\begin{pmatrix} 1 & -\alpha_1 & ... & -\alpha_q \end{pmatrix} \begin{pmatrix} 1 & -c_1 & & . & . & . & & -c_{p+q} \\ 0 & . & . & & & & & . \\ . & . & . & . & & & & . \\ . & & . & . & . & & & . \\ 0 & . & . & 0 & 1 & -c_1 & ... & -c_p \end{pmatrix} =$$

$$\begin{pmatrix} 1 & -\rho_1 & ... & -\rho_p & 0 & ... & 0 \end{pmatrix}$$

5. Construct weight : $w_i(\alpha, \rho) = \frac{|(yc(L))' M_X yc(L)|^{-\frac{1}{2}(T+p+q)}}{|\lambda + (c-\hat{c})'(cov(\hat{c}))^{-1}(c-\hat{c})|^{-\frac{1}{2}(\lambda+p+q)}}$

6. Construct $E(g(\alpha, \rho)) = \frac{\sum_{i=1}^{n} w_i(\alpha,\rho) g_i(\alpha,\rho)}{\sum_{i=1}^{n} w_i(\alpha,\rho)}$

7. Update $\hat{c}$ and $cov(\hat{c})$ using $E(g(\alpha, \rho))$'s and go to step 3.

Note that through the transformation from c to (α,ρ), we implicitly took account of the jacobian of this transformation such that this jacobian, which appears in the posterior in equation (22), does not appear in the weight function, $w(\alpha,\rho)$. The above Importance Sampling scheme is intended for the model in equation (31), the model where the exogenous variables directly influence the disturbances. When we use the Importance Sampling scheme for the specification of the exogenous variables as in equation (16), the weights change to,

$$w_i(\alpha,\rho) = \frac{|Y'M_{XC(L)}Y|^{\frac{1}{2}}|(yc(L))'M_{Xc(L)}yc(L)|^{-\frac{1}{2}(T+p+q)}}{|\lambda+(c-\hat{c})'(cov(\hat{c}))^{-1}(c-\hat{c})|^{-\frac{1}{2}(\lambda+p+q)}}, \tag{38}$$

if instead the model in equation (24) is used, the weights change to,

$$w_i(\alpha,\rho) = \frac{|Y'M_{X\alpha(L)^{-1}}Y|^{\frac{1}{2}}|(yc(L))'M_{X\alpha(L)^{-1}}yc(L)|^{-\frac{1}{2}(T+p+q)}}{|\lambda+(c-\hat{c})'(cov(\hat{c}))^{-1}(c-\hat{c})|^{-\frac{1}{2}(\lambda+p+q)}}. \tag{39}$$

Sampling β can be done from the conditional posteriors from equations (21), (28) and (35) given a value of α and ρ. The generated values of β get the same weight as the values of α and ρ.

Generating the ARMA parameters α and ρ from the AR parameters c enables us to calculate some interesting diagnostic parameters which are helpful in the determination of the lag length of the MA polynomial. In the next section also a posterior odds procedure for determining this lag length is presented but here we a use matrix decomposition to investigate it. This decomposition involves a Töplitz matrix which is used to construct the MA parameters as (it is here assumed that $p \geq q$ but $p < q$ follows trivially),

$$\begin{pmatrix} \alpha_1 & \dots & \alpha_q \end{pmatrix} \begin{pmatrix} c_p & \dots & c_{p+q-1} \\ . & . & . \\ c_{p-q} & \dots & c_p \end{pmatrix} = \begin{pmatrix} c_{p+1} & \dots & c_{p+q} \end{pmatrix} \Leftrightarrow \alpha'\Lambda_C = \varphi_c', \tag{40}$$

where $\Lambda_c = \begin{pmatrix} c_p & \dots & c_{p+q-1} \\ . & . & . \\ c_{p-q+1} & \dots & c_p \end{pmatrix}$ and $\varphi_c = \begin{pmatrix} c_{p+1} & \dots & c_{p+q} \end{pmatrix}'$. When eigenvalues of the Töplitz matrix Λ_c lie in the neighborhood of 0, the lag length q exceeds the MA lag length of the model which generated the analyzed data and elements of α are consequently nearly nonidentified. The parameters contained in α and ρ can be large in that case as an inversion of a near singular matrix is involved. It is therefore important to construct the eigenvalues of the matrix Λ_c to obtain information about the rank of Λ_c. Instead of constructing the eigenvalues we decompose the matrix Λ_c into a product of two triangular matrices, see also [9] and [11] where triangular matrices are used to determine the rank of the long run multiplier in cointegration models, to avoid the identification problems which appear when the eigenvalues need to be ordered in increasing values. This decomposition reads,

