The Science and Art of DSGE Modelling
A Foundations Course

Bayesian Estimation

Maryam Mirfatah, Sahil Ravgotra and Luciano Rispoli

September 9, 2020

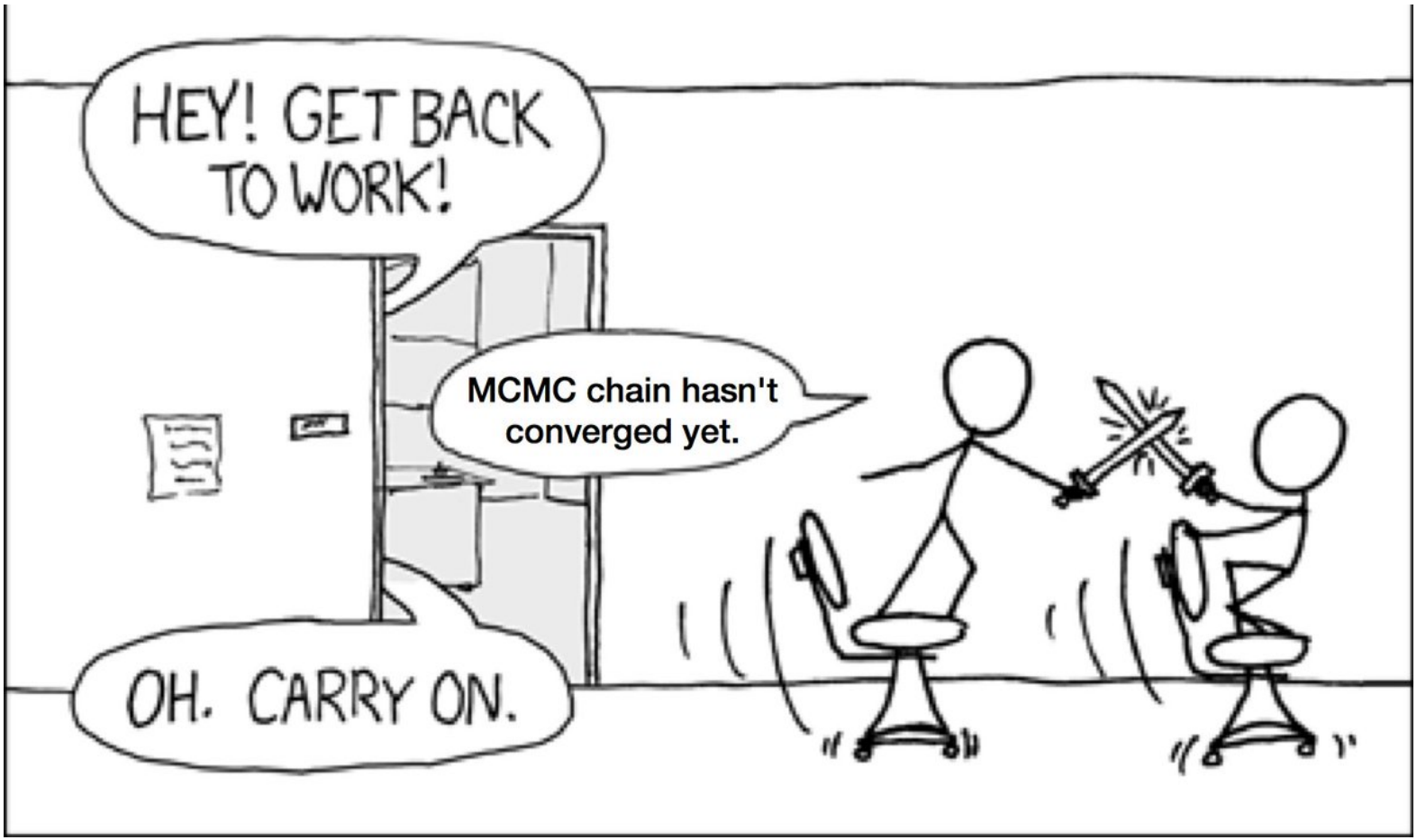
Today's timetable

- 09.30 am - 11.00 am: Session 1 : Lecture
- 11.00 am - 11.30 am: Break
- 11.30 am - 01.00 pm: Session 2 : Lecture
- 01.00 pm - 02.15 pm: Break
- 02.15 pm - 03.45 pm: Session 3 : Lecture /Labs
- 03.45 pm - 04.15 pm: Break
- 04.15 pm - 05.45 pm: Session 4 : Labs

Learning outcomes

- Students will develop an understanding of Bayesian Estimation Methods of DSGE models.
- Students will get familiarized with Bayesian Estimation Methods of DSGE models within MATLAB.
- Students will be able to conduct independent estimation of DSGE models using Bayesian methods within MATLAB.
- Please bear in mind: this is not an exhaustive treatment of Bayesian estimation methods. But you will be equipped with the necessary tools to start conducting independent research.
- More (a lot more !) independent work is needed to fully master these techniques.

Hopefully by the end of the day you will get the following joke...



From Calibration to Systems Estimation

- Up to now we have used **calibration** to pin down parameters in the model.
- We have used the deterministic **steady state** to **solve** for parameter values that result in observed long-run outcomes for macro-variables such as hours worked, the great ratios (consumption, investment shares given a government spending share) and the real interest rate.
- For example, from the notes (section 3.5) you have that in the RBC model the real interest rate can be calibrated as $R = \frac{1}{\beta}$.
- A further example of parameter calibration was given by the paramter $\varrho = \frac{(1-H)\alpha}{(1-H)\alpha+c_y H}$, for given α, H and c_y.

From Calibration to Systems Estimation

- This **first-moment matching** can be extended to a **second-moment matching** of variances, correlations and auto-correlations to calibrate shock processes.
- Current practice in empirical macroeconomics is to replace this **informal moment matching** with **formal systems estimation**.
- Maximum Likelihood (ML), General Method of Moments (GMM) and Bayesian estimation are three widely used **systems estimation methods** for DSGE models.
- Distinguish between **full** and **limited information methods**.
- **Full information methods**: Maximum Likelihood, utilize full likelihood of the DSGE model.
- GMM is a **limited information method**: only uses few equations. Other (limited information) approaches: IRFs matching see Christiano *et al.* (2005). Minimise difference between IRFs to a monetary policy shock obtained from the model and those from a VAR in terms of the observables.
- Limited information methods fail to satisfy likelihood principle (all information in an experiment is contained in the likelihood of parameters).

Key Readings

- Economic Dynamics in Discrete Time by Miao chapters 12 and 15
- Structural Macro-econometrics by DeJong and Dave - chapters 6–9
- The Econometrics of DSGE Models by Fernandez-Villaverde, J.
- Bayesian Estimation of DSGE Models, by Herbst and Schorfheide
- An Introduction to Graphs in Dynare by J. Pfeifer
- A Guide to Specifying Observation Equations for the Estimation of DSGE Models by J. Pfeifer

Taking models to the data

- DSGE models focus on explaining **business cycle fluctuations** around **steady state** or exogenous balance growth values $\Rightarrow$ model **solutions** then imply **stationarity** of the variables
- Actual macro data, however, exhibits both **trends** and **cycles** $\Rightarrow$ we need to **transform** the data to fit properties of the model
- Once **trends** have been satisfactorily **removed**, one can then focus on **cyclical behaviour**
- A key step in the setting up the model for estimation in Dynare is to reconcile model variables and observed variables (data) through '**measurement equations**'
- DSGE models can be estimated with a either **two-step approach**: data is first detrended/filtered (or in our case first-differenced) or demeaned if stationary and then structural parameters are estimated
- Or a **one-step approach**: common trend and means are estimated as part of the system (what we do).

