

# Data Cloning: Maximum Likelihood Estimation of DSGE Models

Luiz Gustavo Cassilatti Furlani      Márcio Poletti Laurini  
Marcelo Savino Portugal

July 14, 2013

## Abstract

In this article, we present evidence supporting the use of the data cloning method of Lele et al. (2007) and Lele et al. (2010) for maximum likelihood estimation of DSGE models. In the data cloning method, maximum likelihood estimators are obtained as the limit of Bayesian estimators, taking advantage of existing computational estimation methods based on Markov Chain Monte Carlo. We show the desirable properties of this methodology based on some experiments. In the first experiment, we estimate the Cash-In-Advance model of Schorfheide (2000) and compare the results with the most important methods of numerical optimization used in maximum likelihood estimation. The data cloning estimators show low variability, unlike the maximum likelihood estimators, suggesting greater robustness of the former to the problem of initial conditions. In the second experiment, we show that the data cloning estimators also present good results in terms of bias and RMSE. The proposed procedures also allow to check the econometric identification of parameters and to determine the appropriate number of clones.

## 1 Introduction

Dynamic Stochastic General Equilibrium (DSGE) models have raised growing interest in the field of macroeconomics. Due to computational advancements, these models have become increasingly more complex and accurate, allowing for an in-depth analysis of the relationships between real and nominal economic variables.

Briefly, a model of this class can be described by the following equations: a Taylor rule, i.e., a monetary policy rule in which the interest rate is the central

bank's instrument for invigorating the economy; a dynamic IS curve, with expectations; and a New-Keynesian Phillips curve, which also considers the expectations of individuals. From the monetary policy rule and the dynamic IS curve, we obtain aggregate demand. Aggregate supply is given by the New-Keynesian Phillips curve. Economic equilibrium is achieved by the relationship between aggregate supply and aggregate demand curves. The estimation of these models is directly based on the solution of dynamic equilibrium, involving steps to find this solution and subsequent parameter estimation based on the representation of the equilibrium of these models.

DSGE models, albeit highly intuitive, are problematic in at least two ways, according to Canova (2007). First, this type of model provides only an approximation to the data generating process, since the vector of the structural parameters generally has a low dimension and thus strong restrictions are imposed in the short and in the long run. Secondly, the number of exogenous variables is usually smaller than the number of endogenous variables, and so the covariance matrix of the endogenous variables becomes singular. Because of these characteristics, the estimation and testing of DSGE models by traditional methods (e.g., maximum likelihood) is very complex. In addition, stochastic singularity prevents numerical routines based on the Hessian from working properly and consequently the maximum of the objective function from being reached, as discussed in Ingram et al. (1994) and Ireland (2004).

A possible solution to these problems is to impose artificial measurement errors for some of the endogenous variables, making the covariance matrix nonsingular, as used in Bencivenga (1992), DeJong et al. (2000a), DeJong et al. (2000b), Ireland (2001), Kim (2000) and Schorfheide (2000). As discussed in Ireland (2004), this strategy allows identifying additional shocks to those observed in real business cycle models and determining how important these shocks are to aggregate fluctuations. However, this strategy has some drawbacks: its implementation requires extending the theoretical model by using ad hoc choices. If these shocks are not important in the theoretical model, their use can introduce noise in the estimation process, leading to incorrect specification, reducing the efficiency of the estimates and the power of hypothesis tests on the parameters.

Another difficulty associated with the estimation of DSGE models using likelihood-based methods is the shape of the likelihood function. The likelihood function of DSGE models is a nonlinear function of parameters and latent variables, obtained from the numerical solution of a dynamic optimization problem. As discussed in Ruge-Murcia (2007) and DeJong and Dave (2007), the shape of the likelihood function can generate nontrivial problems of estimation. In the model proposed in Ruge-Murcia (2007), the log-likelihood function is flat with respect to the discount factor, preventing the correct assessment of this parameter. As mentioned

in DeJong and Dave (2007), the presence of cliffs in the likelihood function can be problematic, since it generates points of non-differentiability, thus violating one of the regularity conditions of maximum likelihood estimators (van der Vaart (1988), Chernozhukov and Hong (2004)). Although some solutions have been proposed, such as the use of more robust methods for discontinuities (e.g. Chris Sims' CSMINWELL algorithm) and the imposition of transformations and restrictions on parameters so as to change the curvature of the likelihood function (e.g. DeJong and Dave (2007)), they do not effectively solve the problems of non-regularity in the likelihood function of DSGE models.

The literature on the estimation of DSGE models focuses on Bayesian estimation methods, usually Markov Chain Monte Carlo (MCMC) methods. According to Herbst (2011) and Herbst and Schorfheide (2013), 95% of papers published in the top economics journals from 2005 to 2010 use the MCMC methods to perform Bayesian estimation of DSGE models. Although part of the motivation for the use of Bayesian methods is related to the dynamic Bayesian learning process in macroeconomic models, as discussed in Poirier (2012), it is possible to speculate that this choice is mostly due to the computational advantages of this method and to the possible use of prior information, which is especially relevant in macroeconomic models with limited sample sizes available. Canova (2007) presents yet another advantage of using Bayesian methods for the estimation of DSGE models: posterior distributions incorporate uncertainty about the parameters and model specification, making them more attractive to macroeconomists. For these reasons, the vast majority of articles in the literature uses Bayesian methods, especially the MCMC algorithms, to estimate DSGE models.

The regularity conditions of MCMC-based Bayes estimators are less restrictive than those of maximum likelihood estimation, as discussed in Robert and Casella (2004) and Chernozhukov and Hong (2004). MCMC-based estimators are valid under general conditions that do not depend on the continuity of the likelihood function. These estimators can be thought of as solutions to global convex optimization problems, avoiding the problems of existing local maxima in numerical algorithms of likelihood optimization. The MCMC methods replace the optimization problem with a corresponding problem of expectations calculation (posterior mean of Markov chains) that is computationally more robust. Other problems with non-regular likelihood models are discussed in Chernozhukov and Hong (2004), for instance, the fact that the asymptotic theory is not standard and overall statistics are not sufficient in this context.

Through Bayesian methods, it is possible to efficiently infer the posterior distribution of the parameters and volatility, even when the covariance matrix of the vector of the endogenous variables is singular, since the Hessian is not necessary. However, these algorithms also have problems in the estimation of DSGE mod-

els. As we will show in the next sections, there is a great dependence on the choice of the prior distribution. This means that if this distribution is informative, the results might be affected. This variability can be problematic, since different prior distributions can lead to different results and conclusions about the object of analysis.