$$\Lambda_c = \begin{pmatrix} \theta_{11} & 0 & . & 0 \\ . & . & . & . \\ . & & . & 0 \\ \theta_{q1} & . & . & \theta_{qq} \end{pmatrix} \begin{pmatrix} 1 & \psi_{12} & . & \psi_{1q} \\ 0 & . & . & . \\ . & . & . & \psi_{q-1q} \\ 0 & . & 0 & 1 \end{pmatrix} \Leftrightarrow \Lambda_c = \Theta_c \Psi_c, \tag{41}$$

where $\Theta_c = \begin{pmatrix} \theta_{11} & 0 & . & 0 \\ . & . & . & . \\ . & & . & 0 \\ \theta_{q1} & . & . & \theta_{qq} \end{pmatrix}$ and $\Psi_c = \begin{pmatrix} 1 & \psi_{12} & . & \psi_{1q} \\ 0 & . & . & . \\ . & . & . & \psi_{q-1q} \\ 0 & . & 0 & 1 \end{pmatrix}$. The number of diagonal elements of Θ_c, θ_{ii}, $i = 1, ..., q$; differing from 0 show the lag length of the MA polynomial. These parameters show the identifiability of the MA parameters and they can be used to test for the identification of these. The diagonal elements in the neighborhood of 0 lead to fat tailed (Cauchy type) posteriors of the ARMA parameters such that we prefer to have no diagonal elements close to 0. Although the diagonal elements of Λ_c are indicative about the lag length, they should always be analyzed jointly with the MA parameters, α. Consider for example an AR(1) model which is estimated using an ARMA(1,1). In this model $\Lambda_c = \rho \neq 0$ but $\alpha = 0$ showing that both α and Λ_c have to be analyzed. In [4], the specification in equation (40) is used to derive classical statistical tests for identification of the parameters in ARMA models.

The use of the Λ_c parameter matrix and the MA coefficients α allows us to conduct inference regarding the orders of the AR and MA polynomials. It remains troublesome, however, to conduct inference regarding both orders simultaneously. The AR and MA parameters are namely up to a large extent determined by the autocovariances of low order. So, if a general to specific approach is conducted to determine the different orders of the polynomials, it is difficult to do so as both AR and MA parameters can explain these autocovariances. In the next section, a posterior odds approach is developed which compares (low order) ARMA models with an equal number of ARMA parameters, $p+q$. In this manner, identification problems resulting from the problem whether the low order autocovariances are explained by MA or AR polynomials are circumvented. As these ARMA models with equal an equal ARMA order, $p+q$, are not nested in one another, it is hard to compare these models using classical statistical analysis.

4.2 AR and MA lag lengths comparison using Posterior Odds Ratios

Difficulties in a general to specific approach in ARMA modeling arise as autocovariances can both be explained using MA and AR polynomials. We therefore propose to perform model selection using posteriors odds ratios of parsimonious models. In this approach models with an equal number of ARMA parameters, $p+q$, are compared. As the number of ARMA parameters does not differ between the models, we specify identical uniform priors and prior odds for the parameters, c, contained in the implicit AR model such that the priors cancel out in the posterior odds ratios (Note that these parameters c lead to different kind of implicit AR models). By comparing parsimonious models (models for which all parameters are more or less different from 0) the difficulties arising in the general to specific approach can be avoided. We compare for example ARMA(2,1) and ARMA(1,2) models with each other and/or AR(3), MA(3) models since these models are all equally capable in explaining the first three autocovariances but are different for higher order autocovariances. If instead a general to specific approach is conducted which starts from an ARMA(3,3) model encompassing all four different models individually, and then afterwards tests for the restrictions imposed by the different models are conducted, it is typically hard to identify especially the MA parameters as the AR parameters can explain part of (short run) behavior resulting from a MA polynomial.

The posterior odds ratios used for comparing the different models can be calculated using Importance Sampling, see [6]. In [5], it is proved that the weights used in the Importance Sampling procedure converge to the ratio of the integral of the posterior and the Importance function and that the limiting distribution is normal,