Data

- In Appendix 6 of the notes you can find all the data sources and transformations used here which follow closely Smets and Wouters (2007)
- In today's codes folder **Data Preparation** you can find the raw data file *raw_data.mat*, and an m-file *filtered_data.m* that filters output using different filter, and saves the filtered data in *us_data.mat* which we will be using to estimate the NK model.
- The program will compare **linear trend removal** (LT),the **Hodrick-Prescott filter** (HP) and **first-differencing** (FD) and reproduce the Figure 3.4 in De Jong and Dave (2007) in Chapter 3 which you should read.
- Our estimation uses **output**, **inflation** and **nominal interest rate** data to estimate the output **trend** and **means** of inflation and nominal interest rate **together** with all the parameters of the model.
-

Overview of Bayesian Estimation

Bayesian analysis requires:

- Initial information $\Rightarrow$ Prior distribution
- Data $\Rightarrow$ Likelihood density or the probability of observing the data given the model and parameters
- Prior and Likelihood $\Rightarrow$ Bayes theorem $\Rightarrow$ Posterior distribution
- Posterior distribution used for confidence intervals for parameters and impulse responses.
- The posterior distribution also provides information regarding **identification** of parameters - how much information does the data provide on parameters?

Bayesian Estimation Versus Maximum Likelihood

- Bayesian estimation is a **full information** systems estimation method (much like ML)
- It is a "hybrid" approach between informal calibration and ML
- In the absence of prior information it converges to ML. If we are sure the priors are correct we are back to calibration.
- It uses **prior information** to identify key structure parameters - use of additional sources of information
- The likelihood surface can be **flat** (or almost flat) in some directions $\Rightarrow$ priors add 'curvature' to likelihood
- The Bayesian approach allows straightforward facilities for the construction of confidence intervals for parameter estimates and impulse responses, forecasting and model comparison.

Bayes Rule

- Bayesian analysis is based on a few simple rules of probability
- **Some notation**: Suppose A, B are random variables (or events), then

$$\begin{aligned} \text{probability of event A} &\equiv p(A) \\ \text{probability of A and B} &\equiv p(A, B) \text{ or } p(A \cap B) \\ \text{probability of A given B} &\equiv p(A|B) = P(A) \text{ if A, B are independent} \end{aligned}$$

- Then by definition of conditional probability

$$p(A|B) \equiv \frac{p(A, B)}{p(B)}$$

- Reversing the roles of events A and B , we also have

$$p(B|A) \equiv \frac{p(A, B)}{p(A)}$$

- Equating these expressions and rearranging, we get the **Bayes rule**:

$$p(B|A) = \frac{p(A|B)p(B)}{p(A)}$$

The Bayesian Approach to Econometrics

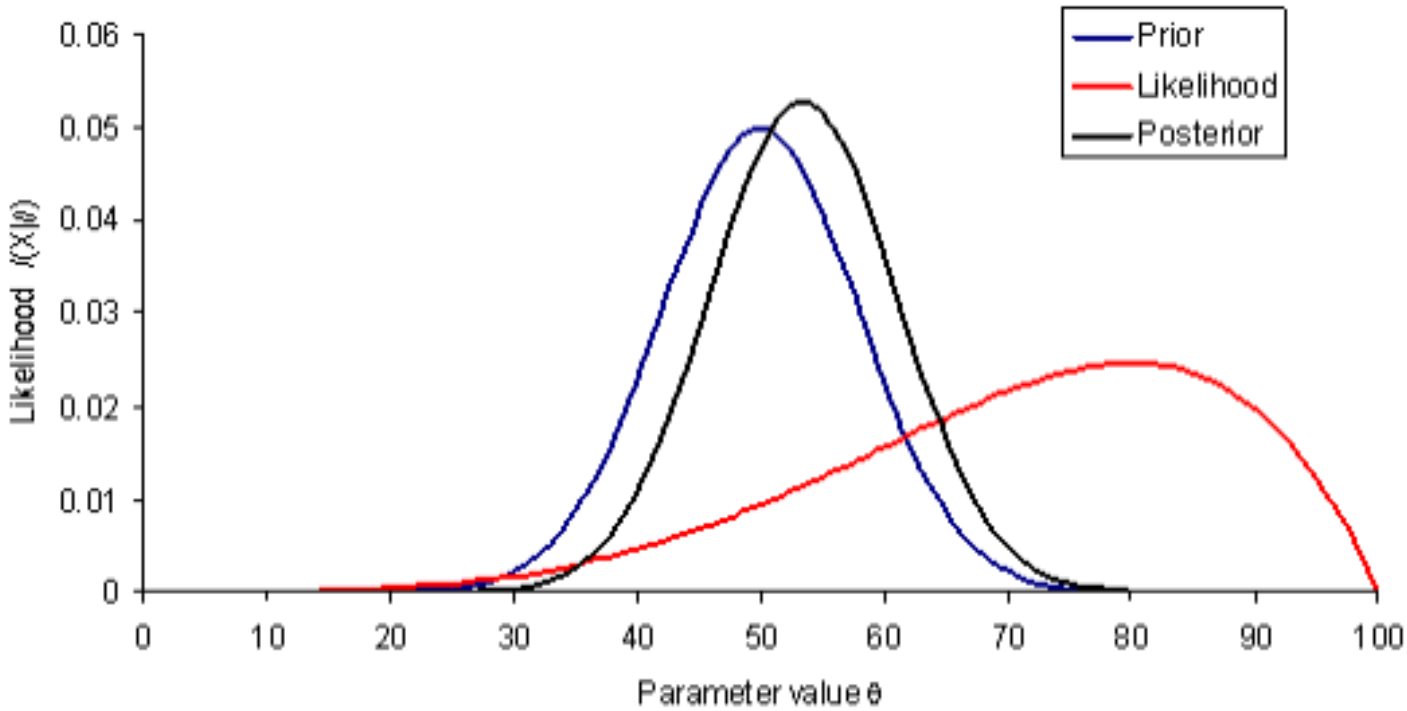
- We want to use data (say T data points $y^T \equiv (y_1, y_2, \cdots y_T)$) to learn about the model's parameters (say n parameters, $\theta = (\theta_1, \theta_2, \cdots \theta_n)$)
- A Bayesian approach does just that: replacing B by θ and A by y

$$p(\theta|y) = \frac{p(y|\theta)p(\theta)}{p(y)}$$

- Our focus is on $p(\theta|y)$: given the data (y), what can we tell about θ?
- Main difference: classical (frequentist) econometrics treats θ as some unknown fixed value(s), whereas Bayesian econometrics assumes that, if θ is unknown, then it should be expressed using rules of probability (i.e., θ is effectively a random object)
- Noting that we're interested in θ, we can drop $p(y)$, so

$$p(\theta|y) \propto p(y|\theta)p(\theta) \quad \text{(posterior kernel)}.$$

- ML maximizes $p(y|\theta)$ wrt θ; Bayesian estimation maximizes $p(\theta|y)$. The prior $p(\theta)$ gives the surface **more curvature**.



Likelihood = red line, Prior = blue line and Posterior = black line. Example of how prior adds more curvature to likelihood.

The Bayesian Approach to Econometrics

- $p(\theta|y)$: **posterior density**
  - summarises what we know about θ after (hence *posterior*) seeing the data
- $p(y|\theta)$: **likelihood density** given the model parameters - also denoted as $L(\theta; y)$
- $p(\theta)$: **prior density**
  - contains all relevant information about θ that does not depend on the data, i.e. what we know about θ *prior* to seeing the data
- $p(\theta|y) \propto p(y|\theta)p(\theta)$ is like an **updating rule**: the data allow us to update our priors about θ, resulting in the posterior, which combines data and non-data information
- The likelihood density is computed using the **Kalman Filter** which is a recursive forecasting procedure for the unobserved states given the observables in the linear state space form (see Miao, ch. 10).

Dynare Steps in the Computation - Bottom up

Below is the Dynare routine for estimation, we will focus on steps **1-6** for now.