The procedure proposed in this paper allows constructing maximum likelihood estimators as the limit of Bayes estimators, from MCMC methods, using the data cloning method proposed in Lele et al. (2007) and Lele et al. (2010). While the construction of likelihood estimators using Bayesian methods is already well established in the literature of asymptotic inference (Walker (1969), Ibragimov and Has'minski (1981), van der Vaart (1988)), the construction of estimators based on data cloning is computationally simpler and allows the use of MCMC estimation algorithms of DSGE models with a slight modification. Basically, it is only necessary to build a vector containing  $k$  replications of the original sample and to make a minor change in the calculation of the Fisher information matrix. It is also easy to show that this method is computationally efficient, even in dynamic models and in the presence of latent variables, as shown in Jacquier et al. (2007) and Laurini (2013). Under standard regularity conditions, (Lele et al. (2010)) with sufficiently large  $k$ , these estimators converge to the maximum likelihood estimators.

As discussed in Lele et al. (2010), the use of data cloning has several other advantages. The algorithm avoids problems in numerical procedures for maximizing the likelihood function, replacing them with MCMC sampling procedures, and thus it is unnecessary to calculate the derivative of the likelihood function of the variance-covariance matrix. This method also avoids the use of diffuse or non-informative priors to provide equivalent frequentist inference, which can be computationally and theoretically problematic, as discussed in Robert and Casella (2004) and Seaman et al. (2012).

This paper is structured as follows. Section 2 presents the data cloning method for obtaining a maximum likelihood estimator that is more robust to the usual problems. In section 3, we show evidence of robustness of the data cloning method when compared to usual maximum likelihood estimators. Section 4 aims to support the use of data cloning for the estimation of DSGE models. Section 5 assesses identification problems. Section 6 presents the prior feedback method, i.e. a computationally efficient way to implement the data cloning method. Section 7 briefly describes implementation details. Finally, section 8 concludes.

## 2 The Data Cloning Method

Consider the following (hierarchical or latent factor) model:

$$\underline{Y} \sim f(\underline{y} | \underline{X}, \varphi) \quad (1)$$

$$\underline{X} \sim g(\underline{x} | \theta) \quad (2)$$

where,  $\underline{Y} = (Y_1, Y_2, \dots, Y_n)$  is the vector of observations,  $f$  and  $g$  are joint probability density functions,  $\underline{X}$  is a vector of latent factors,  $\varphi = (\varphi_1, \varphi_2, \dots, \varphi_q)$  is a vector of fixed unknown parameters that directly affect the observations  $\underline{Y}$  and  $\theta = (\theta_1, \theta_2, \dots, \theta_p)$  is a vector of fixed unknown parameters related to the latent state process  $\underline{X}$ .

The likelihood function for the model described by equations (1) and (2) is given by:

$$L(\theta, \varphi; \underline{y}) = \int f(\underline{y} | \underline{X}, \varphi) g(\underline{X} | \theta) d\underline{X} \quad (3)$$

where,  $\underline{y}$  represents realizations of the random variable  $\underline{Y}$ .

The maximum likelihood estimates of the parameters, defined as:

$$\left( \hat{\theta}, \hat{\varphi} \right) = \left( \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p; \hat{\varphi}_1, \hat{\varphi}_2, \dots, \hat{\varphi}_q \right), \quad (4)$$

are the values of  $(\theta, \varphi) = (\theta_1, \theta_2, \dots, \theta_p; \varphi_1, \varphi_2, \dots, \varphi_q)$  that jointly maximize the likelihood function (3). This maximization, however, is not simple, considering it involves elevated computational cost of high dimensional integration, since it is necessary to marginalize the latent variables present in the likelihood function.

The use of Bayesian methods circumvents this problem. The parameters  $(\theta, \varphi)$  are no longer fixed and unknown, and become random variables. The prior distribution, i.e. the joint distribution of these random variables, is a measure of the analyst's beliefs about the different values of the parameters before observing the data. Its combination with the likelihood function, using the Bayes Theorem, generates the posterior distribution, which represents the update on the analyst's beliefs about the different values of the parameters after observing the data.

Defining the prior distribution as  $\pi(\theta, \varphi)$ , the posterior distribution is given by:

$$h(\theta, \varphi, \underline{X} | \underline{y}) = \frac{f(\underline{y} | \underline{X}, \varphi) g(\underline{X} | \theta) \pi(\theta, \varphi)}{\int f(\underline{y} | \underline{X}, \varphi) g(\underline{X} | \theta) \pi(\theta, \varphi) d\underline{X} d\theta d\varphi} \quad (5)$$

The marginal posterior distribution of the parameters,  $\pi(\theta, \varphi | \underline{y})$  is obtained, then, by integrating the posterior distribution,  $h(\theta, \varphi, \underline{X} | \underline{y})$  for  $\underline{X}$ . Through MCMC algorithms, it is possible to generate random numbers from the posterior distribution using only the numerator of equation (5), precluding the use of integrations in multiple dimensions.

Denote the random numbers generated by the MCMC algorithms as  $(\theta, \varphi, \underline{X})_j$ ,  $j = 1, 2, \dots, B$ , with  $B$  large enough to generate a good estimate of  $h(\theta, \varphi, \underline{X} | \underline{y})$ . Once obtained the posterior distribution, the marginal posterior distribution of  $(\theta, \varphi)$  is obtained simply by dropping the component  $\underline{X}$  of the random numbers  $(\theta, \varphi, \underline{X})_j$ , generating  $(\theta, \varphi)_j$ ,  $j = 1, 2, \dots, B$ . Similarly, the mean and variance of  $\pi(\theta, \varphi | \underline{y})$  are the mean and variance of the random numbers  $(\theta, \varphi)_j$ ,  $j = 1, 2, \dots, B$ .

The estimates from the MCMC algorithms, however, are also problematic, considering that there is a large dependence on the choice of the prior distribution in small samples, as typically observed in the estimation of DSGE models. In other words, if the structure is highly informative, the results might be affected, which may lead to different conclusions about the object of analysis. To take advantage of MCMC computational methods while obtaining maximum likelihood estimators, which do not depend on the use of prior information, one possibility is to create estimators that behave as the limit of Bayesian estimators. A simple way to obtain estimators with these properties is the use of data cloning, developed by Lele et al. (2007) and Lele et al. (2010).

To understand this method, assume a hypothetical situation in which an individual performs simultaneously  $k$  independent statistical experiments, producing exactly the same results,  $\underline{y}$ . The new likelihood function is now given by the original function,  $L(\theta, \varphi; \underline{y})$ , raised to the  $k$ -th power:  $[L(\theta, \varphi; \underline{y})]^k$ . Additionally, assume that this individual gets a posterior distribution given by  $h^{(k)}(\theta, \varphi, \underline{X} | \underline{y})$  and a marginal posterior distribution,  $\pi^{(k)}(\theta, \varphi | \underline{y})$ , using the prior distribution  $\pi(\theta, \varphi)$  and the likelihood function with  $k$  clones of the data.

Formally, the original marginal posterior distribution is given by:

$$\pi(\theta, \varphi | \underline{y}) = \frac{\{\int f(\underline{y} | \underline{X}; \varphi) g(\underline{X} | \theta) d\underline{X}\} \pi(\theta, \varphi)}{C(\underline{y})} \quad (6)$$

where,  $C(\underline{y}) = \int L(\theta, \varphi; \underline{y}) \pi(\theta, \varphi) d\theta d\varphi$  is a normalization constant.