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} w_i(\alpha,\rho) - \frac{\int p(c|data)dc}{\int I(c)dc}\right) \Rightarrow n(0,\omega), \tag{42}$$

where $I(c) = |\lambda+(c-\hat{c})'(cov(\hat{c}))^{-1}(c-\hat{c})|^{-\frac{1}{2}(\lambda+p+q)}$, is the importance function, $\Rightarrow$ indicates weak asymptotic convergence; $\omega = E(w(\alpha,\rho)-E(w(\alpha,\rho)))^2$, $\frac{1}{n}\sum\limits_{i=1}^{n} w_i(\alpha,\rho)^2 - (\frac{1}{n}\sum\limits_{i=1}^{n} w_i(\alpha,\rho))^2 \Rightarrow \omega$; and $p(c|data) = |(yc(L))'M_X yc(L)|^{-\frac{1}{2}(T+p+q)}$, is the kernel of the posterior, see also equations (36) and (37). As the Importance function is a multivariate t density, the integral $\int I(c)dc$ is analytically known, see [21],

$$\int I(c)dc = \pi^{\frac{1}{2}}\frac{\Gamma(\frac{1}{2}\lambda)}{\Gamma(\frac{1}{2}(\lambda+p+q))}|cov(\hat{c})|^{\frac{1}{2}}\lambda^{-\frac{1}{2}\lambda}. \tag{43}$$

Sofar, we represented the kernel of the posterior without its normalizing constants. In the construction of the posterior odds we exactly need to know these normalizing constants. In the construction of the kernel used in the Importance sampling scheme, two analytical integration steps were conducted beforehand. First, the variance of the disturbances σ^2 is integrated out using the normalizing constants of an inverted-Wishart probability density function and second the parameters of the exogenous variables, β, are integrated out using the normalizing constants of a multivariate t probability density function. In the following, these different integration steps are conducted separately, such that we obtain the constants present in the marginal posterior of the ARMA parameters.

Initially, the joint posterior including constants reads,

$$
\begin{aligned}
p(\alpha, \beta, \rho, \sigma^2 | data) = (2\pi)^{-\frac{1}{2}T} \sigma^{-(T+k+p+q+2)} \left| \tfrac{\partial(c_1,\ldots,c_{p+q})}{\partial(\alpha_1,\ldots,\alpha_q,\rho_1,\ldots,\rho_p)} \right| \\
|Y'M_X Y| \ \left| \sum_{t=1}^{T} x_t x_t' \right|^{\frac{1}{2}} \exp\left[-\tfrac{1}{2\sigma^2} \sum_{t=1}^{T} \varepsilon_t^2 \right].
\end{aligned} \tag{44}
$$

By transforming the parameters α and ρ to c, the jacobian appearing in the posterior in equation (44) disappears. Integrating out σ^2 using an inverted-Wishart probability density function results in the joint posterior of c and β,

$$
\begin{aligned}
p(c, \beta | data) = 2^{\frac{1}{2}(k+p+q)} \pi^{-\frac{1}{2}T} \Gamma(\tfrac{1}{2}(T+k+p+q)) \\
|Y'M_X Y|^{\frac{1}{2}} \left| \sum_{t=1}^{T} x_t {x_t}' \right|^{\frac{1}{2}} \left| \sum_{t=1}^{T} \varepsilon_t^2 \right|^{-\frac{1}{2}(T+k+p+q)}.
\end{aligned} \tag{45}
$$

If we further integrate out β from the joint posterior in equation (45) using the conditional posterior of β given c shown in equation (35), which is a multivariate t, the marginal posterior of c is obtained including its normalizing constants,

$$
\begin{aligned}
p(c | data) = 2^{\frac{1}{2}(k+p+q)} \pi^{-\frac{1}{2}(T-k)} \Gamma(\tfrac{1}{2}(T+p+q)) \\
|Y'M_X Y|^{\frac{1}{2}} \ |(yc(L))' M_X yc(L)|^{-\frac{1}{2}(T+p+q)}.
\end{aligned} \tag{46}
$$

The posteriors in the previous equations are all assuming a diffuse (Jeffreys') prior for the implicit AR parameters, c, and Jeffreys' type priors for β and σ^2. For the construction of posterior odds ratios proper priors are needed. We use the posterior odds ratios for comparing models, say H_1 and H_2, with similar exogenous variable structures, $k_{H_1} = k_{H_2}$, and with an equal number of ARMA parameters, $p_{H_1} + q_{H_1} = p_{H_2} + q_{H_2}$. Assuming identical priors, prior odds and parameter regions for the implicit AR parameters, c, under the different hypotheses, the posterior odds ratios do not depend on the choice of the prior. The posterior odds ratio then reads,

$$
P_{H_1|H_2} = \frac{\int p(c|data, H_1) dc}{\int p(c|data, H_2) dc}. \tag{47}
$$