1. Solves the model for a particular parameter vector θ. Currently this is a first-order (linear) solution
2. Evaluates the likelihood density $p(y \mid \theta)$ using the linear Kalman filter and assuming Gaussian shocks (useful assumption as first and second moment pin down the distribution of the process)
3. Maximizes $p(y|\theta)p(\theta)$ numerically to arrive at the mode of θ (repeating 1 and 2 each time)
4. Computes the Hessian H:

$$H = \left(\frac{\partial^2 \log(p(y|\theta)p(\theta))}{\partial\theta\partial\theta'} \right) \tag{1}$$

5. The estimate of covariance matrix of the parameters is the inverse of -H at the mode (see Miao chapter 15, page 403).
6. Output is reported at this stage (the prior mean, the estimated mode, its standard deviation and a t-test). The user can stop here.
7. Proceeds to the computation of the posterior distribution using MCMC **if -H is positive definite**

Quick MLE review : a pencil and paper example I

Suppose that Y is generated by the following process:

$$Y = \theta + N(0, 1)$$

where θ is our parameter vector.
This process has likelihood:

$$L(Y; \theta) = \frac{1}{\sqrt{\pi}} e^{-\frac{(Y-\theta)^2}{2}}$$

and minus of the log-likelihood:

$$-log(L(Y; \theta)) = \frac{(Y-\theta)^2}{2} + log(\pi)/2$$

Note : -log likelihood usually is simpler to deal with algebraically.

Quick MLE review : a pencil and paper example II

- The Jacobian of the log-likelihood denoted with $\nabla_\theta log(L(Y;\theta))$ corresponds to $\frac{\partial log(L(Y;\theta))}{\partial \theta}$, i.e. the vector of first partial derivatives of the log likelihood wrt θ.
- In this very simple case (1 parameter) : $\nabla_\theta log(L(Y;\theta)) = -\theta + Y$
- Set $\nabla_\theta log(L(Y;\theta)) = 0$, to find that the MLE estimate is $\hat{\theta} = Y$
- Question: is this a (local) maximum? Construct the Hessian matrix.
- The Hessian is the matrix of second partial derivatives, denoted as $H_{\theta=\hat{\theta}} = \frac{\partial log(L(Y;\theta))}{\partial \theta \partial \theta'}$.
- For a (local) maximum, the Hessian needs to be negative definite at $\hat{\theta}$, which indicates local concavity (MATLAB will check this for you).
- In this simple case: $H_{\theta=\hat{\theta}} = -1$. Hence we have a maximum.
- Provided some regularity assumption hold, MLE is consistent and efficient (achieves the Cramér–Rao lower bound when the sample size tends to infinity.
- $var(\hat{\theta}) = I(\theta)^{-1}$, where $I(\theta) = -nE[\frac{\partial log(L(Y;\theta))}{\partial \theta \partial \theta'}]$, where $I(\theta)$ is the Fisher information matrix.

A Digression on Unconstrained Optimization

- From Simon and Blume, Mathematics for Economists, 1994, page 396 we have the following necessary and sufficient conditions for the maximum x^* of a function $F(x)$ of n variables $x = (x_1, \cdots, x_n)$
- **First Order Condition**:

$$\frac{\partial F}{\partial x_i} = 0 \text{ for } i = 1, \cdots, n$$

- **Second Order Condition**: The **Hessian** (H) below is **negative definite**:

$$H = \frac{\partial^2 F}{\partial x \partial x'} \equiv \begin{bmatrix} \frac{\partial^2 F}{\partial x_1^2} & \cdots & \frac{\partial^2 F}{\partial x_n \partial x_1} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \frac{\partial^2 F}{\partial x_1 \partial x_n} & \cdots & \frac{\partial^2 F}{\partial x_n^2} \end{bmatrix} \text{ evaluated at } x = x^*$$

- This is an n-variable generalization of the condition for the maximum of a function $F(x)$ of **one variable** x: $\frac{dF}{dx} = 0$, $\frac{d^2F}{dx^2} < 0$.

Constructing the posterior : a pencil and paper example

To get the conditional distribution of the parameters given the data (posterior) we need the distribution of the parameters in the absence of any data. This is called the prior. For this example take the prior of θ to be $N \sim (\mu,\sigma)$.

Prior:

$$p(\theta) = \frac{1}{\sqrt{\pi}} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}$$

Likelihood:

$$L(Y;\theta) = \frac{1}{\sqrt{\pi}} e^{-\frac{(Y-\theta)^2}{2}}$$

Combine **prior information** with the **data** to obtain the **posterior (kernel)** :

$$p(\theta|y) \propto L(Y;\theta)p(\theta) \propto \frac{1}{\sqrt{\pi}} e^{-\frac{(Y-\theta)^2}{2}} \frac{1}{\sqrt{\pi}} e^{-\frac{(\theta-\mu)^2}{2\sigma^2}}$$

up to a constant:

$$p(\theta|y) \propto L(Y;\theta)p(\theta) \propto e^{-\frac{(Y-\theta)^2}{2} - \frac{(\theta-\mu)^2}{2\sigma^2}}$$

In this case, it is straightforward to compute first and second moments from the posterior analytically. Question: is this always the case?

Computation

- Again our focus is on the posterior distribution of $p(\theta|y)$ that summarise what we know about θ conditional on y, this distribution is characterized by (posterior) means, medians, modes, etc (and respective standard deviations)
- Knowing this allows Bayesian inference expressed as $E[g(\theta)|y]$, where $g(\theta)$ is a function of interest (e.g. mean, variance, etc.):

$$E[g(\theta)|y] = \int g(\theta)p(\theta|y)d\theta \tag{2}$$

- Bar a few exceptions, it is often impossible to to evaluate the integral analytically $\Rightarrow$ recur to simulation methods (Monte Carlo), drawing from the posterior density $p(\theta|y)$.
- As the number of draws (N) increases, then we can invoke the Law of Large Numbers and the Central Limit Theorem

Example

- Say, we are interested in characterizing the posterior distribution by its mean and variance. We need to compute the mean:

$$E(\theta|y) = \int \theta p(\theta|y) d\theta$$

- and the variance

$$Var(\theta|y) = E(\theta^2|y) - [E(\theta|y)]^2$$

$$= \int \theta^2 p(\theta|y) d\theta - [E(\theta|y)]^2$$

- In both cases we need to evaluate an integral that usually cannot be worked out analytically!

Solution

- Solution, focus on the integral of the form

$$E(g(\theta)|y) = \int g(\theta)p(\theta|y)d\theta$$

- If we had i.i.d. draws from the posterior, we could simply use a law of large numbers:

$$E(g(\theta)|y) = \int g(\theta)p(\theta|y)d\theta \approx \frac{1}{S}\sum_{s=1}^{S} g(\theta_s) = \hat{g}(s)$$

- Replace integral by sum over S draws from the posterior distribution $p(\theta|y^T)$
- This is called Montecarlo integration
- Problem : how do we sample from a distribution which is untractable? MCMC - Metropolis Hasting, more in a bit.

Digression: why would we want to focus on the posterior mode instead of the posterior mean ?

- If the posterior distribution is normal, there is no difference between mode and mean.
- Typically the distribution of the posterior is unknown, then we need to pick a measure.
- With irregular (not symmetric) posteriors the mode is a better measure of central tendency.
- **Example**: suppose the posterior follows a binomial distribution with p(x_1)=0.6. The two outcomes are $x_0 = 0$ or $x_1 = 1$. In this case, the mode will be 1 while the mean will be 0.6. Which measure of central tendency describes best the distribution?
- Obviously the mode. In this case the mean does not even capture a plausible value obtainable in the sample (note: in the sample we either observe 0 or 1).

Bayesian Maximum Likelihood

- The most conventional approach to estimation is to maximize the likelihood $L(y|\theta)$
- If we have a data-set of time series data $y^T = \{y_1, y_2, ..., y_T\}$, then using Bayes Theorem it is straightforward to show that

$$L(y_{t+1}|y^t, \theta) = \frac{L(y^{t+1}|\theta)}{L(y^t|\theta)}$$

hence

$$L(y^{t+1}|\theta) = L(y_{t+1}|y^t, \theta)L(y^t|\theta)$$

so that by induction we have

$$L(y^t|\theta) = \prod_{k=2}^{t} L(y_k|y^{k-1}, \theta)L(y_1|\theta)p(\theta)$$

Calculation of the Likelihood Function

For linear models, with Gaussian shocks the model is solved for given θ along the saddle path, and written in **state space form** as:

$$x_{k+1} = Ax_k + B\varepsilon_{k+1} \qquad y_k = Cx_k$$

Note: usually we only observe a subset of state variables x, these observables are stored in the vector y! Thus, we need a filtering procedure: **Kalman Filter**.

