Faculdade de Economia
da Universidade de Coimbra

Grupo de Estudos Monetários e Financeiros
(GEMF)
Av. Dias da Silva, 165 – 3004-512 COIMBRA,
PORTUGAL

gemf@fe.uc.pt
http://gemf.fe.uc.pt

ANTÓNIO ALBERTO SANTOS

MCMC, likelihood estimation and identifiability problems in DLM models

ESTUDOS DO GEMF
N.º 12 2010

PUBLICAÇÃO CO-FINANCIADA PELA FUNDAÇÃO PARA A CIÊNCIA E TECNOLOGIA

Impresso na Secção de Textos da FEUC
COIMBRA 2010

MCMC, likelihood estimation and identifiability problems in DLM models

António Alberto Ferreira Santos*

Faculty of Economics/Monetary and Financial Study Group (FEUC/GEMF),

University of Coimbra, Av. Dias da Silva, 165, 3004-512 Coimbra, Portugal

Abstract

In this article we deal with the identification problem within the Dynamic Linear Models family and show that using Bayesian estimation procedures we can deal better with these problems in comparison with the traditional Maximum Likelihood estimation approach. Using a Bayesian approach supported by Markov chain Monte Carlo techniques, we obtain the same results as the Maximum likelihood approach in the case of identifiable models, but in the case of non-identifiable models, we were able to estimate the parameters that are identifiable, as well as to pinpoint the troublesome parameters. Assuming a Bayesian approach, we also discuss the computational aspects, namely the ongoing discussion between single- versus multi-move samplers. Our aim is to give a clear example of the benefits of adopting a Bayesian approach to the estimation of high dimensional statistical models.

Key Words: Bayesian Statistics, DLM Models, Markov chain Monte Carlo, Maximum Likelihood, Model Identification.

*tel.: +351 790 526; e-mail: aasantos@fe.uc.pt

1 Introduction

There has been some debate related to the algorithms based on MCMC techniques used to estimate high dimensional models. One example is the Dynamic Linear Models (DLM). The parameters of these models can be estimated using Maximum Likelihood (ML) estimation. We can factorize the joint density of the sample and define the likelihood function. The likelihood function is far from being trivial because the parameters of the conditional distributions used to define the joint distributions are defined recursively, and hence a high nonlinear and dimensional objective function must be maximized. Even if this a well established fact, finding the maximum numerically is not too difficult in most cases. The difficult part is to establish whether the parameters are identified. Using the ML, sometimes, we find likelihoods with flat regions and consequently with high standard errors associated with the estimators. These flat regions are due to the phenomenon of non-identifiable parameters and therefore the failure of the maximization process. However, without analyzing carefully the structure of the model, we have no guidance concerning the parameters causing the problems.

Another way to estimate this type of models is to adopt a Bayesian approach and define the posterior distribution for the parameters of interest, which associated with a given loss function will give a point estimate. If in the previous case (Maximum Likelihood estimation) identifiability problems may be hard to detect, using MCMC techniques they will be detected quite easily. The main hint is the lack of mixing in the chains reveled by high levels of auto-correlation and high correlations between chains associated with different parameters. This aspect also leads to the ongoing discussion between the use of single- versus multi-move samplers.

2 DLM model and the identification problem

Let us consider the constant DLM. Adopting the notation of West and Harrison (1997), with the DLM represented by the quadruple $M = \{F, V, G, W\}$, we have

$$y_t = F\alpha_t + v_t, \quad v_t \sim N(0, V) \tag{1}$$

$$\alpha_{t+1} = G\alpha_t + w_t, \quad w_t \sim N(0, W). \tag{2}$$

This parametrization is not unique and an equivalent model can be considered, yielding the same forecasting density, $f(y_{t+1}|D_t)$, where $D_t = \{y_t, y_{t-1}, \ldots, y_1\}$. Different distributions to the states, filter and smoothing, are obtained but the original ones can be recovered through the distributions associated with the reparameterized model.

Now, let us consider a slightly different model in which we have

$$\begin{aligned} y_t &= F\alpha_t^* + v_t, \quad v_t \sim N(0, V) \\ \alpha_{t+1}^* &= G\alpha_t^* + w_t, \quad w_t \sim N(0, W) \end{aligned}$$

where, $\alpha_t^* = H^{-1}\alpha_t$. In this case it is easy to recover the initial parametrization. Consider $F^* = FH^{-1}$, $G^* = HGH^{-1}$ and $w_t^* = Hw_t$. Then

$$\begin{aligned} y_t &= F^*\alpha_t + v_t, \quad v_t \sim N(0, V) \\ \alpha_{t+1} &= G^*\alpha_t + w_t^*, \quad w_t \sim N(0, W^*). \end{aligned}$$

This DLM is defined by the quadruple $M^* = \{F^*, V, G^*, W^*\}$ with $W^* = HWH^\top$. The filter and smoothing distributions of α_t depend on F, G, and W, but also on H. In both parametrization the one-step ahead distribution does not depend on H. This is the essential distribution needed to define the likelihood function if the parameters are assumed unknown, and are to be estimated using available data. We are facing an indeterminacy as the likelihood function does not depend on H, and we cannot distinguish between M and M^*.