Using the limiting behavior of the weight factors, this posterior odds ratio can be approximated by,

$$
\frac{\frac{1}{n_1} \sum_{i=1}^{n_1} w_i(\alpha, \rho, H_1)}{\frac{1}{n_2} \sum_{i=1}^{n_2} w_i(\alpha, \rho, H_2)} \frac{\Gamma(\frac{1}{2}(\lambda_2 + p + q)) \Gamma(\frac{1}{2}\lambda_1) \lambda_2^{\frac{1}{2}\lambda_2}}{\Gamma(\frac{1}{2}(\lambda_1 + p + q)) \Gamma(\frac{1}{2}\lambda_2) \lambda_1^{\frac{1}{2}\lambda_1}} \left(\frac{|cov(\hat{c}_1)|}{|cov(\hat{c}_2)|} \right)^{\frac{1}{2}} \Rightarrow P_{H_1|H_2}, \tag{48}
$$

where $w_i(\alpha, \rho, H_j)$ are the weights for model j; n_1, n_2 are the number of Importance Sampling drawings from model 1 resp.2; λ_1, λ_2 are the degrees of freedom of the Importance functions used for model 1 resp.2; $cov(\hat{c}_1), cov(\hat{c}_2)$ are the covariance functions of the Importance functions used for model 1 resp. 2. When the degrees of freedom of the Importance Functions are equal to one another, the weight ratio approximating the posterior odds ratio simplifies to,

$$\frac{\frac{1}{n_1}\sum_{i=1}^{n_1} w_i(\alpha,\rho,H_1)}{\frac{1}{n_2}\sum_{i=1}^{n_2} w_i(\alpha,\rho,H_2)}\left(\frac{|cov(\hat{c}_1)|}{|cov(\hat{c}_2)|}\right)^{\frac{1}{2}} \Rightarrow P_{H_1|H_2}. \tag{49}$$

When the model in the denominator is an AR model, such that $p_{H_2} = p_{H_1} + q_{H_1}$, the weight ratio approximating the posterior odds ratio further simplifies as the integral in the denominator can be constructed analytically. In section 5, these posterior odds ratios are used to compare ARMA models for the (extended) Nelson-Plosser data, see [14], [18] and [3], and monthly observations of the 3-month and 10 year US interest rates.

4.3 Example

To show the consequences of the use of a diffuse or Jeffreys' prior, we compare the posteriors of the parameters for an artificial time series, generated from an ARMA(1,1) model, see equation (8), with $\rho = 0.6$, $\alpha = 0.4$, $\sigma^2 = 0.005$, $T = 200$ ($\theta = \rho - \alpha = 0.2$). We calculated the posteriors for the parameters of an ARMA(1,1) model both using a diffuse and a Jeffreys' prior. For the diffuse prior, the posteriors are calculated using the analytical expression of the bivariate posterior of (α, ρ), which is proportional to the conditional likelihood. For the Jeffreys' prior, the Importance Sampling Algorithm from section 4.1 is used.

Figures 1 to 7 contain the marginal posteriors of the parameters of an ARMA(1,1) model for the artificially generated time-series. Table 1 contains the posterior means and standard deviations of the different marginal posteriors. The bivariate posterior of θ and α and its contourlines are shown in figures 1,2 (diffuse prior) and 3,4 (Jeffreys' prior). The bivariate posterior and its contourlines show that the bivariate posterior using the diffuse prior is constant in the direction of α around $\theta = 0$. So, the posterior using the diffuse prior has much more probability mass at $\theta = 0$ compared to the posterior using a Jeffreys' prior. The marginal posteriors of θ shown in figure 5 confirm this as the marginal posterior using the diffuse prior has a secondary mode at $\theta = 0$ such that this posterior has more probability mass at $\theta = 0$. In theory the value at $\theta = 0$ of this posterior is infinite as we have integrated over a parameter, α, which does not influence the nonzero joint posterior of (θ, α) at $\theta = 0$. We have chosen a finite parameter region of α $(-1.3, 1.3)$, however, such that the posterior in figure 5 is finite at $\theta = 0$ as the integral of a constant function over a finite region is finite. Note the direct linkage between the size of the parameter region and the posterior value at $\theta = 0$. So, when using a diffuse prior there is a implicit favor for $\theta = 0$. The larger probability mass at $\theta = 0$ also reflects itselves in the marginal posterior of α and ρ, shown in figures 6 and 7. For both figures, it holds that the marginal posterior

prior\parameter	ρ	α	θ
diffuse	0.32 0.49	0.19 0.49	0.12 0.068
Jeffreys'	0.38 0.37	0.22 0.36	0.16 0.062

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

using the diffuse prior has much fatter tails and also shows some irregularities at the boundary of the stationary (invertible) parameter region, see also [2] and [17]. The posteriors using the Jeffreys' prior have a more regular behavior though.