The **log-likelihood** is then given by:

$$lnL(y|\theta) = -\frac{Tr}{2} ln(2\pi) - \frac{1}{2}\sum_{k=1}^{T}(det(F_k) + e_k^T F_k^{-1} e_k) + lnp(\theta)$$

where r is the number of measurements at each period, and e_k, F_k are obtained from the **Kalman Filter** recursions

$$\begin{aligned} e_k &= y_k - Cx_{k,k-1}\,; \quad F_k = CP_kC^T \\ x_{k+1,k} &= Ax_{k,k-1} + AP_kC^T F_k^{-1} e_k \\ P_{k+1} &= AP_kA^T - AP_kC^T F_k^{-1} CP_kA^T + Bcov(\varepsilon)B^T \end{aligned}$$

subject to the initial conditions $x_{1,0} = 0$, and P_1 being the solution of the **Lyapunov equation** $P_1 = AP_1A^T + Bcov(\varepsilon)B^T$.

First Stage Estimation - Maximizing the Likelihood

- The first thing done by Dynare in the estimation stage is to maximize the Bayesian likelihood
- Recall we obtain $L(y^T|\theta)$ via the KF
- This yields the ML estimates, with parameter standard errors obtained from the information matrix I_N, which corresponds to the Cramer-Rao lower bound
- For a given model M_i, we can write the Bayesian likelihood as $L(y^T|\theta, M_i)$, and the marginal likelihood of model M_i is given by $\int L(y^T|\theta, M_i)d\theta$ (e.g. by integrating out the parameter vector θ)
- Different models M_i may have some parameters fixed at 0, or estimated under different information sets. The econometrician will prefer the model with lowest marginal likelihood
- The Laplace approximation to the log marginal likelihood is given by $\frac{N}{2}ln(2\pi) + lnL(y^T|\theta^*, M_i) - \frac{1}{2}ln(det(I_T))$ where I_T is the information matrix evaluated at the maximum θ^*.

Problems with Maximizing the Likelihood

- Recall that we are usually interested in quantities such as

$$E[g(\theta)|y] = \int g(\theta)p(\theta|y)d\theta$$

- e.g. mode, mean, variance, etc
- For complex models, with nonlinear effects of parameters, finding the mode is not straightforward
- The main problem is that the algorithm may have converged to a local maximum of the likelihood
- Even changing the initial parameter values is not an assured method of hitting a global maximum
- Instead it is useful to sample the likelihood function over a large range of parameter draws
- The objective when performing this sampling is to ensure that the frequency of sampling a draw should exactly match the probability of that draw
- The most commonly used method is the MCMC algorithm

Recall : Dynare Steps in the Computation

1. Solves the model for a particular parameter vector θ. Currently this is a first-order (linear) solution
2. Evaluates the likelihood density $p(y \mid \theta)$ using the linear Kalman filter and assuming Gaussian shocks
3. Maximizes $p(y|\theta)p(\theta)$ numerically to arrive at the mode of θ (repeating 1 and 2 each time)
4. Computes the Hessian H:

$$H = \left(\frac{\partial^2 \log(p(y|\theta)p(\theta))}{\partial\theta\partial\theta'} \right) \tag{3}$$

5. The estimate of covariance matrix of the parameters is the inverse of -H at the mode (see Miao chapter 15, page 403).
6. Output is reported at this stage (the prior mean, the estimated mode, its standard deviation and a t-test). The user can stop here.
7. Proceeds to the computation of the posterior distribution using MCMC **if -H is positive definite**

MCMC Metropolis-Hastings algorithm

- MCMC methods: samplers wandering over the posterior, taking most draws from high probability areas
- Not easy to draw directly from $p(\theta|y)$ - we need methods that work well for any case $\Rightarrow$ MH MCMC, drawing from a candidate ("transition") distribution
- "Markov Chain" bit: a given draw θ^* depends on $\theta^{(s-1)}$
- "Monte Carlo" bit: θ^* is drawn at random from a candidate proposal (or transition) distribution $\alpha(\theta^{(s-1)}, \theta^*)$, and then (see below) $\theta^{(s)}$ is either θ^* or $\theta^{(s-1)}$
- Idea: to specify a transition density for a MC such that, starting from some initial value θ_0 and iterating a number of times, we produce a limiting distribution which is the target distribution we need to sample from.
- Usually one discards the first several thousand draws to ensure that the sequence is not dependent on the starting draw

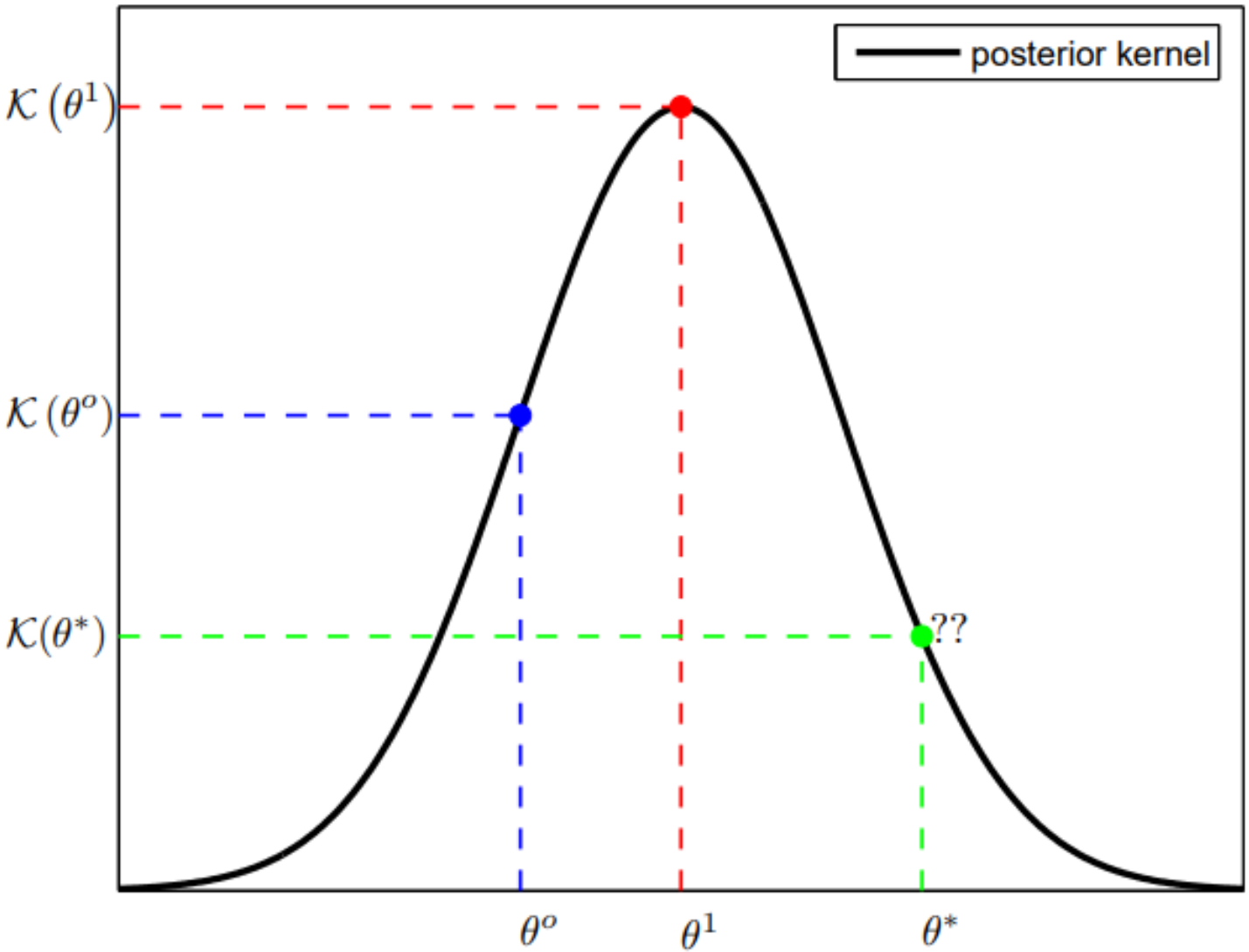
MCMC Metropolis-Hastings Algorithm - cont.