3 A simplified DLM model

Let us consider the simple model

$$\begin{aligned} y_t &= \beta\alpha_t + \zeta_t, \quad \zeta_t \sim N(0, \sigma_\zeta^2) \\ \alpha_{t+1} &= \phi\alpha_t + \eta_t, \quad \eta_t \sim N(0, \sigma_\eta^2). \end{aligned}$$

If no restrictions are imposed on the parameters, this model is not identifiable. By considering the vector of parameters $\theta = (\beta, \sigma_\eta^2, \phi, \sigma_\eta^2)$, and the information set $D_t = \{y_t, y_{t-1}, \ldots, y_0\}$, the joint density assumes the form

$$f(y_1, \ldots, y_n|\theta) = f(y_1|\theta) \prod_{t=2}^{n} f(y_t|D_{t-1}). \tag{3}$$

In this setting, it assumes the form of a multivariate normal density. However, the vector of means and variance-covariance matrix depend in a highly non-linear manner on the parameters. The main problem rises from the signal-to-noise ratio $\sigma_\eta^2/\sigma_\zeta^2$ and its relation with the parameter β. If one wants to estimate the signal-to-noise ratio, the parameter β can also be estimated. However, we cannot distinguish between σ_η^2 and σ_ζ^2 without considering any restriction on β.

To give a more precise idea of the complexity involved, let us consider $n = 2$. The log-likelihood function, assuming $\alpha_0 \sim N(0, 1)$, is given by,

$$\begin{aligned} l(\beta, \phi, \sigma_\zeta, \sigma_\eta|y_1, y_2) &= -\frac{1}{2}\ln(\phi^2 + \sigma_\zeta^2) - \frac{y_1^2}{2(\phi^2 + \sigma_\zeta^2)} \\ &- \frac{1}{2}\ln\left(\phi^2\left(\phi^2 + \sigma_\eta^2 - \frac{(\beta^3(\phi^2 + \sigma_\eta^2) + \beta\sigma_\zeta^2)^2}{\beta^2(\phi^2 + \sigma_\eta^2) + \sigma_\zeta^2}\right) + \sigma_\zeta^2\right) \\ &- \frac{\left(y_2 - \frac{\beta\phi(\beta^3(\phi^2+\sigma_\eta^2)+\beta\sigma_\zeta^2 y_1)}{\beta^2(\phi^2+\sigma_\eta^2)+\sigma_\zeta^2}\right)^2}{2\phi^2\left(\phi^2 + {\sigma_\eta}^2 - \frac{\left(\beta^3(\phi^2+{\sigma_\eta}^2)+\beta\,{\sigma_\zeta}^2\right)^2}{\beta^2(\phi^2+{\sigma_\eta}^2)+{\sigma_\zeta}^2}\right) + {\sigma_\zeta}^2}. \end{aligned}$$

Using this log-likelihood, it is not possible to identify all parameters. Only imposing restrictions on some parameters can the others be identified.

Using the general notation presented above, we have $F = \beta$, $V = \sigma_\zeta^2$, $G = \phi$ and $W = \sigma_\eta^2$, which yields the following re-parametrization, $F^* = \beta H^{-1}$, $G^* = \phi H H^{-1} = \phi$ and $W^* =$

$\sigma_\zeta^2 HH^\top$. With these results the problems of identification associated with the maximum likelihood estimation method become then much clear. The parameters β and σ_η^2, using just the time series $\{y_t\}_{t=1}^n$ without any constraints, cannot be identified. With an adequate rate of convergence, and σ_η tending to zero and β tend to infinity, or vice-versa, does not change the value of the likelihood function. Using just the likelihood function, it is not easy to find why the maximization algorithm fails. An inadequate algorithm to the problem or lack of identification of the model can be a possible explanation for the problem.

4 MCMC and Bayesian estimation

An alternative to ML estimation is to use MCMC techniques within a Bayesian framework. The aim is to approximate the joint posterior distribution of the unknowns, states and parameters. Parameters point estimates are obtained through the joint marginal posterior distribution of the parameters associated with a given loss function. By considering a quadratic loss function, the point estimates are given by the vector of means of the respective posterior distribution. Here, we are dealing basically with the Gaussian distribution, and if we assume a diffuse prior information, all relevant information is given by the likelihood. In the Bayesian framework the posterior distribution of the parameters is proportional to the likelihood function, and so the mode, the maximum of the likelihood function, coincides with the mean of the posterior distribution. Essentially, we obtain the same point estimates. However, the output of the two procedures are substantially different and, even for the assumed loss function, much more can be done with the output of a Bayesian analysis.

Let us consider the model above with $n = 500$. In this case the posterior distribution of interest is of dimension 504, $f(\alpha_1, \ldots, \alpha_{500}, \beta, \sigma_\zeta, \phi, \sigma_\eta | y_1, \ldots, y_{500})$. The calculations needed are far from trivial. The main idea is use simulations to approximate it numerically. This task is not an easy one due to two main reasons: first, the dimension of the distribution; second, the mixing of mean and variance components, which results in a mixing of Gaussian and

Gamma distributions for the unknowns. Plain vanilla simulation methods that give independent and identically random observations, like the inversion method and the accepting-rejection method, will fail in this environment. The way forward is to divide the main problem into simpler problems. This introduces some dependence in the random observations but allows the treatment of more complex models. By analyzing the model, we can explore the two separate nature of the unknowns, states and parameters. We are going to simulate the states conditioned on the parameters and the parameters on the states. Later we break further these problems. Simulating from the parameters given the states is a trivial task. More difficult is to simulate from the states given the parameters.