Equation (6) can be rewritten as:

$$\pi(\theta, \varphi | \underline{y}) = \frac{L(\theta, \varphi; \underline{y}) \pi(\theta, \varphi)}{C(\underline{y})} \quad (7)$$

Lele et al. (2007) prove that if  $k$  is large, the marginal posterior distribution,  $\pi^{(k)}(\theta, \varphi | \underline{y})$ , is concentrated around the maximum likelihood estimates. The likelihood function with  $k$  clones of data has the same maximum point of the original function, given by  $(\hat{\theta}, \hat{\varphi})$ , and the Fisher information matrix is  $k$  times the information matrix based on the likelihood function.

Assuming the usual identification conditions of the maximum likelihood estimator, it is possible to show that the marginal posterior distribution of  $\theta$ , from the sample containing  $k$  replications of the original data, is given by:

$$\pi^{(k)}(\theta, \varphi | \underline{y}) = \frac{\{f(\underline{y} | \underline{X}; \varphi) g(\underline{X} | \theta) d\underline{X}\}^k \pi(\theta, \varphi)}{C(k, \underline{y})} \quad (8)$$

where,  $C(k, \underline{y}) = \int L(\theta, \varphi; \underline{y})^k \pi(\theta, \varphi) d\theta d\varphi$  is a normalization constant.

Lele et al. (2010) prove that, under regularity conditions, with large number of clones,  $\pi^{(k)}(\theta, \varphi | \underline{y})$  converges to a multivariate Normal distribution with mean equal to the maximum likelihood estimate and variance equal to  $\frac{1}{k} I^{-1}(\hat{\theta}, \hat{\varphi})$ , where  $I(\hat{\theta}, \hat{\varphi})$  is the Fisher information matrix corresponding to the original likelihood function. Formally:

$$\pi^{(k)}(\theta, \varphi | \underline{y}) \xrightarrow{k \rightarrow \infty} N\left(\left(\hat{\theta}, \hat{\varphi}\right), \frac{1}{k} I^{-1}\left(\hat{\theta}, \hat{\varphi}\right)\right) \quad (9)$$

Evidently, in practice we do not have  $k$  independent experiments, but we can proceed by using  $k$  replicates of the original sample (clones). In this situation, according to Lele et al. (2007) and Lele et al. (2010), we do not have the convergence in probability used in Walker (1969), but the deterministic convergence of a sequence of functions.

The results of Lele et al. (2007) and Lele et al. (2010) show that it is possible to sample sequences of values  $(\theta, \varphi)_j$  from the marginal posterior distribution,  $\pi^{(k)}(\theta, \varphi | \underline{y})$ , and obtain the maximum likelihood estimator as the average of these values, as well as the variance-covariance matrix as  $k$  times the variance-covariance matrix of this sample:

$$E^{(k)}(\theta, \varphi | \underline{y}) \xrightarrow{k \rightarrow \infty} \left(\hat{\theta}, \hat{\varphi}\right) \quad (10)$$

$$Var^{(k)}(\theta, \varphi | \underline{y}) \xrightarrow{k \rightarrow \infty} \sqrt{k} \left(\hat{\theta}, \hat{\varphi}\right) \quad (11)$$

Proof of these properties can be found in Lele et al. (2010) and Baghishani and Mohammadzadeh (2011). Note that several Monte Carlo optimization procedures, similar to the convergence of posterior probability to maximum likelihood estimator, had already been discussed in the literature, e.g. Pincus (1968), ? and D'Epifanio (1996). See also the discussion on the method of Prior Feedback in Robert and Casella (2004), pps 169-173 and in Section 6.

As pointed out in Lele et al. (2010), one should note that data cloning methods are only an approximation mechanism in order to obtain maximum likelihood estimators using Bayesian estimators. Thus, the usual limits and properties in

finite samples of likelihood estimators are still valid. These methods are not meant to reduce the bias of maximum likelihood estimators.

The quasi-Bayesian estimators proposed in Chernozhukov and Hong (2003) and Chernozhukov and Hong (2004) are also related to the data cloning method, using MCMC to construct maximum likelihood estimators. In this case, a Random Walk-Metropolis Hastings algorithm is applied to a Laplace approximation of an objective function, such as likelihood functions. The main difference is that this method is directly equivalent to the maximum likelihood estimator when using weighting functions equivalent to improper priors. In the case where (weighted) proper priors are used, the quasi-Bayesian estimator is only asymptotically equivalent to the maximum likelihood estimator, but as they are based on the original data sample without replications, they still carry the information of the pseudo-priors used. In the event that these pseudo-priors are flat, mimicking the maximum likelihood estimator, the method can suffer from existing computational problems in Bayesian estimators, exploring very inefficiently the likelihood function and requiring a very large number of samples to obtain estimators with adequate accuracy in the approximation of the likelihood function. Note that the concept of data cloning can be used in the estimator proposed in Chernozhukov and Hong (2003) and Chernozhukov and Hong (2004), applying the procedure of cloning for the sample used in the Laplace transformation of the objective function, as stated in Corollary 5.11 in Robert and Casella (2004), discussing the theoretical properties of Prior Feedback method.

In some preliminary experiments, we noted that the method of Chernozhukov and Hong (2003) and Chernozhukov and Hong (2004) was inaccurate and unstable in the estimation of the model used in this article. In order to get more computationally stable estimation, it was necessary to use informative weights (pseudo-priors), which did not allow a direct comparison with the data cloning method, that eliminates the influence of priors. Although the method proposed in Chernozhukov and Hong (2003) and Chernozhukov and Hong (2004) is more general than the data cloning method, as it can be used for any type of econometric objective function, these computational limitations favor the use of data cloning. The fact that there exists a vast literature on the estimation of DSGE models, as well as comprehensive computational implementation of general solution methods and of MCMC-based Bayesian estimation, are also favorable to the use of data cloning, due to its ease of implementation

### 3 Estimation Results

After introducing the data cloning method, we propose a comparative exercise with the usual and most important maximum likelihood estimation meth-

ods. To achieve this objective, we use the database of Schorfheide (2000) and the Cash-in-Advance model (hereafter CIA) presented by the author (model 1), briefly reproduced below.

The economy in this model is composed of a representative agent, a firm, and a financial intermediary. The product is determined by the following Cobb-Douglas function:

$$Y_t = K_t^\alpha (A_t N_t)^{1-\alpha} \quad (12)$$

where,  $Y_t$  is the product at  $t$ ,  $K_t$  denotes the capital stock,  $N_t$  is the measure of labor and  $A_t$  represents technology, evolving according to the following exogenous process:

$$\ln A_t = \gamma + \ln A_{t-1} + \epsilon_{A,t} \quad (13)$$

where,  $\epsilon_{A,t} \sim N(0, \sigma_A^2)$ .