- Intuition: we want to sample from the region with highest posterior probability, but we also want to visit the whole parameter space as much as possible
- given that there is a discrepancy between the candidate and target densities, the MCMC will not take the correct draws $\Rightarrow$ MH algorithm corrects this by calculating an acceptance probability and eventually discarding some draws
- Because it is difficult to find a good candidate density, we usually employ a Random Walk Chain MH algorithm

$$\theta^* = \theta^{(s)} = \theta^{(s-1)} + z \tag{4}$$

- sampler wanders in random directions, thus visiting most of the parameter space
- $z \sim N(\theta_{s-1}, c\hat{\Sigma}_\theta)$, key choice is the **scaling parameter** c

A walk down the posterior...



MCMC Metropolis-Hastings Algorithm - cont.

1. Choose starting value θ^0 (usually posterior mode from first stage and loop over steps 2 to 4)
2. Draw a candidate from the jumping distribution θ^* in (4)
3. Given θ_i^*, solve the model and use the KF to compute $p(\theta_i^*|y)$
4. For each draw i, accept $(\widehat{\theta_i} = \theta_i^*)$ with probability r; reject the draw (i.e. $\widehat{\theta_i} = \theta_{i-1}$) with probability $1 - r$
   - The acceptance probability of each new draw is defined by:

$$r = min\left[\frac{p(\theta_i^*|y)}{p(\theta_{i-1}|y)}, 1\right] \tag{5}$$

   - Notice that the lower $\frac{p(\theta_i^*|y)}{p(\theta_{i-1}|y)}$, the higher chance of rejection as r is proportional to this ratio.
   - Back to step 2
5. After repeating these steps so that sufficient draws are generated, build posterior from empirical distribution

Some remarks

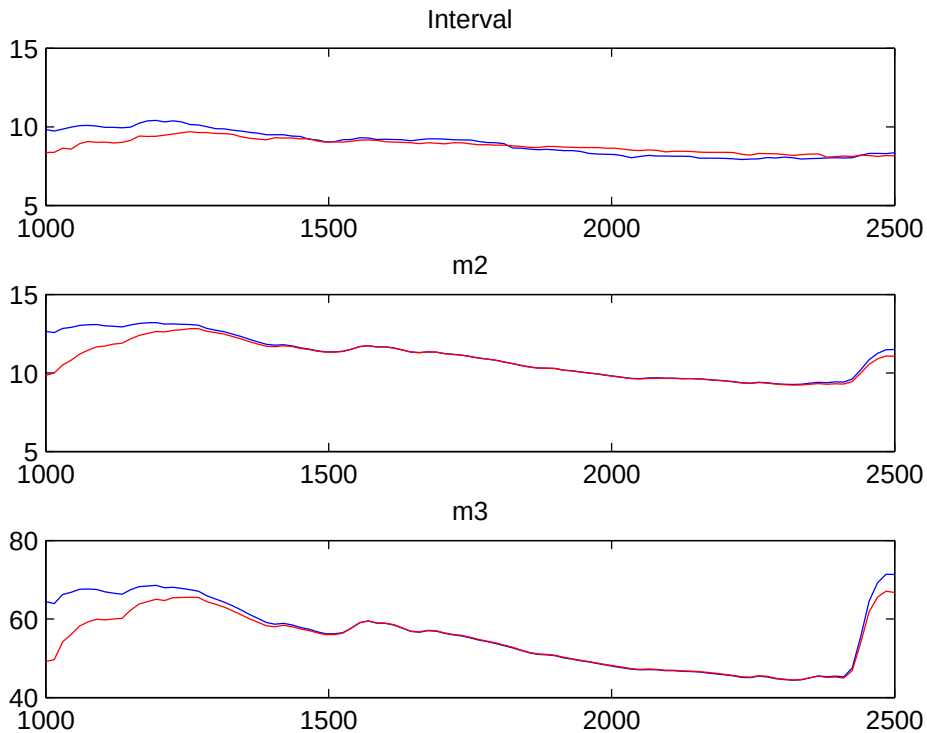
- This acceptance rule ensures that the entire domain is visited
- The **acceptance rate** is dependent on the choice of c:
  - if the jump is too small, the fraction of accepted draws is too high, i.e. the chain will get stuck around a local maximum and won't visit the tails
  - if c is too large, the acceptance rate will be low, i.e. the draws will come from regions of lower probability and the chain is stuck in the tails
- Acceptance rate: ideally 20-40% $\Rightarrow$ each move goes a reasonable distance in parameter space, but not so low as to reject too frequently

Testing for Convergence

- Testing for convergence of the posterior distribution is notoriously difficult, and Dynare utilises some indicative statistics, summarised by diagrams, as recommended by Brooks and Gelman (1998b). These diagrams are made up of
  - 3 multivariate figures (%80 interval, m2 and m3) representing convergence indicators for all parameters considered together (the posterior kernel), generated if mh_nblocks$>$ 1 and mh_repl$>$ 1000.
  - 3 figures for each parameter, representing univariate convergence indicators (%80 interval, m2 m3), triggered if mh_nblocks$>$ 1 and mh_repl$>$ 2000
  - For each of these indicators, dynare computes within-sequence statistics (red) and statistics based on pooled draw from all sequences (blue)
  - If the chains have converged, the two lines should stabilize horizontally and should be close to each other.
- Dynare will generate independent chains that should converge to the same ergodic distribution, provided that a sufficient number of draws is generated

Example of Convergence - Multivariate Measures

The diagnostics below, for the NK linear model estimated later, are generated by the estimation command if mh_replic is larger than 2000, mh_nblocks=2 and if option '*nodiagnostic*' is not used.
'Interval' refers to the 80% quantile range, 'm2' and 'm3' refer to the second and third moments



What to do about Lack of Convergence

- Convergence is a notorious problem for MCMC, and the only theorem is that if convergence occurs, it is to the correct distribution.
- Crucially, one would want multivariate convergence
- Improving convergence could be done in one of two ways:
  - Increase the number of draws
  - Increase the 'scale factor' for the Monte Carlo part. This increases the range of search but at the expense of reducing the acceptance ratio

Marginal Likelihood and model comparison - Geweke (1999)

- The computations thus far do not rely on the marginal likelihood $p(y)$, which appears in the denominator of Bayes Theorem.
- Recall: we focused on the posterior kernel $p(\theta|y) \propto p(\theta)p(y|\theta)$
- Marginal likelihoods play an important role in assessing the relative fit of models because they are used to turn prior model probabilities into posterior probabilities.
- The most widely used marginal likelihood approximation in the DSGE model literature is the modified harmonic mean estimator proposed by Geweke (1999).
- This estimator is based on the identity:

$$\int \frac{f(\theta)}{p(y)} d\theta = \int \frac{f(\theta)}{p(\theta)p(y|\theta)} p(\theta|y) d\theta \tag{6}$$

- where : $f(\theta)$ has the property that $\int f(\theta)d\theta = 1$.
- The identity is obtained by rewriting Bayes Theorem, multiplying both sides with $f(\theta)$ and integrating over θ.

Marginal Likelihood and Model comparison - cont'd

- Suppose we have two models, Model 1 and Model 2.
- Then, if marginal data density Model 1 greater than Model 2 conclude that model 1 fits better than model 2.
- Can use this to compare across two different models, or to evaluate contribution to fit of various model features: habit persistence, adjustment costs, etc.
- In Dynare, you will see after mode finding: Log data density [Laplace approximation] is 1186.624917. Another estimate is provided after MCMC algorithm.
- The Laplace approximation of marginal density is stored in :
- **oo_.MarginalDensity.LaplaceApproximation**
- The Modifed Harmonic Mean of marginal density is stored in:
- **oo_.MarginalDensity.ModifiedHarmonicMean** which is used with
- mh_replic>0 or load_mh_file option.