In phenomenons like financial volatility evolution and economic time series forecasting, using income or inflation, where this kind of model reveals its usefulness, it is common to find processes with high persistence, with a value for ϕ near 1, where ϕ acts like a parameter of correlation. Using the benchmark of a bivariate Gaussian distribution, as the correlation coefficient becomes near one in absolute value, more stretched are the contours associated with the distribution, and the chain can move less freely. With only small steps allowed, the chains present a high level of autocorrelation and inferences become less reliable. What has been proposed is to consider the problem in higher dimensions. There was been research trying to establish the differences in performance between the so-called single- and multi-move samplers. In some problems the chains associated with a single-move sampler present a high degree of autocorrelation making inferences very difficult to implement.

What we analyze here is the separation between the high degree of autocorrelation of the chains related to some kind of identifiability problem and the one related to the nature of the problem, namely the high memory of the subjacent processes. We are going to exemplify these facts by estimating two models using ML and MCMC. In the first, not all the parameters are identifiable, in the second, all the parameters are identifiable. In the first, we cannot obtain ML estimates and the ones resulting from MCMC must be used with care, as it is difficult to accept that all chains of interest have converged. With the second model, ML estimates can be found

and with a large sample, with diffuse prior distribution for the parameters, Bayesian estimates will be near the ML estimates.

5 MCMC algorithms

Simulating from the parameters conditioned on the states is a trivial task. Let us consider the original model (possible non-identifiable) and let us define the samplers. Assuming a given vector of states $\alpha = \{\alpha_1, \ldots, \alpha_n\}$ and a vector of observations $y = \{y_1, \ldots, y_n\}$, the posterior distribution of β can be written as

$$\beta|\alpha, y, \sigma_\zeta \sim N\left(\frac{\sum_{t=1}^n y_t\alpha_t}{\sum_{t=1}^n \alpha_t^2}, \frac{\sigma_\zeta^2}{\sum_{t=1}^n \alpha_t^2}\right).$$

By considering

$$S_\beta = \sum_{t=1}^n (y_t - \beta\alpha_t)^2,$$

we can establish the posterior distribution

$$\sigma_\zeta^2|y, \alpha, \beta \sim IG\left(\frac{n}{2}, \frac{S_\beta}{2}\right).$$

To the parameters associated with the states, the procedures are identical, and for ϕ we have

$$\phi|\alpha, \sigma_\eta \sim N\left(\frac{\sum_{t=2}^n \alpha_t\alpha_{t-1}}{\sum_{t=2}^n \alpha_{t-1}^2}, \frac{\sigma_\eta^2}{\sum_{t=2}^n \alpha_{t-1}^2}\right).$$

By considering

$$S_\phi = \sum_{t=2}^n (\alpha_t - \phi\alpha_{t-1})^2,$$

we can establish the posterior distribution

$$\sigma_\eta^2|\alpha, \phi \sim IG\left(\frac{n}{2}, \frac{S_\phi}{2}\right).$$

Here for all the parameters we are assuming a priori diffuse distributions.

Simulating from the states given the parameters is much more difficult. The posterior distribution of the states given the parameters is a multivariate Gaussian distribution. If we

knew the vector of means and the variance-covariance matrix, using a Cholesky decomposition, it would not be too difficult to generate draws from the corresponding distribution. The problem here is the definition of the vector of means and variance-covariance matrix. These are defined recursively and take into account the information associated with previous observations, which makes difficult to express them analytically.

5.1 The single move sampler

One way to move forward is to break the n-dimensional problem into n problems of dimension one. This can be done by considering one dimensional conditional distribution (Gibbs Sampling) for the states with the distribution

$$\alpha_t|y_t, \alpha_{t-1}, \alpha_{t+1}, \beta, \sigma_\zeta, \phi, \sigma_\eta \sim N(m_t, s_t^2),$$

where

$$m_t = \frac{\beta y_t \sigma_\eta^2 + \phi \sigma_\zeta^2(\alpha_{t-1} + \alpha_{t+1})}{(1+\phi^2)\sigma_\zeta^2 + \beta^2\sigma_\eta^2}$$

and

$$s_t^2 = \frac{\sigma_\eta^2 \sigma_\zeta^2}{(1+\phi^2)\sigma_\zeta^2 + \beta^2\sigma_\eta^2}.$$

These results can be accomplished using two different strategies. Consider the generic expression,

$$f = -\frac{(y_t - \beta\alpha_t)^2}{2\sigma_\zeta^2} - \frac{(\alpha_t - \phi\alpha_{t-1})^2}{2\sigma_\eta^2} - \frac{(\alpha_{t+1} - \phi\alpha_t)^2}{2\sigma_\eta^2},$$

which is proportional to the logarithm of the conditional density given above. Mean and variance parameters can be defined by expanding the expression, collecting the factors depending on α_t^2 and α_t and arranging the expression as $(-1/2)a(\alpha_t - b)^2 + c$, where a is the precision, b the mean and c a constant that is absorbed by the constant of proportionality needed to define the density function. Another strategy is to find the maximum by deriving with respect to α_t, equalize to zero and resolve in order to α_t. Evaluating $-1/(d^2f/d\alpha_t^2)$ at the point obtained

previously, at the maximum of the function, and at the mean of the conditional distribution, we obtain exactly s_t^2.

5.2 Multi-move samplers

The recursive nature of the process can be used by applying the Kalman filter to define successive posterior distributions and then move backward to define the smoothing distribution, which will be used within the so-called Simulation Smoother. Different algorithms have been proposed to perform this task in the most efficient manner. However, by taking into account the simple fact that only Gaussian distributions are here considered, the main issue is to define the appropriate vector of means and variance-covariance matrix.