In each period, the Central Bank adjusts the rate of growth of the money stock, disturbing the economy with a second exogenous process, according to the following monetary policy rule:

$$\ln m_t = (1 - \rho) \ln m^* + \rho \ln m_{t-1} + \epsilon_{M,t} \quad (14)$$

where,  $\epsilon_{M,t} \sim N(0, \epsilon_M^2)$  and changes in  $m^*$  and  $\rho$  are associated to rare regime shifts.

At the beginning of period  $t$ , the representative agent inherits the money stock of the economy,  $M_t$ . It is easy to see, from equations (13) and (14) that all decisions of the agent reflect shocks - i.e. surprises - to the growth rate of capital stock and technology. The representative agent determines the portion of the capital stock deposited at the bank,  $D_t$ . The bank receives the deposits from the agent, paying interest  $R_{H,t} - 1$ , the Central Bank's monetary injection,  $X_t$ , and grants loans to the firm at rate  $R_{F,t} - 1$ .

The firm then begins production, requiring hours of work from the representative agent. The amount borrowed by the bank is used to pay wage  $W_t H_t$  to the agent, where  $W_t$  is the nominal wage per hour and  $H_t$  denotes the hours worked. The agent's cash balance in the period  $t$  is therefore given by  $M_t - D_t + W_t H_t$ . It is important to note that, in the CIA model, the agent cannot leverage consumption with loans from the bank. Expenses must be financed by the agent's accumulated resources in that period.

At the end of period  $t$ , the agent receives dividends from the firm, represented by  $F_t$ . He also receives, from the bank, the amount deposited at the beginning of the period - these deposits earn interest  $R_{H,t} - 1$  -, and dividends,  $B_t$ .

In summary, in each period, the representative agent chooses the amount consumed,  $C_t$ , hours worked,  $H_t$ , and the money deposited in the bank,  $D_t$  to maximize

the present value of the sum of expected future utility. Formally, agent's problem can be represented by the following set of equations:

$$\begin{aligned}
& \max_{\{C_t, H_t, M_{t+1}, D_t\}} E_0 \left[ \sum_{t=0}^{\infty} \beta^t [(1 - \phi) \ln C_t + \phi \ln (1 - H_t)] \right] \quad (15) \\
& \text{s.a. } P_t C_t \leq M_t - D_t + W_t H_t \\
& M_{t+1} = (M_t - D_t + W_t H_t - P_t C_t) + R_{H,t} D_t + F_t + B_t \\
& D_t \geq 0
\end{aligned}$$

where,  $P_t$  is the price level.

The maximization problem of the firm involves the choice of the capital stock of the period  $t + 1$ ,  $K_{t+1}$ , labor demand,  $N_t$ , payment of dividends,  $F_t$ , and the amount of loans,  $L_t$ . Formally:

$$\begin{aligned}
& \max_{\{F_t, K_{t+1}, N_t, L_t\}} E_0 \left[ \sum_{t=0}^{\infty} \beta^{t+1} \frac{F_t}{C_{t+1} P_{t+1}} \right] \quad (16) \\
& \text{s.a. } F_t \leq P_t [K_t^\alpha (A_t N_t)^{1-\alpha} - K_{t+1} + (1 - \delta) K_t] - W_t N_t - L_t R_{F,t} \\
& W_t N_t \leq L_t
\end{aligned}$$

The financial intermediary solves the problem:

$$\begin{aligned}
& \max_{\{B_t, L_t, D_t\}} E_0 \left[ \sum_{t=0}^{\infty} \beta^{t+1} \frac{B_t}{C_{t+1} P_{t+1}} \right] \quad (17) \\
& \text{s.a. } B_t = D_t + R_{F,t} L_t - R_{H,t} D_t - L_t + X_t \\
& L_t \leq X_t + D_t
\end{aligned}$$

where,  $X_t = M_{t+1} - M_t$  is the monetary injection from the Central Bank. In equilibrium,  $H_t = N_t$ ,  $P_t C_t = M_t + X_t$ ,  $C_t + (K_{t+1} - (1 - \delta) K_t) = K_t^\alpha (A_t H_t)^{1-\alpha}$  e  $R_{F,t} = R_{H,t}$ .

According to Schorfheide (2000), to solve the model, one should derive optimality conditions for the maximization problems. After detrending the variables, it is possible to find a deterministic steady state. The system can be log-linearized around it to obtain the solution by elimination of unstable roots. Using the notation of Schorfheide (2000), the state-space representation of the log-linearized DSGE model is given by:

$$y_t = \Xi_0 + \Xi_1 s_t + \Xi_* \epsilon_t \quad (18)$$

$$s_t = \Psi_1 s_{t-1} + \Psi_* \epsilon_t \quad (19)$$

$$\epsilon_t \sim iidN(0, \Sigma_t)$$

where,  $y_t$  is a  $nx1$  vector of observable variables,  $\epsilon_t = [\epsilon_{A,t}, \epsilon_{M,t}]'$ ,  $\Sigma_t$  is the diagonal matrix with elements  $\sigma_A^2$  and  $\sigma_M^2$ ,  $s_t$  is a vector of deviations of the detrended model from its steady state, and the matrices  $\Xi_0$ ,  $\Xi_1$ ,  $\Xi_*$ ,  $\Psi_1$  and  $\Psi_*$  are functions of the structural DSGE model parameters.

In our experiment, we use the CIA model just described above to make estimates with 50 different initial values for the most important methods of maximum likelihood estimation implemented in Dynare, which are: the Nelder-Mead simplex method, the CSMINWELL algorithm that is robust to cliffs in the likelihood function of Chris Sims, the covariance matrix adaptation evolution strategy (CMA-ES) method, an evolutionary optimization algorithm developed for non-linear non-convex optimization problems, and Marco Ratto's newrat method. All these algorithms are designed to be robust to poor initial values and discontinuities in the likelihood function. The results of our experiment are compared with data cloning estimates, with 50 and 100 replications. In the data cloning estimation, each different initial value is used in two ways - as an average of prior distributions and as initialization of the MCMC algorithm. The structure of the prior distributions follows Schorfheide (2000). The random sampling of initial values is made by adding a uniform  $[-0.07, 0.07]$  error to the initial values assumed in Schorfheide (2000).

As mentioned earlier, the database in our experiment is fixed and simulated with the estimated values of Schorfheide (2000), and the same simulated database is used in the estimation with distinct initial values. Therefore, by using these parameters in the simulation we have a 'true value' of the vector of parameters to compare with the values estimated by the different methods and different initial values. At each replication, we change the vector of initial values, generating a random sample around the true vector of parameters to initialize the usual maximum likelihood estimates, the priors and the Markov chains in the data cloning method. In each replication, all methods use the same vector of initial values. The results in Table 1 show the mean and standard deviation of the parameters obtained in the  $n$  replications of the experiment, and in figure ?? we show the boxplots with distributions of estimated parameters. In the experiment with the likelihood methods, we used the point estimation of the optimization algorithm. The average posterior of the Markov chains, with 20,000 samples, after discarding the first 10,000 samples, was used as point estimator of the data cloning method.