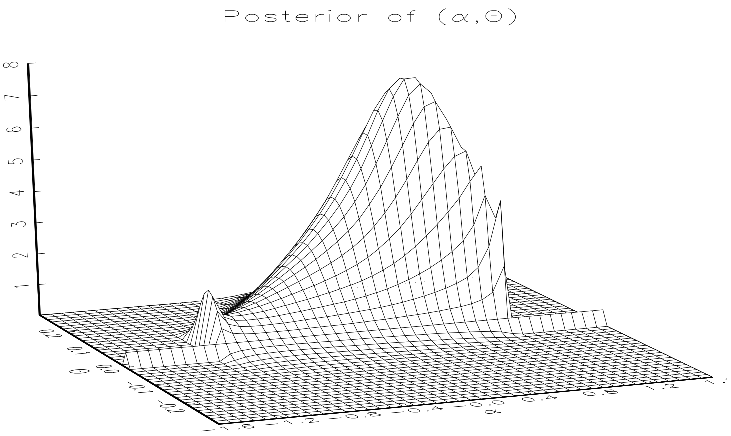

Figure 1: Bivariate posterior (α, θ), artificial time series, diffuse prior

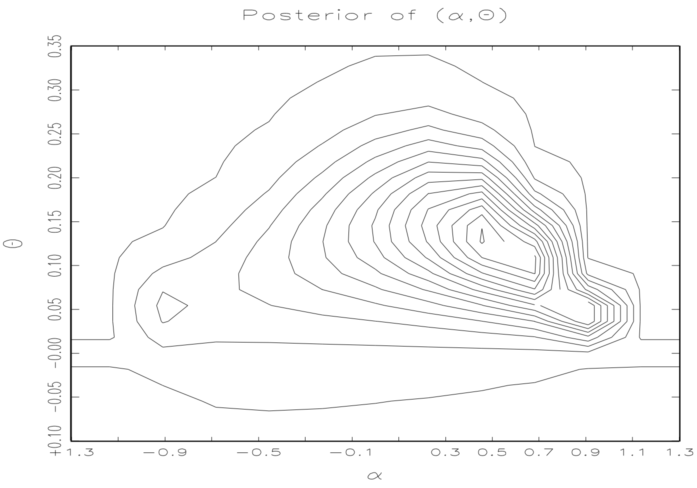

Figure 2: Contourlines bivariate posterior (α, θ), artificial time series, diffuse prior

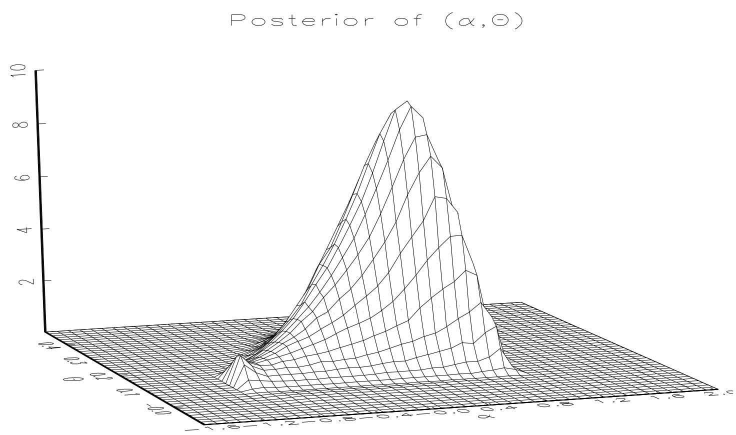

Figure 3: Bivariate posterior (α, θ), artificial time series, Jeffreys' prior

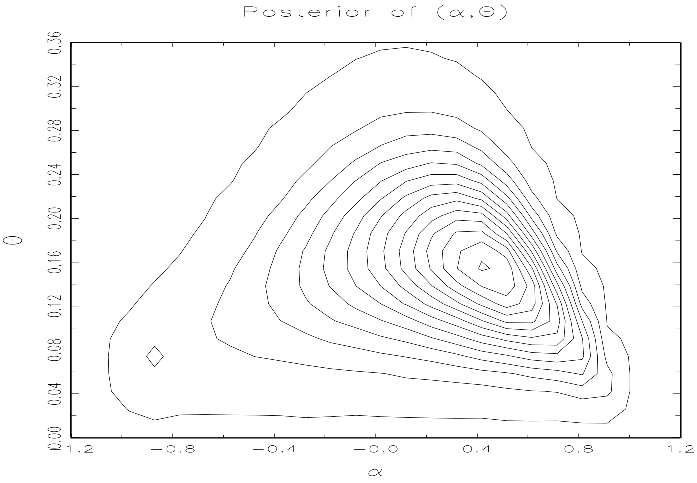

Figure 4: Contourlines bivariate posterior (α, θ), artificial time series, Jeffreys' prior