5.2.1 The multivariate approach

The procedure defined above can be extended through the maximum of the function f in the bivariate case or even in the multivariate case. An analytical expression to the vector of means and variance-covariance matrix is available. However, these would be too complicated for being operational in a computational sense. Let us take the example of the distribution

$$\alpha_1, \alpha_2 | \alpha_3, y_1, y_2, \beta, \sigma_\zeta, \phi, \sigma_\eta \sim N(\mu, \Sigma)$$

where

$$\mu = \begin{bmatrix} \dfrac{\phi^2\alpha_3\sigma_\zeta^4 + \beta\left(\left(1+\phi^2\right)y_1 + \phi y_2\right)\sigma_\zeta^2\sigma_\eta^2 + \beta^3 y_1\sigma_\eta^4}{\phi^4\sigma_\zeta^4 + \beta^2\left(1+2\phi^2\right)\sigma_\zeta^2\sigma_\eta^2 + \beta^4\sigma_\eta^4} \\ \\ \dfrac{\phi^3\alpha_3\sigma_\zeta^4 + \beta\phi\left(y_1 + \phi y_2 + \beta\alpha_3\right)\sigma_\zeta^2\sigma_\eta^2 + \beta^3 y_2\sigma_\eta^4}{\phi^4\sigma_\zeta^4 + \beta^2\ \left(1+2\phi^2\right)\sigma_\zeta^2\sigma_\eta^2 + \beta^4\ \sigma_\eta^4} \end{bmatrix}.$$

We could continue to obtain the distribution of

$$\alpha_1, \ldots, \alpha_n | D_n, \alpha_{n+1}, \beta, \sigma_\zeta, \phi, \sigma_\eta \sim N(\mu, \Sigma)$$

with the parameters μ and Σ defined in an analytical manner, but this would lead to an unmanageable expression when n assumes a large value. A better computational solution would

be to evaluate the vector of means and variance-covariance matrix numerically. As the vector of derivatives is easy to define, the Hessian (H) matrix is tridiagonal and relates to the variance-covariance matrix as $\Sigma = -H^{-1}$, passing an adaptation period, the number of iterations needed to find the maximum (vector of means) is small. The algorithm becomes very efficient for simulating from the joint smooth distribution of the states given the parameters.

At each iteration, for the vector of states, sub-iterations are needed to find the vector of means, which are associated with the definition of the gradient. A vector of dimension n must be evaluated at each sub-iteration. This is not too complicated as the structure of the vector is very simple, and with current computational power, this does not represent a great computational burden. At each sub-iteration, the inverse of the hessian matrix need not to be recomputed as it only depends on the parameters of the model. Being a tridiagonal matrix, symmetric and negative definite, by considering the Cholesky decomposition, $-H = C^\top C$, with C a bidiagonal matrix, calculating the inverse of $-H$ as the product of the inverse of two bidiagonal matrices facilitates greatly the computations.

5.2.2 The simulation smoother

Several simulation techniques have been proposed to simulate from high dimensional smoothing distributions within a DLM. As mentioned above, it corresponds to simulate from a multivariate normal distribution with a given vector of means and variance-covariance matrix, then as this problem is standard in statistics the main difference between algorithms is the efficiency associated with the definition of such draws. Using the statistical structure of the model, different algorithms were proposed and one is presented in West and Harrison (1997), p. 570. Two widely referred algorithms were proposed by Frühwirth-Schnatter (1994) and Carter and Kohn (1994). Another common algorithm is the one proposed in de Jong and Shephard (1995). These algorithms are referenced in Koopman and Durbin (2000), Durbin and Koopman (2001) and Durbin (2004). In particular the one proposed in Durbin (2004), being very efficient, is very easy to codify.

Following closely the notation in Durbin (2004) and making the necessary adaptations for the constant DLM presented in section 2, we present the Kalman Filter, the States Smoother and the Simulation Smoother algorithms. Consider $a_{t+1} = E(\alpha_{t+1}|D_t)$ and $P_{t+1} = Var(\alpha_{t+1}|D_t)$, that can be obtained using the following recursions,

$$\begin{aligned} u_t &= y_t - Fa_t \\ Z_t &= FP_tF^\top + V \\ K_t &= GP_tF^{-1} \\ L_t &= G - K_tF \\ a_{t+1} &= Ga_t + K_tu_t \\ P_{t+1} &= GP_tL_t^\top + W, \quad t = 1, \ldots, n. \end{aligned}$$

With the States Smoother, the aim is to calculate $\hat{\alpha}_t = E(\alpha_t|D_n)$ and $\hat{p}_t = Var(\alpha_t|D_n)$ by the backwards recursion,

$$\begin{aligned} r_{t-1} &= F^\top Z_t^{-1}u_t + L_t^\top r_t \\ N_{t-1} &= F^\top Z_t^{-1}F + L_t^\top N_tL_t \\ \hat{\alpha}_t &= a_t + P_tr_{t-1} \\ \hat{p}_t &= P_t - P_tN_{t-1}P_t, \quad t = n, \ldots, 1 \end{aligned}$$

with $r_n = 0$ and $N_n = 0$.