Measurement Equations

- Inflation is in gross terms and is computed as the ratio of a price index P in two subsequent periods: $\Pi_t^{obs} = \frac{P_t}{P_{t-1}}$
- In net terms, inflation can be approximated by the log of gross inflation, $\pi_t^{obs} = \log \Pi_t^{obs}$ or, equivalently, $\pi_t^{obs} = \log P_t - \log P_{t-1}$
- Interest rates (unlike inflation) are measured net and in annual terms, while models are usually in quarterly terms - if the raw data is in annualised percentage points, then you typically divide the data by 400
- If in a two-step procedure using a filter, the resulting series will have a zero mean, whereas with first-differences you will be left with the series' average growth rate
- If the model is entered in log-linearized form, then the filtered variables correspond to model variables, as deviations from the steady state

Measurement Equations

We estimate the linear NK model using SW data with output in first-differences

Observables: output, inflation and interest rate

The corresponding measurement equations for the 3 observables are:

$$\begin{bmatrix} D(\log GDP_t) * 100 \\ \log(GDPDEF_t / GDPDEF_{t-1}) * 100 \\ FEDFUNDS_t / 4 * 100 \end{bmatrix} = \begin{bmatrix} Y_t - Y_{t-1} + \text{trend growth} \\ \Pi_t + \text{constant}_{\Pi} \\ R_{n,t} + \text{constant}_{R_n} \end{bmatrix}$$

- Note: the quarterly **trend growth rate** in real GDP; the quarterly **steady-state inflation rate** and the **steady-state nominal interest rate** are estimated together with the other parameters.
- Sample: 1984:1-2008:2 which starts at observation 144 in the data file.
- There is a pre-sample period of 4 quarters so the observations actually used for the estimation go from 148:245.

Priors

- For Bayesian estimation we need parameter 'priors' (location) and their distributions (shape)
- Where do we get the priors from? Micro estimates, calibration, existing studies...
- Typically the prior mean is centered around calibrated value. Std. errors reflect subjective or objective (to cover the range of existing estimates)
- The shape of the distribution
- General guidance - *inverse gamma* distributions are used as priors when non-negativity constraints are necessary, *beta* distributions for fractions or probabilities, *normal* distributions are used when more informative priors seem to be necessary (*uniform* or 'flat' priors if there is little information about the parameter)
- Options in Dynare are normal, gamma, beta, inverse gamma and uniform distribution

Estimation in Dynare

- Observed variables are declared after *varobs* and must be available in the data file (can be a .mat file or a .m file or even an .xls file) with the same name
- Estimated parameters are declared in an *estim_params; ... end;* block.
- For each estimated parameter, declare the initial value and, optionally, a lower and upper bound for the ML estimation.
- If their prior distributions are further declared, Dynare chooses to perform the Bayesian estimation
- Computing the estimation is triggered by *estimation* and required option in brackets after *estimation*: *datafile=FILENAME*

Estimation in Dynare: options

Commonly used options are:

- *prefilter*: a value of 1 means that the estimation procedure will demean the data
- *first_obs*: specifies the first observation to be used
- *mode_compute*: specifies the optimizer. For example: 0: switch mode computation off; 1: fmincon; 4: csminwel
- *mh_replic*: sets the number of replications for MH algorithm - longer chains are more likely to have converged
- *mh_drop*: sets the percentage of discarded draws
- *mh_jscale*: specifies the scale to be used for the jumping distribution in MH algorithm
- *presample*: number of initial periods that don't enter into likelihood computation (to initialize Kalman)

Some other useful options

- *mode_file=FILENAME_mode*: reloads the computed posterior mode (stored in FILENAME_mode.mat in the working dir.)
- *load_mh_file*: recovers the estimates using the existing MH replications
- *mode_check*: plots the objective function around the computed mode
- *forecast = INTEGER*: posterior distribution of out-of-sample forecast
- *moments_varendo*: triggers the computation of the posterior distribution of the theoretical moments of the endogenous variables

Some Dynare Tricks

- Even though we are interested in estimating (some) parameters, it is still useful to fully calibrate the model and initially keep the sequence '*steady*; *check*; *stoch_simul*;'
- Before estimating, you want to be sure that the steady state can be computed, the BK conditions are met and that calibrated IRFs make sense.
- Recall that from BK conditions:
  - **Stability**: This implies that when the economy is pushed off its steady state following a shock, it converges back to it.
  - **Determinacy**: this implies that when a shock displaces the economy from its steady state, there is a unique path leading back to equilibrium.
- Occasionally it is convenient to define 'model-local variables' - usually composite parameters: these are define inside the 'model' block, preceded by the '#' symbol - however, these are not estimated!

Estimating the Linearized NK model

With the model in state space form with observable variables, the steps of the dynare procedure are as follows:

- Dynare computes log-likelihood functions running the Kalman recursion as discussed previously.
- The mode of the posterior is estimated using a numerical optimisation procedure
- Chris Sim's csminwel (mode compute 4) is quick but may not produce the required positive definite minus of the Hessian
- If this is the case mode compute 6 should produce a positive definite result but it takes a lot of time!
- The posterior distribution is obtained with MCMC-MH using the inverse Hessian at the estimated posterior mode as the covariance matrix of the jumping distribution
- Fine-tuning *mh_jscale*; *mh_replic*; *mh_drop* in order to remove any dependence of the chain from its starting values

Dynare Files for Estimating the NK model I

The .mod files for this analysis are:

- **NKlinearEst_Identification.mod**- provides results for identification
- **NKlinearEst.mod** provides full estimation results for the mode and the posterior distribution with 10,000 mcmc draws
- **NKlinear_Est.mod** has been estimated using an initial saved mode-compute 6 which is designed to give a negative definite Hessian. The mode file saved as **NKlinear_Est_mode_saved**
- Estimation results for 100,000 mcmc draws are reported in a separate log file **NKlinear_Est_100Kmcmc_jscale_.40.log**
- The convergence test are shown in the figure **NKlinear_Est_100Kmcmc_jscale_.40.fig**.
- This shows how **more draws** may be necessary to achieve convergence.

Dynare Files for Estimating the NK model II

- **NKlinearEst.mod** then proceeds to extract the mean of the estimates and then computes **impulse response functions (irfs)** using stoch_simul.
- **us_data.mat** is the data file used for all our results
- Note : various files produced by dynare of its own saved results (not indicated as saved) of interest including the filename.log file that saves whatever appear on the screen, the filename.mode file that saves the mode and filename_results.mat that saves all results including the irfs.

Estimation output

- Results from posterior optimization (also for maximum likelihood)
- Marginal data density (modified harmonic mean estimator - Geweke, 1999)
- mean and confidence interval from posterior simulation
- Graphs with prior, posterior and mode
- Graphs of smoothed shocks and smoothed observation errors
- Convergence diagnostics (MCMC replications)
- Graphs of posterior IRFs (optional)
- All results are stored in **FILENAME_results.mat** (in particular, in structure array *oo_*)
- All MH draws are saved in the subfolder under the path: *FILENAME\metropolis* – post. median

A summary of Dynare Bayesian estimation

1. Transform the actual data to fit properties of the model
2. Specify prior distributions
3. Dynare computes the log-likelihood numerically via the Kalman filter
4. Finds the maximum of the likelihood and posterior mode
5. Draws posterior sequences and simulates posterior distribution with Metropolis algorithm
6. Computes various statistics on the basis of the posterior distribution (post. moments)
7. Estimates the posterior marginal density to compare models
8. One remaining issue: **Identification**

Identification

- It is necessary to confront the question of parameter identifiability in DSGE models before taking them to the data.
- As model or parameter identification is a prerequisite for the informativeness of different estimators, and their effectiveness when one uses the models to address policy questions.
- The sources of identification failure could be :
  - **marginalisation** (from the model structure: i.e. mapping the deep parameters to the reduced form coefficients of the solution and mapping the solution to the population objective function)
  - or **lack of information** (from the data: i.e. mapping the population to the sample objective).
- The lecture and the provided notes provide an overview that discusses about the former.
- The standard remedy suggested by most empirical DSGE literature is to fix some (potentially non-identifiable) parameters and re-maximize with the parameters that are well-identified. This approach can be however problematic see Canova and Sala (2009).