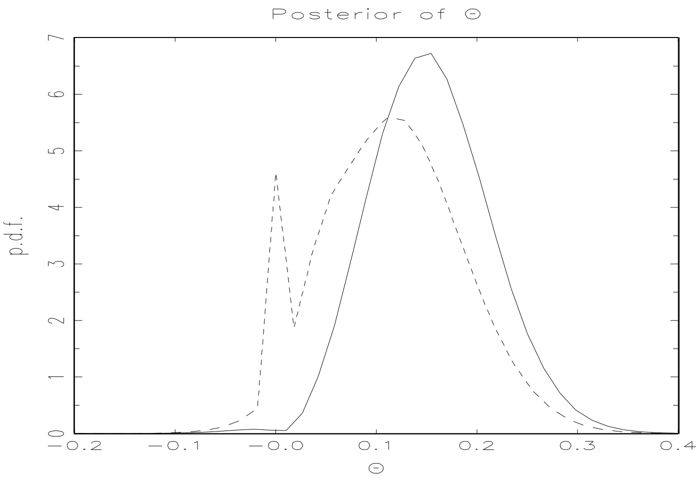

Figure 5: Marginal posterior θ, artificial time series, diffuse (- -) and Jeffreys' (—) prior

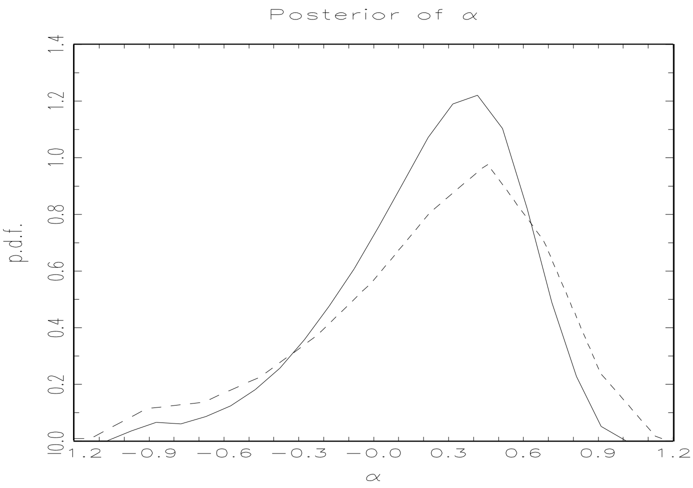

Figure 6: Marginal posterior α, artificial time series, diffuse (- -) and Jeffreys' (—) prior

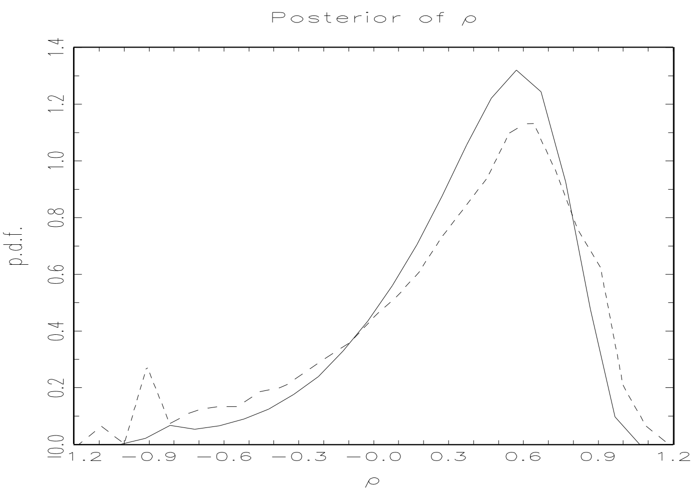

Figure 7: Marginal posterior ρ, artificial time series, diffuse (- -) and Jeffreys' (—) prior

5 Applications

To show the applicability of the derived theory and simulation procedures, we applied them to two data sets. The first data set consists of yearly observations of 14 Macro-Economic time series and is investigated by Nelson and Plosser, see [14]. This data set originally only contained data until 1970 but is later extended until 1988, see [18]. We use this extended data set. The second data set consists of monthly observations from January 1957 to April 1989 of the U.S. three month treasury bill rate and of interest rates having a maturity of ten years. We start by analyzing the first data set.