The Simulation Smoother can be implemented using the following steps:

1. Obtain random draws for v_t^* and w_t^* from the distributions $N(0, V)$ and $N(0, W)$, respectively, and, using the parameters of the model (1)-(2), obtain for $t = 1, \ldots, n$, the simulated values of α_t^* and y_t^*.
2. Compute for $t = 1, \ldots, n$, $\hat{\alpha}_t = E(\alpha_t|D_n)$ and $\hat{\alpha}_t^* = E(\alpha_t^*|D_n^*)$.
3. Take a draw from the distribution of $\alpha_t|D_n$, as

$$\tilde{\alpha}_t = \hat{\alpha}_t - \hat{\alpha}_t^* + \alpha_t^*,$$

for $t = 1, \ldots, n$.

As well known, when MCMC methods are applied to state space models, the most difficult part is to simulate the states. We present the single-move and a multi-move sampler to guarantee that the results we aim for are not influenced by the algorithm used to simulate the states of the DLM. We argue that MCMC techniques within the Bayesian Statistics are better equipped to deal with models with possible non-identifiable parameters than standard ML techniques. Using the DLM, we show in the next section that this is independent of the algorithm used to simulate the states.

6 Empirical application

We use simulated data to illustrate the differences in the properties of the chains in the case of an identified model versus a non-identified one. The procedures implemented were codified using the statistical package `R`. New code was written to implement the estimation procedures. The output of the Bayesian analysis is dealt with using the library `coda` of the referred statistical package. The code and data are available upon request.

When maximum likelihood methods are used, applied to a non-identifiable model, most probably the algorithm used in the maximization process gives a meaningless output, sometimes without giving any clue for the reason of such result. There are many reasons for a numerical maximization algorithm not to converge to a unique optimal solution. One of them is the lack of identification of all parameters.

Using Bayesian estimation and MCMC techniques, we obtain an output from which the parameter estimates can be obtained. The chains may or may not converge. However, an output is always given. The lack of convergence may be a symptom of misspecification, namely, by the use of a model with non-identifiable parameters. What can be found here is that using these more sophisticated techniques, checks for identification can be performed, while at the same time we can obtain reliable estimates to the identifiable parameters and pinpoint the parameters causing

the identification problem. A second step could be followed, fixing or imposing restrictions to those parameters as a way of specifying an identifiable model. In contrast to ML techniques, an output is obtained and this can be helpful in the model redefinition if necessary.

One illustration uses a simulated series with $n = 500$ observations. This allow us to compare the parameters estimates with the true values, considering an identifiable model and a non-identifiable one. Let us use the following model,

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, 0.3^2) \\ \alpha_{t+1} &= 0.97\alpha_t + \eta_t, \quad \eta_t \sim N(0, 0.15^2). \end{aligned}$$

We consider first the vector of parameters $(\beta, \sigma_\varepsilon, \phi, \sigma_\eta) = (1.0, 0.3, 0.97, 0.15)$ associated with the non-identifiable model, and second, $(\sigma_\varepsilon, \phi, \sigma_\eta) = (0.3, 0.97, 0.15)$ for the identifiable one. With the first parametrization the sample does not contain sufficient information for the estimation of all the parameters.

In a non-identifiable model the parameters in the likelihood function are connected in a manner that they cannot be distinguished, different combinations lead to the same value of the log-likelihood. Using Bayesian Statistics associated with MCMC techniques has two main advantages: it identifies the troublesome parameters and deals smoothly with the estimation issue.

To illustrate these aspects, let us simulate from the model above and then estimate the parameters using ML and MCMC. First, we ignore the identifiability problem and try to estimate all four parameters. Using ML the maximization procedure, it ends suddenly without giving any explicit clue. Depicted in Figure 1 are the problems associated with the chains of β and σ_η, which do not converge. However, in the case of parameters that are causing no problem, meaningful estimates are obtained through chains that seem to converge. Using MCMC techniques, in both cases, we use simulated chains with 10000 draws.

After detecting the problem of non-identifiable parameters, we have to proceed by redefining the model or imposing more informative prior distributions. It was clear from above that the

Figure 1: Chains and respective densities for the parameters β (beta), σ_ε (sigma_u), ϕ (phi) and σ_η (sigma_v) of the DLM