Identification in Dynare

- The default syntax for the identification procedures in DYNARE:

```
varobs dy pinfobs robs;
identification(advanced=1);
```

- point identification at the prior mean
- the MC exploration using the draws from the prior distribution
- identification strength measured at the mean and weighted by the prior standard deviation.

Identification Example: NK Model

- The output of this procedure then reveals whether or not there are identification problems with parameter estimation:

```
==== Identification analysis ====
Testing prior mean
Testing prior mean  WARNING: Komunjer and Ng (2011) failed: The
are more shocks and measurement errors than observables, this i
not implemented (yet). Skip identification analysis based on mi
state space system. The number of moments with non-zero derivat
is smaller than the number of parameters Try increasing ar = 2
```

- Our model has 4 shocks and 3 observables. The RE solution cannot be invertible (see Fernandez-Villaverde *et al.* (2007)) and the standard perfect information assumption on the part of agents is inconsistent with the information of the econometrician. (See Levine *et al.* (2019)).

Fernandez-Villaverde *et al.* (2007)- The PMIC

The solution to a (log-)linearized RE model has a state space representation:

$$x_{t+1} = Ax_t + B\varepsilon_{t+1} \tag{7}$$

$$y_{t+1} = Cx_t + D\varepsilon_{t+1} \tag{8}$$

where : where x_t is an $n \times 1$ vector of possibly unobserved state variables, y_t is a $k \times 1$ vector of variables observed by an econometrician, and ε_t is an $m \times 1$ vector of economic shocks impinging on the states and observables.

- Under which condition do the shock in the DSGE model map into the shock from a VAR in terms of the observables? i.e. under which conditions does the econometrician have the same info set as agents?
- (Pre-)Condition 1: D is a square matrix and D^{-1} exists. This condition requires the number of economic shocks to be the same as the number of observables ($m = k$) and that the shocks are linearly independent.
- Condition 2: the Poor Man's Invertibility Condition (PMIC). This requires the matrix $F = (A - BD^{-1}C)$ to be stable, e.g. this requires all the eigenvalues of F to be less than 1 in modulus.
- If the PMIC condition is satisafied the econometrician's and the agent's information set are aligned.

Komunjer and Ng (2011)

In the case where the D matrix in is square and D^{-1} exists, the impulse responses to the structural shocks ε_t are given by the following VMA representation if the eigenvalues of A are inside the unit-circle:

$$y_t = D(L)\varepsilon_t = \sum_{j=0}^{\infty} d_j L^j \varepsilon_{t-j} \tag{9}$$

where L is the lag operator $d_0 = D$, $\sum_{j=0}^{\infty} tr(d_j d_j') < \infty$, $d_j = CA^{j-1}B$ for $j \geq 1$. The transfer (impulse response) function associated to this :

$$\mathcal{D}(z) = D + C(I_{n_x} - A)^{-1}B = \sum_{j=0}^{\infty} d_j L^j z^{-j} \tag{10}$$

with spectral densities:

$$\Omega(z) = \Gamma(0) + \sum_{j=1}^{\infty} \Gamma(j) z^{-1} + \sum_{j=1}^{\infty} \Gamma(-j) z^{-1} = \mathcal{D}(z)\Sigma_\varepsilon \mathcal{D}(z^{-1})' \tag{11}$$

Komunjer and Ng (2011)

- When $m > k$, equation 9 is no longer the Wold representation for y_t and ε_t is no longer fundamental.
- More precisely, the transfer function of a non-singular system with more shocks than observables is not left-invertible.
- This means that the shock vector ε_t cannot be recovered from the t-dated history of the vector y_t. In other words, the econometrician information set is narrower than the agent's.
- Dynare is informing you that this situation / type of identification analysis is not (yet) supported in Dynare, so it will skip the checks based on the minimal state space representation.

Identification Example: NK Model

- Continuing with the identification tests:

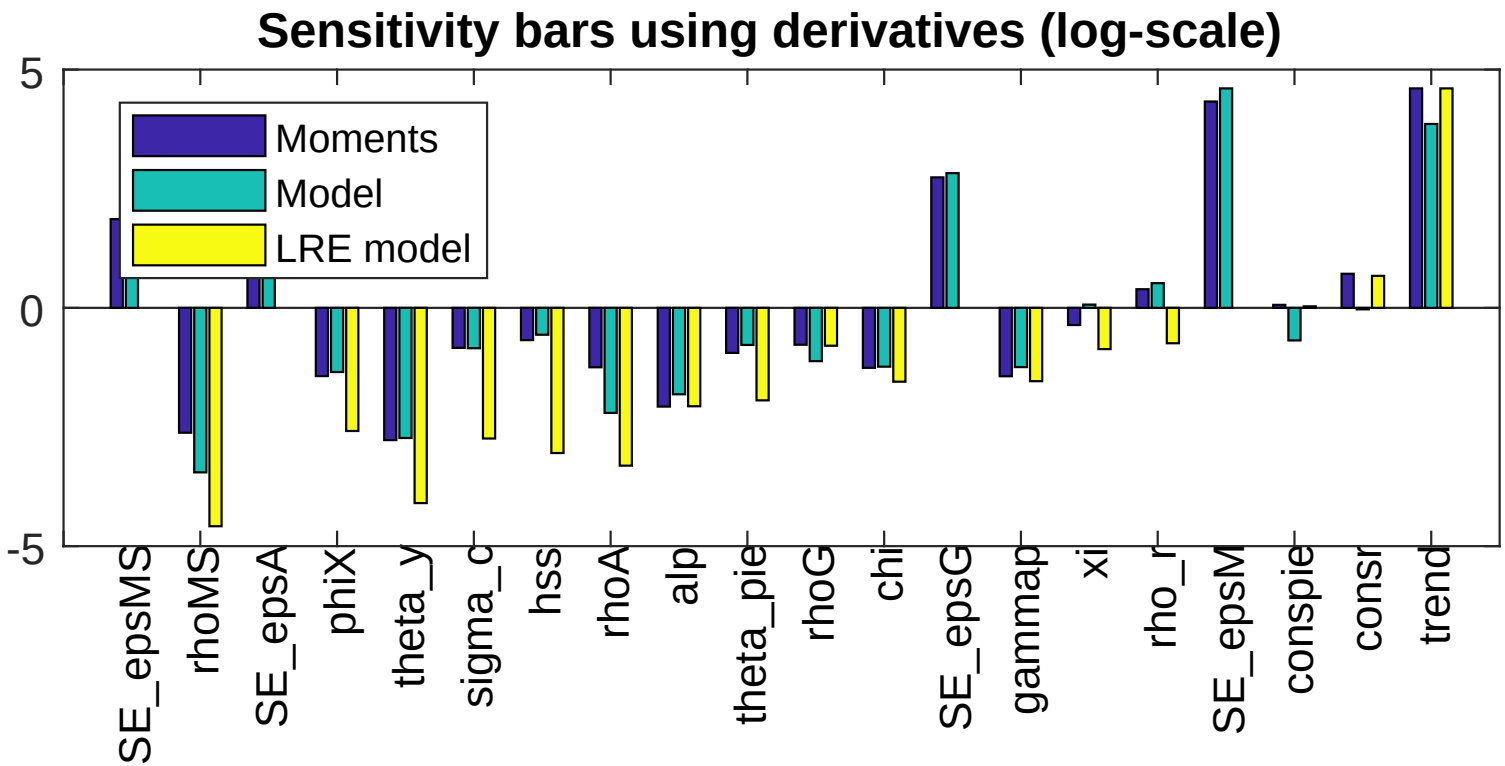
```
==== Identification analysis ====REDUCED-FORM:
All parameters are identified in the
Jacobian of steady state
and reduced-form solution matrices (rank(Tau)
is full with tol = robust).
SPECTRUM (QU AND TKACHENKO, 2012):
All parameters are identified in the Jacobian of mean
and spectrum (rank(Gbar) is full with tol = robust).
MOMENTS (ISKREV, 2010):
All parameters are identified in the Jacobian of
first two moments (rank(J) is full with tol = robust).
```

- If the rank condition fails in any of these two procedures, the procedure indicates which parameters are responsible for identification problems.