We model the (extended) Nelson-Plosser series using ARMA models with three ARMA parameters ($p + q = 3$) and these models contain a constant term and a linear trend. In section 3.3, we argued that for these kind of exogenous variables, the marginal posteriors of the ARMA paramters are not affected by the chosen specification of the exogenous variables. So, we choose the simplest model, the model with direct influence of the exogenous variables on the disturbances,

$$\rho(L)y_t = \alpha(L)(\mu + \delta t + \varepsilon_t), \tag{50}$$

where $\rho(z) = 1 - \sum_{i=1}^{p} \rho_i z^i$, $\alpha(z) = 1 - \sum_{i=1}^{q} \alpha_i z^i$, $p + q = 3$. Using the priors from equations (33) and (34) for the different parameters in these models, we constructed the posterior odds ratios from the weights resulting from the Importance Sampling Algorithm using the expression from equation (48). 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 and all contain a constant and linear time trend. So, posterior odds ratios are calculated for ARMA(3,0) (=AR(3)), ARMA(2,1), ARMA(1,2) and ARMA(0,3) (=MA(3)) models. The resulting ratios are listed in table 2.

The Posterior Odds Ratios from table 2 are quite surprising as for almost all series, an ARMA(2,1) model is more likely than an AR(3). For some series, the ARMA(2,1) is definitely preferable to an AR(3) given the value of the posterior odds, like Industrial Production, Employment, Unemployment, Consumer Price Index, Interest and the Standard and Poor 500, but for other series the odds indicate that both models are more or less equally likely. The ARMA(2,1) model can also be approximated by a high order AR model but an important difference between AR and MA lies in their consequences for the long run behavior of the series, MA components lead to exponentially decaying correlations which can still be significant over long horizons and can easily be misinterpreted as evidence for unit root behavior of the AR polynomial. So, it is interesting to investigate the influence of the MA parameters on the parameters reflecting the long run be-

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

Table 2: Posterior Odds Ratios Extended Nelson-Plosser series

havior of the analyzed series, like the unit root parameter, $\sum_{i=1}^{p} \rho_i$. We performed such an analysis and the results are listed in table 3, which contains the posterior means and standard deviations (below the posterior means) of the ARMA model that is preferred by the posterior odds ratios from table 2. Note that a MA(3) model is implausible for all series since this model leads to a very restricted type of long run behavior of the analyzed series.

For all series, except the Consumer Price Index (CPI), the MA parameter, α_1, has a positive correlation with the unit root parameter. The posterior mean of the unit root parameter of the ARMA(2,1) is therefore for all series, except CPI, smaller than the posterior mean of the unit root parameter of the AR(3) model. Depending on the size of the MA parameter, this decrease of the MA parameter can be quite large and it is most pronounced for the unemployment series, whose 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) but 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. These MA components also explain the long run memory in the first differences of these series, like inflation.

The parameter θ_{11}, see equation (41) 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 properly identified as the posterior mean lies more than two posterior standard deviations from 0. Exceptions are Industrial Production and

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

Table 3: Posterior Moments ARMA parameters Nelson-Plosser series

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

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

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

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

Velocity, for which an AR(3) is preferred. For Industrial Production holds that there is some posterior probability for zero values of θ_{11} leading to fat tailed behavior of the posteriors. This behavior is lost when we choose an ARMA(1,1) model which is sensible since the posterior mean of ρ_2 lies close to 0. In the resulting ARMA(1,1), α_1 is properly identified, see table 5. If the posteriors of an ARMA(2,1) model for velocity are calculated, the posterior of θ_{11} has a considerable amount of probability mass close to zero leading to fat tailed posteriors of the other parameters. This also indicates that an ARMA(2,1) is not the appropriate model for velocity which can also be concluded from the posterior odds ratios from table 2.

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

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

Table 5 shows that the summed posterior mean changes of ρ_1 and α_1 of the

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

Table 6: Posterior Odds Ratios Interest Rate Series

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

Table 7: Posterior Moments ARMA models interest rates

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 identification 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 parameter is similar though between both models, indicating that the long run behavior is not much affected by the deletion of ρ_2.

We also calculated the posteriors of the parameters of ARMA models for U.S. short and long term interest rates. Again the orders of the ARMA models, $p+q$, are supposed to equal 3. The difference with the model in equation (50) therefore solely results from the deterministic components as the model for the interest rates only contains a constant term while the model for the (extended) Nelson-Plosser data both contains a constant term and a linear time trend. To determine the favored univariate ARMA model for both interest rates we calculated the posterior odds ratios for all models with ARMA order, $p+q$, equal to 3. These posterior odds ratios are listed in table 6.