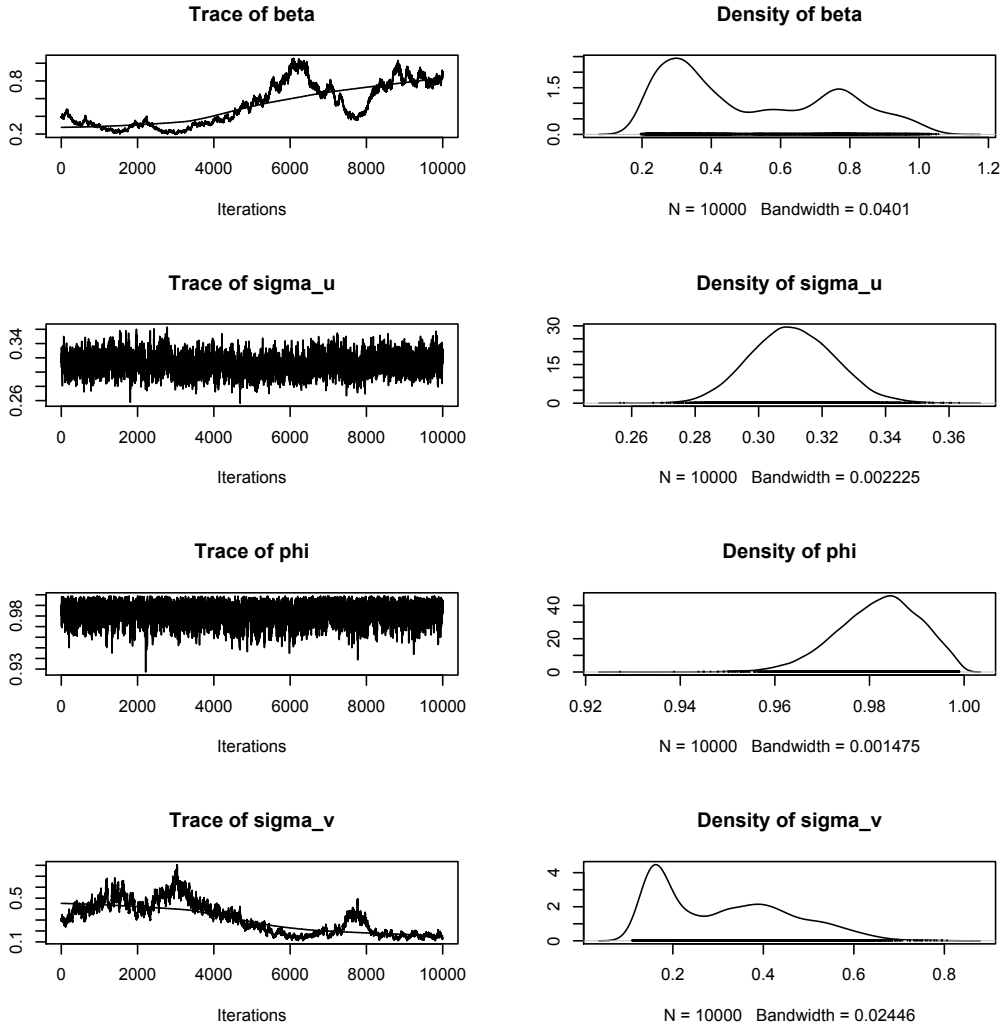

Figure 2: Chains and respective densities for the parameters σ_ε (sigma_u), ϕ (phi) and σ_η (sigma_v), of the DLM, in the case of an identifiable model

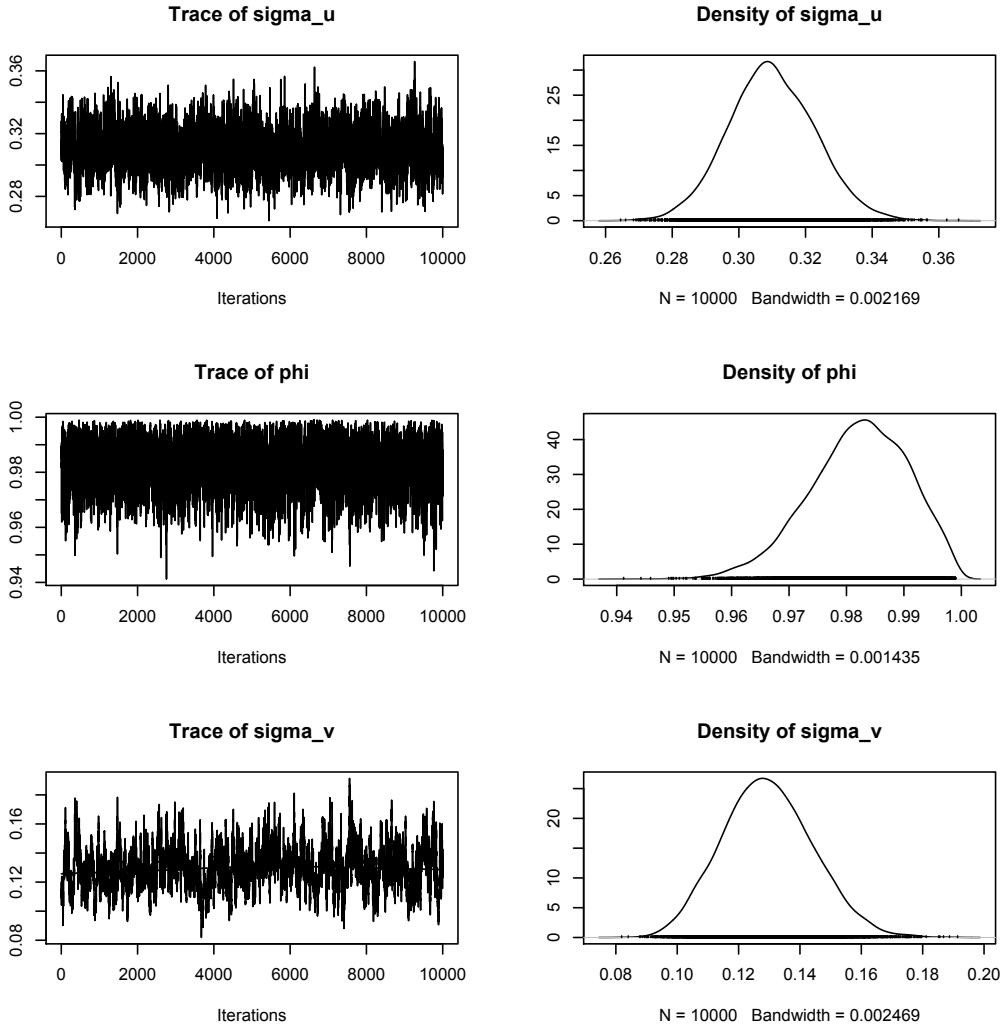

Table 1: Summary of the parameters estimates for the non-identifiable model. Note: Due to the problems posed by the parameters β and σ_η, it is not possible to estimate the standard-error of such parameters. As a result the ESS is extremely low.