After analyzing the table, it is easy to note significant differences in the estimation results obtained from the various maximum likelihood methods. This suggests that there is a great dependence of initial values on these methods. All numerical likelihood optimization methods used in this experiment have been designed to be robust, but in practice it is not possible to check this property for the proposed

	simplex		csminewell		cma-es		newrat		dcmle 50		dcmle 100	
	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd	mean	sd
$\alpha$	0.6155	0.1572	0.4983	0.1795	0.7226	0.0745	0.5093	0.0852	0.4900	0.0184	0.4884	0.0201
$\beta$	0.9139	0.0740	0.9239	0.0958	0.9166	0.0465	0.9761	0.0144	0.9854	0.0025	0.9861	0.0019
$\gamma$	0.0054	0.0067	0.0087	0.0137	0.0048	0.0016	0.0048	0.0006	0.0038	0.0006	0.0039	0.0006
$\ln m^*$	0.0101	0.0003	0.0100	0.0002	0.0101	0.0000	0.0101	0.0000	0.0101	0.0000	0.0101	0.0000
$\rho$	0.6130	0.3030	0.3465	0.3231	0.7423	0.2454	0.6092	0.2878	0.8367	0.0428	0.8413	0.0364
$\phi$	0.6149	0.3256	0.6306	0.0635	0.7198	0.2348	0.7338	0.2472	0.7897	0.0505	0.7908	0.0561
$\delta$	0.0318	0.0664	0.0771	0.1192	0.0004	0.0009	-0.0000	0.0000	0.0011	0.0013	0.0012	0.0014
$\sigma_a$	0.0263	0.0104	0.0274	0.0268	0.0342	0.0111	0.0167	0.0048	0.0153	0.0007	0.0153	0.0007
$\sigma_m$	0.0050	0.0021	0.0157	0.0150	0.0037	0.0010	0.0048	0.0016	0.0034	0.0007	0.0034	0.0006

Table 1: Sensitivity to initial conditions - Mean values and standard deviations for parameters estimated using MLE by numerical optimization and data cloning. Values used in the simulation -  $\alpha=0.4168$ ,  $\beta=0.9901$ ,  $\gamma=0.0038$ ,  $\ln m^*=0.0141$ ,  $\rho=0.8623$ ,  $\phi=0.6837$ ,  $\delta=0.0020$ ,  $\sigma_a=0.0133$ ,  $\sigma_m=0.0029$ .

model. The experiment results also show that the data cloning estimators have low variability and are closer to the true values used in the simulation, highlighting the robustness of this method to the problem of initial conditions. The variability in the data cloning estimators is of order  $\sqrt{n}$ , generated by the Monte Carlo approximation in the MCMC method.

It is also important to note that the computational cost favors the data cloning method. The MCMC sampling procedure is very efficient, allowing a low computational time for each sample. In the numerical optimization methods, the convergence procedure is time consuming and oftentimes it is necessary to restart the algorithm, especially in the presence of cliffs on the likelihood function (e.g. CSMINWELL).

## 4 Monte Carlo Evidence

Based on the favorable evidence concerning the data cloning method presented in the previous section, we conducted a new experiment, in order to support the use of this method in the estimation of DSGE models. We performed a standard Monte Carlo experiment to check the properties of the estimators obtained from simulations of the 'true model' samples, instead of comparing the effects of the vector of initial parameters using a fixed sample.

We simulated observations using the model of Schorfheide (2000), utilizing the same vector of parameters shown in section 3 as 'true' parameters, and estimated the parameters by the default maximum likelihood estimation method of Dynare - CSMINWELL - and the data cloning algorithm with 50 replications of the original sample. We used two sets of initial values in each method, to check its sensitivity in the estimates: the 'true vector' of parameters, as the first vector of initial values,

and a second vector of parameters, using the mean for the priors assumed in the Bayesian estimation reported in Schorfheide (2000). The initial values are also used as the mean values for the priors in the data cloning method. In Table 2, we show the mean error (bias) and the root mean square error - henceforth RMSE - in relation to the parameters used in the simulation exercise, and in figure 2 boxplots of the distributions of estimated parameters obtained in this experiment.

	mle 1		mle 2		dcmle 1		dcmle 2	
	ME	RMSE	ME	RMSE	ME	RMSE	ME	RMSE
$\alpha$	-0.0529	0.0876	-0.0587	0.0852	0.0073	0.0518	-0.0014	0.0568
$\beta$	-0.0029	0.0180	-0.0095	0.0143	-0.0019	0.0064	0.0021	0.0045
$\gamma$	0.0014	0.0021	-0.0001	0.0010	-0.0001	0.0009	-0.0000	0.0009
$\ln m^*$	0.0001	0.0001	-0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
$\rho$	0.3083	0.4660	-0.0915	0.1041	0.0251	0.0823	0.0200	0.0729
$\phi$	0.1190	0.1608	-0.1028	0.1057	0.1233	0.1466	0.0508	0.1556
$\delta$	-0.0165	0.0544	-0.0028	0.0206	-0.0038	0.0215	-0.0082	0.0236
$\sigma_a$	-0.0029	0.0053	-0.0017	0.0040	-0.0005	0.0016	-0.0003	0.0021
$\sigma_m$	-0.0043	0.0053	0.0002	0.0004	-0.0001	0.0004	-0.0001	0.0003

Table 2: Mean Error and Root Mean Square Error - Monte Carlo Analysis of CSMINWELL and Data Cloning Likelihood Estimation of CIA Model. Parameters in simulations -  $\alpha=0.4168$ ,  $\beta=0.9901$ ,  $\gamma=0.0038$ ,  $\ln m^*=0.0141$ ,  $\rho=0.8623$ ,  $\phi=0.6837$ ,  $\delta=0.0020$ ,  $\sigma_a=0.0133$ ,  $\sigma_m=0.0029$ .

Table 2 shows that, even when initializing the parameters with their real value, there is a significant difference in the results obtained with different initial values using the maximum likelihood method with numerical optimization. The estimation results for the data cloning method suggest low sensitivity to initial conditions and good results in terms of bias and RMSE, supporting the use of this method in the estimation of DSGE models.

Finally, we use the data cloning method with 50 and 100 replications to estimate the maximum likelihood of the model presented in Section 3, using the same database and priors used in Schorfheide (2000), comparing the results with the posterior values estimated by Schorfheide (2000), using the Bayesian MCMC estimation. Table 3 shows the parameters and standard deviations reported in Schorfheide (2000) and the results obtained by the data cloning estimators. It is easy to note that the results of the data cloning estimation are consistent with the general results obtained in Schorfheide (2000), except for the parameter  $\phi$ . The value obtained for this parameter, however, is closer to one, which could indicate an outcome that is more consistent with the economic theory.