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

The posterior moments in table 7 show that the ARMA(1,2) model for long term interest rates has properly identified MA parameters as both identification parameters θ_{11} and θ_{22} have almost no probability mass at 0 as indicated by the posterior means and standard deviations of θ_{11} and θ_{22}. As the MA parameters of the ARMA model for the long term interest rates differ strongly from 0, the long run behavior of the long term interest rate will significantly differ from an standard random walk model. Furthermore since the AR parameter of the long term interest rate, ρ_1, lies close to 1, the long term interest rate can be very well characterized by an IMA(2) model.

6 Conclusions

A Bayesian analysis of parameters of ARMA models using Jeffreys' priors is shown to be applicable quite straightforwardly. This analysis directly allows for the construction of diagnostic parameters signaling the identification of the MA parameters and does not lead to a favor for nonidentified parameter values as is the case when a diffuse prior is used. For all applications, the Importance Sampling Algorithm converged rapidly. Quite surprisingly, in the applications we found that a lot of series which are traditionally modeled using AR contain strong MA components. These MA components can influence the long run parameters such that the use of MA components are important for forecasting purposes, see also [3].

In future work, we will extend the analysis to ARMA models containing seasonal lags and Vector ARMA models. As the Importance Sampling Algorithm performs very well, we will analyze the use of the Importance Function to generate parameters directly from the posterior using acception-rejection sampling. This will allow for model extensions as the posterior using Jeffreys' prior can then be used in a Gibbs Sampling framework.

References

[1] Chib, S. and E. Greenberg, 1994, Bayesian inference in regression models with ARMA(p,q) errors, *Journal of Econometrics,* 64, 183-206

[2] DeJong, D.N. and C.H. Whiteman, 1993, Estimating Moving Average Parameters : Classical Pileups and Bayesian Posteriors, *Journal of Business & Economic Statistics,* 11, 311-317

[3] Franses, P.H., and F. Kleibergen, 1995, Unit Roots in the Nelson-Plosser data : Do they matter for Forecasting?, *International Journal of Forecasting,* forthcoming

[4] Galbraith, J.W., and V. Zinde-Walsh, 1995, *Simple Estimation and Identification Techniques for general ARMA models,* Working Paper, McGill University, Canada

[5] Geweke, J., 1989, Bayesian Inference in Econometric Models using Monte Carlo Integration, *Econometrica,* 57, 1317-1339

[6] Geweke, J., 1989, Exact Predictive Densities for Linear Models with ARCH disturbances, *Journal of Econometrics,* 40, 63-86

[7] Harvey, A., 1981, *Time Series Models,* Philip Allan, London

[8] Kadane, J.B., 1974, The Role of Identification in Bayesian Theory, in: S.E. Fienberg and A. Zellner eds., *Studies in Bayesian Econometrics and Statistics,* North-Holland Publishing Co., Amsterdam

[9] Kleibergen, F., and H.K. van Dijk, 1994, On the Shape of the Likelihood/Posterior in Cointegration Models, *Econometric Theory,* 10, 514-551

[10] Kleibergen, F., and H.K. van Dijk, 1994, Direct Cointegration Testing in Error Correction Models, *Journal of Econometrics,* 63, 61-103

[11] Kleibergen, F., and H.K. van Dijk, 1994, *Bayesian Analysis of Simultaneous Equations Models using Noninformative Priors,* Tinbergen Institute Research Papers

[12] Kloek, T. and H.K. van Dijk, 1978, Bayesian Estimates of Equation System Parameters : An Application of Integration by Monte-Carlo, *Econometrica,* 46, 1-19

[13] Monahan, J.F., 1983, Fully Bayesian Analysis of ARMA Time Series Models, *Journal of Econometrics,* 21, 307-331

[14] Nelson, C.R. and C.I. Plosser, 1982, Trends and Random Walks in Macro Economic Time Series : Some Evidence and Implications, *Journal of Monetary Economics,* 10, 139-162

[15] Phillips, P.C.B., 1989, Partially Identified Econometric Models, *Econometric Theory,* 5, 181-240

[16] Phillips, P.C.B., 1991, To criticize the critics : An objective Bayesian analysis of Stochastic Trends, *Journal of Applied Econometrics,* 6, 333-364

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

[18] Schotman, P.C. and H.K. van Dijk, 1991, On Bayesian Routes to Unit Roots, *Journal of Applied Econometrics,* 6, 387-402

[19] Schotman, P.C., 1994, Priors for the AR(1) model : Parameterization Issues and Time Series Considerations, *Econometric Theory,* 10, 579-595

[20] Uhlig, H., 1994, On Jeffreys' Prior When using the Exact Likelihood Function, *Econometric Theory,* 10, 633-644

[21] Zellner, A., 1971, *Introduction to Bayesian Inference in Econometrics,* John Wiley, New York