	Mean-S	SE-S	ESS -S
β	0.5191	na	2
σ_ε	0.3107	5.64e-4	56
ϕ	0.9819	2.64e-4	1337
σ_η	0.3105	na	8

troublesome parameters are β and σ_η. Using the results from section 2, it is clear that was established that only β and the signal-to-noise parameter $\sigma_\eta^2/\sigma_\varepsilon^2$ can be estimated, not all the three parameters. However, by fixing $\beta = 1$, we can see, in Figure 2, that using MCMC the chains associated with the three parameters converge and meaningful results are obtained.

The differences between single- versus multi-move samplers will not be developed further. Applying the multi-move sampler to the identified model, no substantial differences in the results were obtained when compared with the single-move sampler. Through a graphical analysis (trace and density), not depicted here, we can see no difference. We present the estimates using both algorithms and the difference that is noticed is related to the Effective Sample Size (ESS). The ESS is calculated using `coda` and give us an idea of the efficiency of the simulations, when compared with a sample of n independent and identically distributed random variables, which corresponds an ESS of n. Lower values for the ESS represent more inefficient algorithms.

In Table 1, only the output associated with the Bayesian analysis is presented. It was not possible to obtain ML estimates due to the lack of identification of the parameters. In Table 1, as in Figure 1, the troublesome parameters are clearly identified. As the chains associated with

Table 2: Summary of the parameters estimates for the identifiable model. Estimation used MCMC with single-move sampler (S), multi-move sampler (M) and Maximum Likelihood (ML). Notes: Estimates using MCMC are presented at Mean- with the respective Standard Error (SE -) and Effective Sample Size (ESS-). To compare the estimates, the last two columns are related to Maximum Likelihood estimation.

	Mean-S	SE-S	ESS -S	Mean-M	SE-M	ESS-M	ML	SE -ML
σ_ε	0.309	5.06e-4	1029	0.310	4.06e-4	1118	0.310	1.75e-4
ϕ	0.982	2.45e-4	1659	0.982	2.17e-4	1553	0.982	8.29e-4
σ_η	0.1295	1.07e-4	288	0.129	8.59e-4	357	0.128	2.46e-4

β and σ_η did not converge, it was not possible to calculate the standard-error of the estimates. Then looking at the ESS of these parameters, low values are obtained, a sign of chains associated with the troublesome parameters did not converge.

In Table 2 we present the summary of the estimation using MCMC (single- and multi-move) and ML. As the model is identified, meaningful results were obtained for the three parameters. The parameters estimates are similar using all the three procedures, exactly what we were expecting, which reinforces the usefulness to consider different estimation techniques and not be restricted only to ML or OLS, which may yield meaningless results.

7 Conclusion

To Bayesian estimation, the lack of identification is not such a restrictive problem as in other estimation procedures. With the likelihood assuming very complicated forms, it can be difficult to identify the parameters that potentially or effectively cause identification problems. By maximizing the likelihood function associated with an unidentified model, it is likely that no

parameter estimate will be obtained. With a Bayesian estimation procedure based on MCMC techniques, we can identify the troublesome parameters and deal with unidentified models using convenient informative priors or any other suitable restriction. In this article, we present a simple model, in which, theoretically, we can establish the constraints that yield an identified model. Unidentified parameters lead to Markov chains that do not converge. The chains associated with non-identifiable parameters are necessarily highly correlated as a consequence of the lack of convergence. Bayesian Statistics associated with MCMC techniques is a very robust way to deal with non-identifiable models. An erroneous idea sometimes found in the literature is that MCMC methods are too complicated to be applied widely. We try to demonstrate that this is not necessarily true. With the DLM, all the distributions associated with the states are gaussian, which is straightforward to simulate from. In other cases, for example, stochastic volatility models, the gaussian distribution may constitute a meaningful approximation. The outputs are considerable more richer than in more common estimation procedures, as we demonstrated in the case of the identification problem. The procedures were implemented using an interpretation language like `R`, which demonstrates the feasibility of such approach.

References

Carter, C. K., and R. Kohn (1994), "On Gibbs sampling for state space models," *Biometrika*, 81, 541–553.

de Jong, P., and N. Shephard (1995), "The Simulation Smoother for Time Series Models," *Biometrika*, 82, 339–350.

Durbin, J. (2004), "Introduction to state space time series analysis," in *State Space and Unobserved Component Models*, ed. by A. Harvey, S. J. Koopman, and N. Shephard, pp. 3–25. Cambridge University Press.

Durbin, J., and S. J. Koopman (2001), *Time Series Analysis by State Space Methods*. Oxford University Press, Oxford.

Frühwirth-Schnatter, S. (1994), "Data Augmentation and Dynamic Linear Models," *Journal of Time Series Analysis*, 15, 183–202.

Koopman, S. J., and J. Durbin (2000), "Fast Filtering and Smoothing for Multivariate State Space Models," *Journal of Time Series Analysis*, 21 281–296.

West, M., and P. J. Harrison (1997), *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York.

ESTUDOS DO G.E.M.F.