	Bayes	Bayes SD	dcmle 50	dcmle50 SD	dcmle 100	dcmle100 SD
$\alpha$	0.4168	0.0218	0.4876	0.0087	0.4876	0.0098
$\beta$	0.9901	0.0021	0.9855	0.0259	0.9862	0.0222
$\gamma$	0.0038	0.0010	0.0036	0.0071	0.0035	0.0102
$\ln m^*$	0.0141	0.0017	0.0101	0.0001	0.0101	0.0002
$\rho$	0.8623	0.0343	0.8539	0.1988	0.8548	0.2626
$\phi$	0.6837	0.0479	0.8911	0.5263	0.8973	0.2000
$\delta$	0.0020	0.0011	0.0001	0.0013	0.0000	0.0013
$\sigma_a$	0.0133	0.0009	0.0147	0.0056	0.0147	0.0078
$\sigma_m$	0.0029	0.0002	0.0030	0.0012	0.0030	0.0012

Table 3: Estimation Results - Data Cloning using 50 and 100 replications for the model and dataset of Schorfheide (2000). Bayes corresponds to the Bayesian estimation reported in Schorfheide (2000), and dcmle to the data cloning estimation, using 50 and 100 replications (clones) of the original sample.

## 5 Econometric Identification

A typical problem in the implementation and estimation of DSGE models is verifying the identification conditions, i.e. if there is only one vector of parameters consistent with the maximum of the likelihood function. Identification is one of the fundamental regularity conditions on maximum likelihood estimation, e.g. van der Vaart (1988). If this condition is violated, the Hessian of the optimization problem is singular or near-singular, causing problems in numerical optimization methods and in the calculation of the variance of the parameters. As discussed in Canova and Sala (2009), observational equivalence, partial and weak identification problems are widespread, difficult to identify empirically and may lead to biased estimation, unreliable hypothesis testing and to the choice of false models. Thus, it is essential to check the identification in DSGE models.

As discussed in Lele et al. (2010), it is possible to evaluate the identification of parameters directly through the data cloning procedure. According to the authors, a parameter is nonidentifiable if its variability does not converge to zero with the increase in the number of replications. This identification check can be easily performed graphically, using the standard deviations of the chains sampled in the MCMC procedure. We conducted this analysis for the estimates of the model of Schorfheide (2000), showing in Figure 3 (a) the standard deviation of all parameters as a function of the number of clones. In figure 3 (a), we normalized the standard deviation to enhance visualization. Through the analysis of this figure, it is clear that all parameters in our experiment are identified, as decreased variability is evident from the increased number of replications, showing convergence

to the point mass maximum likelihood estimator.

Lele et al. (2010) also discuss that a global identification measure may be obtained from the maximum eigenvalue of the covariance matrix of the parameters in each data cloning procedure. Similarly to the former identification scheme, the model is globally identified if the maximum eigenvalue converges to zero with the increase in the number of replications. Figure 3 (b) confirms the previous evaluation regarding the identification of the model, by showing that the maximum eigenvalue converges to zero with the increase in the number of replications. These two analytical identification procedures are immediate by-products of the data cloning method, allowing an empirical identification evaluation procedure of the model and responding to a fundamental problem in the estimation of DSGE models.

## 6 Prior Feedback

Prior feedback is a complementary mechanism for the data cloning estimation (Robert and Soubiran (1993), Robert and Casella (2004)). This procedure uses an increasing number of replications, initializing the prior distribution of the estimation with  $k$  replications, with the posterior distribution obtained in the estimation performed with  $k - 1$  replications. This algorithm is initialized with one replication (the standard Bayesian estimation). The number of replications is increased until the parameters converge by a predetermined measure (e.g. Robert and Casella (2004)). The great advantage of this estimation method is that it is more robust to possible problems of convergence of the MCMC chains when the procedure is initialized directly with a large number of clones, as discussed in Robert (2010). Additionally, this estimation procedure inherits some properties of robustness from simulated annealing optimization methods, as discussed in Robert and Soubiran (1993).

In this section, we propose a final analysis, using the procedure briefly described in the previous paragraph, with the original dataset and with the initial vector of prior distributions of Schorfheide (2000) to make the estimation. We define convergence when the variation in the marginal likelihood obtained from the MCMC estimation between the models with  $k$  and  $k + 1$  clones is less than 0.001. When this convergence is obtained, the algorithm is terminated. Figure 4 shows the value of the marginal likelihood of the model obtained for each number of replications. With these criteria, convergence is obtained with only 21 clones. Note that we can define other metrics of convergence, such as some rule of variation in the parameters, similar to the procedures for convergence in the function and the parameters in numerical optimization methods. It is advisable to use the prior feedback procedure when there are convergence problems in the MCMC algo-

rithm, allowing for local exploration of the likelihood function. Note that it is also possible to use a more complex version of this algorithm, decreasing the number of clones and resetting the estimation if a problem is encountered in the exploration of the likelihood function with MCMC methods, as in the SAME algorithm proposed in Doucet et al. (2002). It is important to observe, though, that this version is computationally more intense.

allowing to

## 7 Implementation Details

All of the analysis were implemented using Dynare (version 4.3.3, Adjemian et al. (2011)) and the Octave programming language (Eaton (2002)). The data cloning procedure was implemented by simply replicating the original data sample and stacking  $k$  clones. The Bayesian estimation procedure was carried out in the conventional form, by adjusting the calculation of the variance of parameters as described in Section 2. The selection of several numerical optimization methods was made by the mode compute option in the Dynare estimation procedure.

The calibration acceptance rate of the Metropolis-Hastings algorithm (mode computation) can be performed in the standard Bayesian estimation (no clones) and reused in the estimation with  $k$  clones, or calibrated directly in the estimation with  $k$  clones. The results were essentially equivalent. In the prior feedback procedure, in addition to using the posteriors obtained from the estimation with  $k - 1$  clones as priors for the estimation with  $k$  clones, we also calibrated the Metropolis-Hastings algorithm via its acceptance rate using the result of the Bayesian estimation with the original sample (without clones), in order to preclude the use of recalibration in each stage of the prior feedback procedure. The implementation of the data cloning procedure is quite simple and can easily be an additional option for maximum likelihood estimation in Dynare, complementing the several methods that have already been implemented.

## 8 Conclusions

In this paper, we presented the data cloning method of Lele et al. (2007) and Lele et al. (2010) to obtain a maximum likelihood estimator as the limit of a Bayesian estimator, showing evidences that this method is more robust than the traditional likelihood estimators.

These evidences were obtained from three experiments. In the first, we used the dataset of Schorfheide (2000) and the CIA model presented by the author in a comparative exercise with the most important likelihood estimation methods.

We found significant differences in the estimation results of the most important likelihood methods and great dependence on the vector of initial values. The data cloning estimators presented low variability, supporting the robustness of this method to the problem of initial conditions.

The second experiment aimed to support the use of the data cloning method in the estimation of DSGE models. Once again, the results suggested that the data cloning estimators are less sensitive to initial conditions and show good results in terms of bias and RMSE, unlike the traditional maximum likelihood estimators.

We showed evidence of identification of all parameters in our data cloning estimations, as presented in Lele et al. (2010). We also evaluated convergence in a third experiment, which was reached when the variation in the marginal likelihood in each number of clones was less than 0.001, using the prior feedback procedure. By this criterion, only 21 clones were required to obtain convergence.

## Bibliography

- Adjemian, S., H. Bastani, M. Juillard, M. Karamé, F. Mihoubi, J. Perendia, G. Pfeifer, M. Ratto, and S. Villemot (2011). Dynare: Reference manual, version 4. Dynare Working Papers, 1, CEPREMAP.
- Baghishani, H. and M. Mohammadzadeh (2011). A data cloning algorithm for computing maximum likelihood estimates in spatial generalized linear mixed models. *Computational Statistics and Data Analysis* 55(4), 1748–1759.
- Bencivenga, V. (1992). An econometric study of hours and output variation with preference shocks. *International Economic Review* 33, 449–471.
- Canova, F. (2007). *Methods for Applied Macroeconomics Research*. Princeton University Press.
- Canova, F. and L. Sala (2009, May). Back to square one: Identification issues in DSGE models. *Journal of Monetary Economics* 56(4), 431–449.
- Chernozhukov, V. and H. Hong (2003). An MCMC approach to classical estimation. *Journal of Econometrics* 115(2), 293–346.
- Chernozhukov, V. and H. Hong (2004). Likelihood estimation and inference in a class of nonregular econometric models. *Econometrica* 72(5), 1445–1480.
- DeJong, D., B. Ingram, and C. Whiteman (2000a). A Bayesian approach to dynamic macroeconomics. *Journal of Econometrics* 98, 203–233.

- DeJong, D., B. Ingram, and C. Whiteman (2000b). Keynesian impulses versus solow residuals: identifying sources of business cycle fluctuations. *Journal of Applied Econometrics* 15, 311–329.
- DeJong, D. N. and C. Dave (2007). *Structural Macroeconometrics*. Princeton University Press.
- D’Epifanio, G. (1996). Notes on a recursive procedure for point estimation. *TEST* 5, 203–225.
- Doucet, A., . S. J. Godsill, and C. Robert (2002). Marginal maximum a posteriori estimation using markov chain monte carlo. *Statistics and Computing* 12, 77–84.
- Eaton, J. W. (2002). *GNU Octave Manual*. Network Theory Limited.
- Herbst, E. (2011). Gradient and Hessian-based MCMC for DSGE models. Manuscript, University of Pennsylvania.
- Herbst, E. and F. Schorferheide (2013). Sequential Monte Carlo sampling for DSGE models. Manuscript, University of Pennsylvania.
- Ibragimov, I. and R. Has’minski (1981). *Statistical Inference: Asymptotic Theory*. Springer-Verlag.
- Ingram, B., N. Kocherlakota, and N. Savin (1994). Explaining business cycles: a multiple-shock approach. *Journal of Monetary Economics* 34, 415–428.
- Ireland, P. (2001). Sticky-price models of the business cycle: specification and stability. *Journal of Monetary Economics* 47, 3–18.
- Ireland, P. (2004). A method for taking models to the data. *Journal of Economic Dynamics & Control* 28, 1205–1226.
- Jacquier, E., M. Johannes, and N. Polson (2007). MCMC maximum likelihood for latent state models. *Journal of Econometrics* 137(2), 615–640.
- Kim, J. (2000). Constructing and estimating a realistic optimizing model of monetary policy. *Journal of Monetary Economics* 45, 329–359.
- Laurini, M. P. (2013). A hybrid data cloning estimator for stochastic volatility models. *Journal of Time Series Econometrics* forthcoming.
- Lele, S. R., B. Dennis, and F. Lutscher (2007). Data Cloning: easy maximum likelihood estimation for complex ecological models using Bayesian Markov chain Monte Carlo methods. *Ecology Letters* 10, 551–563.

- Lele, S. R., K. Nadeem, and B. Schmuland (2010). Estimability and Likelihood Inference for Generalized Linear Mixed Models Using Data Cloning. *Journal of the American Statistical Association* 105(492), 1617–1625.
- Pincus, M. (1968). A closed form solution of certain programming problems. *Operations Research* 18, 1225–1228.
- Poirier, D. J. (2012). What is sensible for your agents should be sensible for yourself. *Journal of Econometrics* 170, 249 – 250.
- Robert, C. P. (2010). Feedback on data cloning. Comment - <http://xianblog.wordpress.com/2010/09/22/feedback-on-data-cloning/>.
- Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods*. Springer.
- Robert, C. R. and C. Soubiran (1993). Estimation of a normal mixture model through gibbs sampling and and prior feedback. *Test* 2, 125–146.
- Ruge-Murcia, F. J. (2007). Methods to estimate dynamic stochastic general equilibrium models. *Journal of Economic Dynamics & Control* 31, 2599–2636.
- Schorfheide, F. (2000). Loss function-based evaluation of DSGE models. *Journal of Applied Econometrics* 15, 645–670.
- Seaman, J. W., J. W. Seaman, and J. D. Stamey (2012). Hidden dangers of specifying noninformative priors. *The American Statistician* 66(2), 77–84.
- van der Vaart, A. W. (1988). *Asymptotic Statistics*. Cambridge University Press.
- Walker, A. M. (1969). On the asymptotic behavior of posterior distributions. *Journal of the Royal Statistical Association* 31(1), 80–88.