(Available on-line at http://gemf.fe.uc.pt)

2010-12 *MCMC, likelihood estimation and identifiability problems in DLM models*
- António Alberto Santos

2010-11 *Regional growth in Portugal: assessing the contribution of earnings and education inequality*
- Adelaide Duarte & Marta Simões

2010-10 *Business Demography Dynamics in Portugal: A Semi-Parametric Survival Analysis*
- Alcina Nunes & Elsa Sarmento

2010-09 *Business Demography Dynamics in Portugal: A Non-Parametric Survival Analysis*
- Alcina Nunes & Elsa Sarmento

2010-08 *The impact of EU integration on the Portuguese distribution of employees' earnings*
- João A. S. Andrade, Adelaide P. S. Duarte & Marta C. N. Simões

2010-07 *Fiscal sustainability and the accuracy of macroeconomic forecasts: do supranational forecasts rather than government forecasts make a difference?*
- Carlos Fonseca Marinheiro

2010-06 *Estimation of Risk-Neutral Density Surfaces*
- A. M. Monteiro, R. H. Tütüncü & L. N. Vicente

2010-05 *Productivity, wages, and the returns to firm-provided training: who is grabbing the biggest share?*
- Ana Sofia Lopes & Paulino Teixeira

2010-04 *Health Status Determinants in the OECD Countries. A Panel Data Approach with Endogenous Regressors*
- Ana Poças & Elias Soukiazis

2010-03 *Employment, exchange rates and labour market rigidity*
- Fernando Alexandre, Pedro Bação, João Cerejeira & Miguel Portela

2010-02 *Slip Sliding Away: Further Union Decline in Germany and Britain*
- John T. Addison, Alex Bryson, Paulino Teixeira & André Pahnke

2010-01 *The Demand for Excess Reserves in the Euro Area and the Impact of the Current Credit Crisis*
- Fátima Teresa Sol Murta & Ana Margarida Garcia

2009-16 *The performance of the European Stock Markets: a time-varying Sharpe ratio approach*
- José A. Soares da Fonseca

2009-15 *Exchange Rate Mean Reversion within a Target Zone: Evidence from a Country on the Periphery of the ERM*
- António Portugal Duarte, João Sousa Andrade & Adelaide Duarte

2009-14 *The Extent of Collective Bargaining and Workplace Representation: Transitions between States and their Determinants. A Comparative Analysis of Germany and Great Britain*
- John T. Addison, Alex Bryson, Paulino Teixeira, André Pahnke & Lutz Bellmann

2009-13 *How well the balance-of- payments constraint approach explains the Portuguese growth performance. Empirical evidence for the 1965-2008 period*
- Micaela Antunes & Elias Soukiazis

2009-12 *Atypical Work: Who Gets It, and Where Does It Lead? Some U.S. Evidence Using the NLSY79*
- John T. Addison, Chad Cotti & Christopher J. Surfield

2009-11 *The PIGS, does the Group Exist? An empirical macroeconomic analysis based on the Okun Law*
- João Sousa Andrade

2009-10 *A Política Monetária do BCE. Uma estratégia original para a estabilidade nominal*
- João Sousa Andrade

2009-09 *Wage Dispersion in a Partially Unionized Labor Force*
- John T. Addison, Ralph W. Bailey & W. Stanley Siebert

2009-08 *Employment and exchange rates: the role of openness and technology*
- Fernando Alexandre, Pedro Bação, João Cerejeira & Miguel Portela

2009-07 *Channels of transmission of inequality to growth: A survey of the theory and evidence from a Portuguese perspective*
- Adelaide Duarte & Marta Simões

2009-06 *No Deep Pockets: Some stylized results on firms' financial constraints*
- Filipe Silva & Carlos Carreira

2009-05 *Aggregate and sector-specific exchange rate indexes for the Portuguese economy*
- Fernando Alexandre, Pedro Bação, João Cerejeira & Miguel Portela

2009-04 *Rent Seeking at Plant Level: An Application of the Card-De La Rica Tenure Model to Workers in German Works Councils*
- John T. Addison, Paulino Teixeira & Thomas Zwick

2009-03 *Unobserved Worker Ability, Firm Heterogeneity, and the Returns to Schooling and Training*
- Ana Sofia Lopes & Paulino Teixeira

2009-02 *Worker Directors: A German Product that Didn't Export?*
- John T. Addison & Claus Schnabel

2009-01 *Fiscal and Monetary Policies in a Keynesian Stock-flow Consistent Model*
- Edwin Le Heron

2008-08 *Uniform Price Market and Behaviour Pattern: What does the Iberian Electricity Market Point Out*
- Vítor Marques, Isabel Soares & Adelino Fortunato

2008-07 *The partial adjustment factors of FTSE 100 stock index and stock index futures: The informational impact of electronic trading systems*
- Helder M. C. V. Sebastião

2008-06 *Water Losses and Hydrographical Regions Influence on the Cost Structure of the Portuguese Water Industry*
- Rita Martins, Fernando Coelho& Adelino Fortunato

2008-05 *The Shadow of Death: Analysing the Pre-Exit Productivity of Portuguese Manufacturing Firms*
- Carlos Carreira & Paulino Teixeira

2008-04 *A Note on the Determinants and Consequences of Outsourcing Using German Data*
- John T. Addison, Lutz Bellmann, André Pahnke & Paulino Teixeira

2008-03 *Exchange Rate and Interest Rate Volatility in a Target Zone: The Portuguese Case*
- António Portugal Duarte, João Sousa Andrade & Adelaide Duarte

2008-02 *Taylor-type rules versus optimal policy in a Markov-switching economy*
- Fernando Alexandre, Pedro Bação & Vasco Gabriel

2008-01 *Entry and exit as a source of aggregate productivity growth in two alternative technological regimes*
- Carlos Carreira & Paulino Teixeira

A série Estudos do GEMF foi iniciada em 1996.