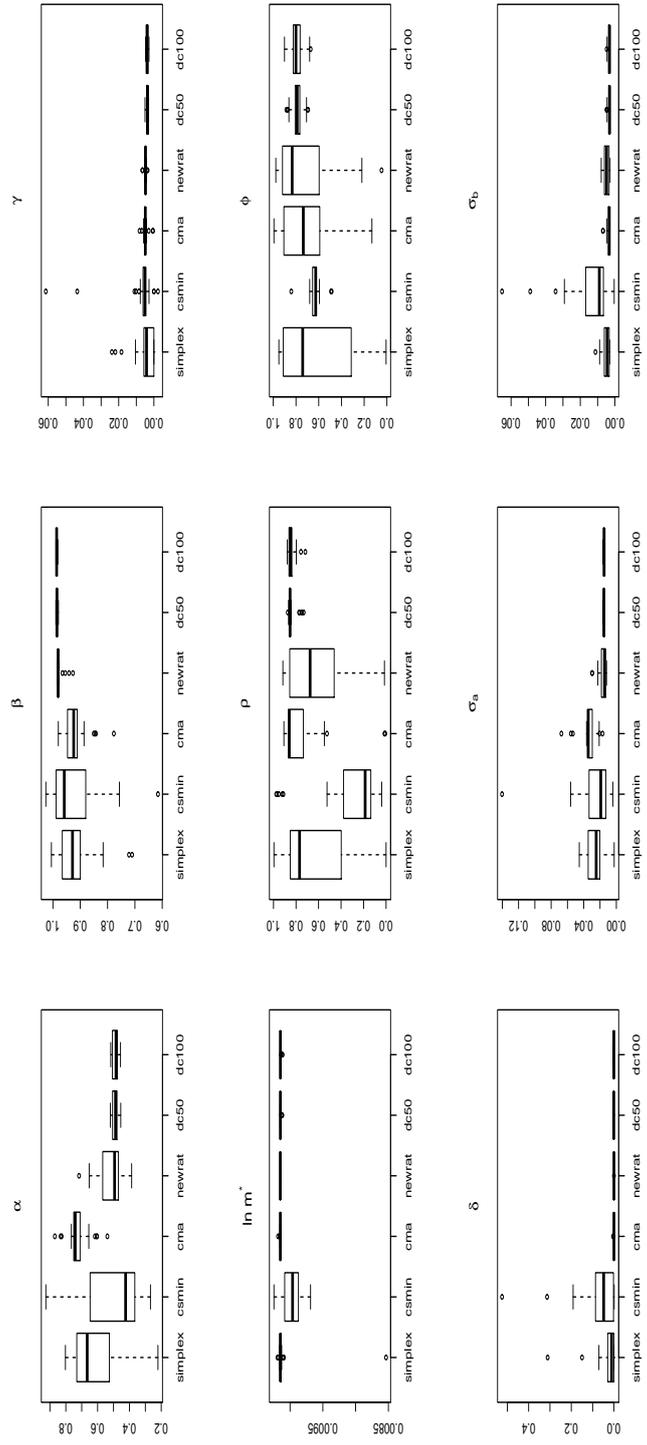


Figure 1: Sensitivity to initial conditions

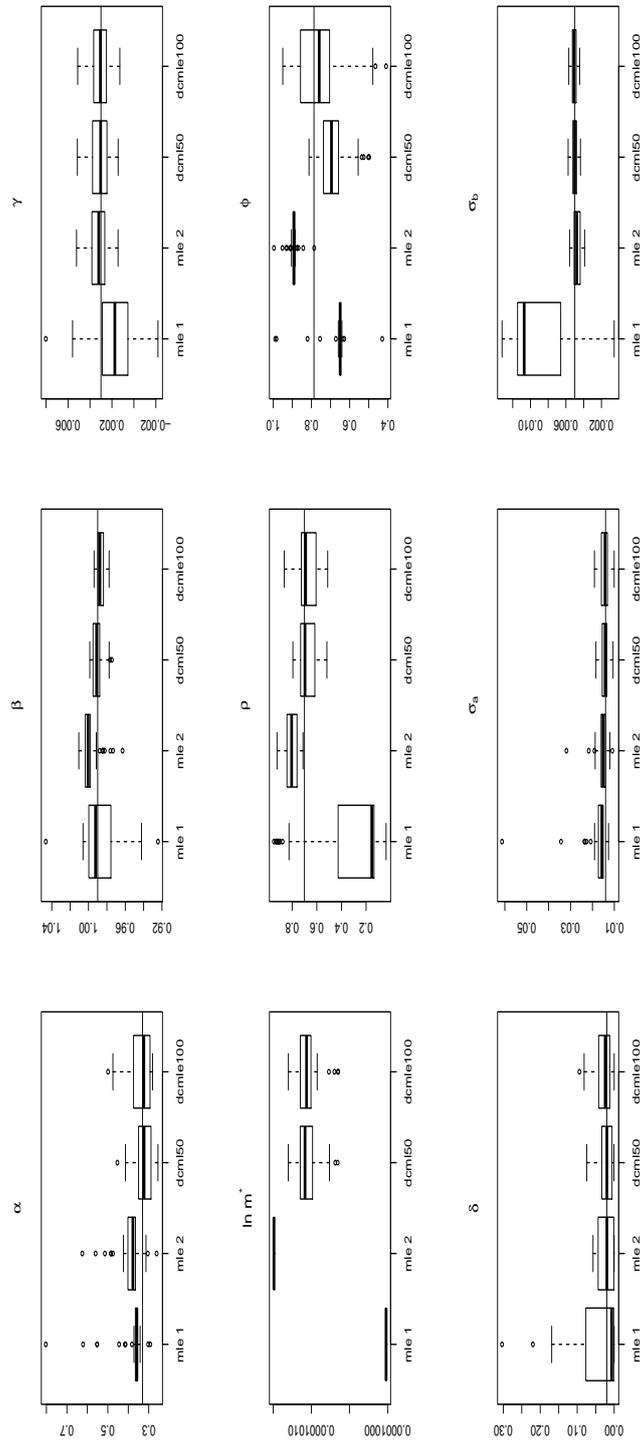
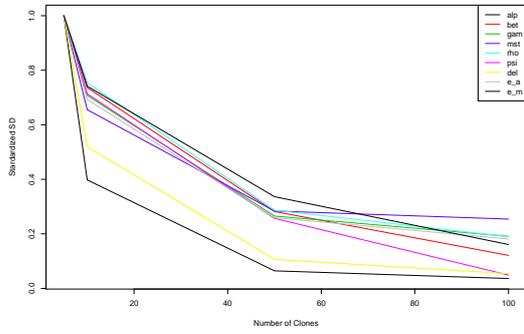
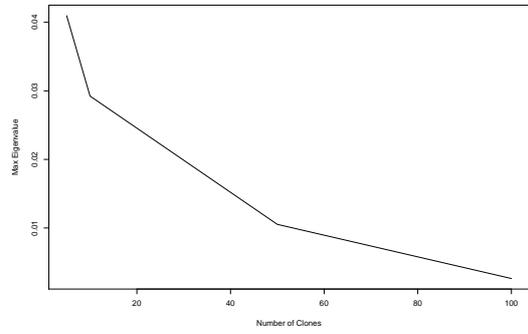


Figure 2: Distributions - Monte Carlo Analysis



(a) MCMC Standard Deviation



(b) Maximum Eigenvalue

Figure 3: Identification

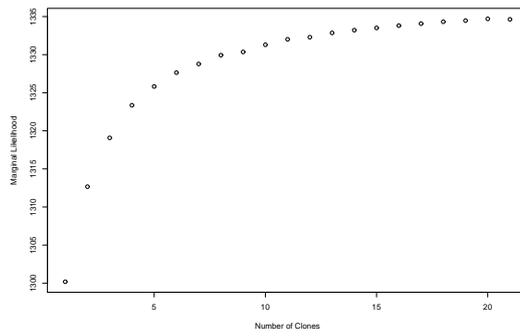


Figure 4: Marginal Likelihood