Identification Example: NK model

- The following Figure shows an aggregate measure of how changes in the elements of the parameter vector θ impact on the model moments. The impact is measured locally using the Jacobian.
- The problem is that the derivatives are not scale invariant so not easily comparable. For this reason here Dynare uses a normalization/standardization procedure described in Iskrev Ratto (2011).
- Dynare plots here three different measure of sensitivity. The bars depict the norm of the columns of three different standardized Jacobian matrices for the respective parameter shown on the x-axis. The respective Jacobian refer to:
  1. the moments matrix, indicating how well a parameter can be identified due to the strength of its impact on moments
  2. the solution matrices (model), indicating how well a parameter could in principle be identified if all state variables were observed
  3. the Linear Rational Expectation (LRE) model, indicating trivial cases of non-identifiability due for example to the fact that some parameters always show up as a product in the model equations.

Identification Example: NK model



Identification Example: NK model

- To completely rule out a flat likelihood at the local point one can also check collinearity between the effects of different parameters on the likelihood.
- If there exists an exact linear dependence between a pair and among all possible combinations their effects on the moments are not distinct and the violation of this condition must indicate a flat likelihood and lack of identification.
- The details of collinearity analysis require the advanced analysis option $((advanced = 1))$ which prints the results of the brute force search for the groups of parameters whose columns in the Jacobian matrix best explain each column of the Jacobian (i.e. best reproducing the behavior of each single parameter).

Identification Example: NK model

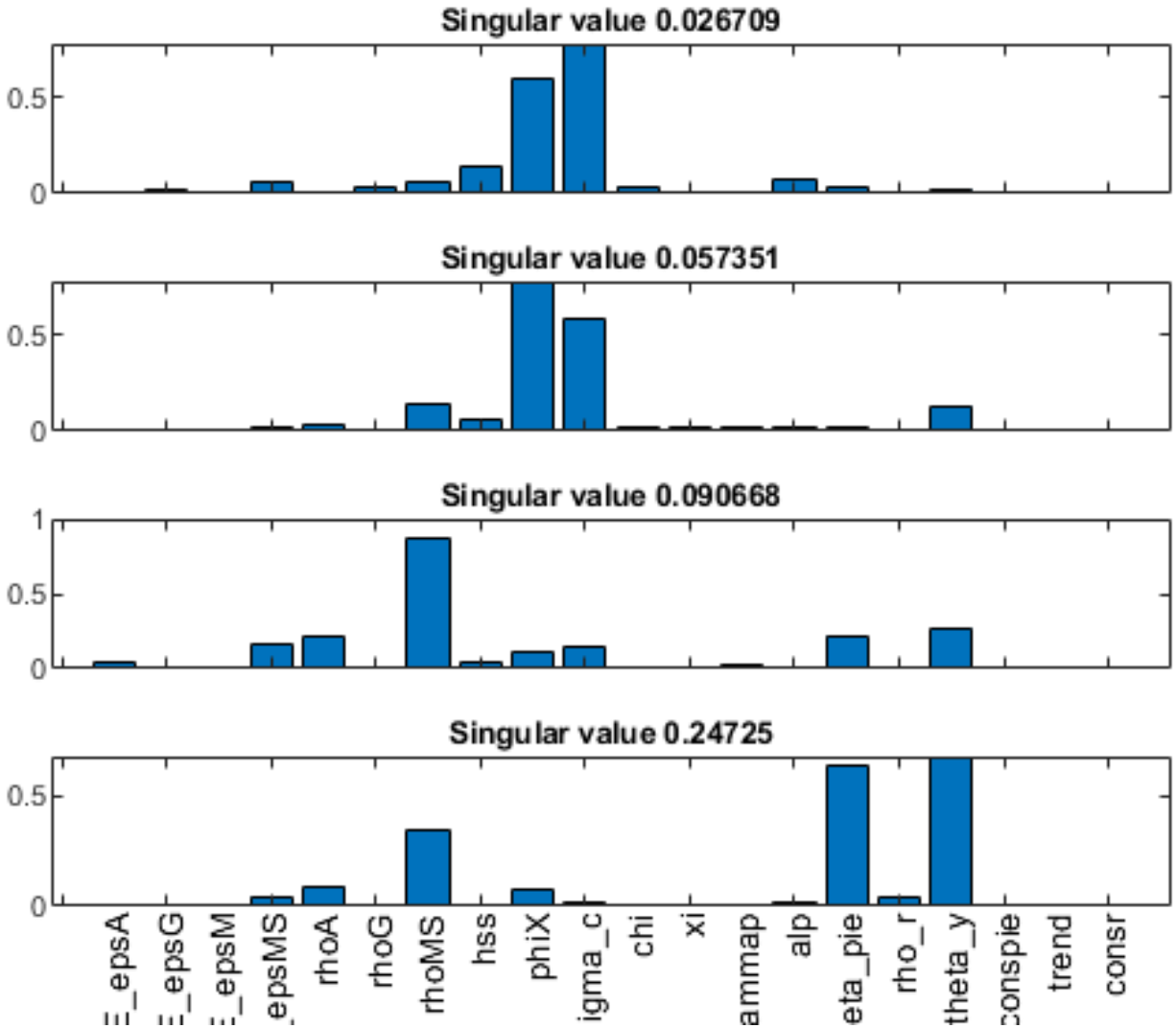
```
Collinearity patterns with 1 parameter(s)

A Selection of High Collinearity patterns with 2 parameter(s)

Parameter [ Expl. params ] cosn
SE_epsA [ SE_epsMS hss ] 0.9978085
SE_epsM [ phiX rho_r ] 0.9656062
SE_epsMS [ SE_epsA hss ] 0.9973889
rhoA [ rhoMS phiX ] 0.9958508
rhoMS [ rhoA phiX ] 0.9956177
hss [ sigma_c xi ] 0.9867869
sigma_c [ SE_epsG hss ] 0.9878160
alp [ SE_epsA hss ] 0.9522590
theta_pie [ hss rho_r ] 0.9985808
rho_r [ hss theta_pie ] 0.9983547
theta_y [ chi rho_r ] 0.9414656
```

Andrle (2010)

- Following Andrle (2010), identification can also be judged from a singular value decomposition (SVD) of the information matrix
- Provides the size of the singular values and the associated eigenvectors (i.e. parameters)
- Parameter combinations associated with the smallest singular values are closest to being perfectly collinear and thus redundant
- Singular value of 0 implies that the parameter is completely unidentified as it is responsible for the information matrix being rank deficient
- Example, a the next figure : consr, conspie and trend are not identifiable.


Singular value 0.026709
Singular value 0.057351
Singular value 0.090668
Singular value 0.24725
epsMS
rhoA
rhoG
rhoMS
hss
phiX
chi
xi
alp
rho_r
trend
consr

Validation based on standard moment criteria

- We now examine the ability of the estimated model to predict second moments (the absolute fit)
- Three sets of second moments:
  - Volatility - Standard Deviations
  - Co-Movement - Cross Correlations
  - Persistence - Autocorrelation
- To generate moments of endogenous variables in Dynare we simply use stoch_simul:
- Uses post-estimation solution based on posterior modes or means of the model to produce the three moment above.

Results and plots

- Again all simulation outputs are stored in the **FILENAME_results.mat** in the working directory $\Rightarrow$ reload it to extract useful information (in the structure array *oo_*)
- e.g. the simulated auto-correlation function can be found on the diagonal of the field *oo_.autocorr*
- Need sub-functions **acfcomp.m** and **autocov.m** to compute the sample ACF.
- In the working directory, **acfs_plot_.m** plots the sample ACFs and estimated ACFs from the model

Validation - matching moments

Standard Deviation			
Model	Output	Inflation	Interest rate
Data	0.5398	0.2400	0.6142
Model	0.7469	0.4115	0.4706
Cross-correlation with Output			
Data	1.00	-0.3199	-0.0064
Model	1.00	-0.2465	-0.3030
Autocorrelations (Order=1)			
Data	0.1466	0.5204	0.9371
Model	0.6285	0.6826	0.9128

Table: Selected Second Moments of the Model

Validation - ACF plots

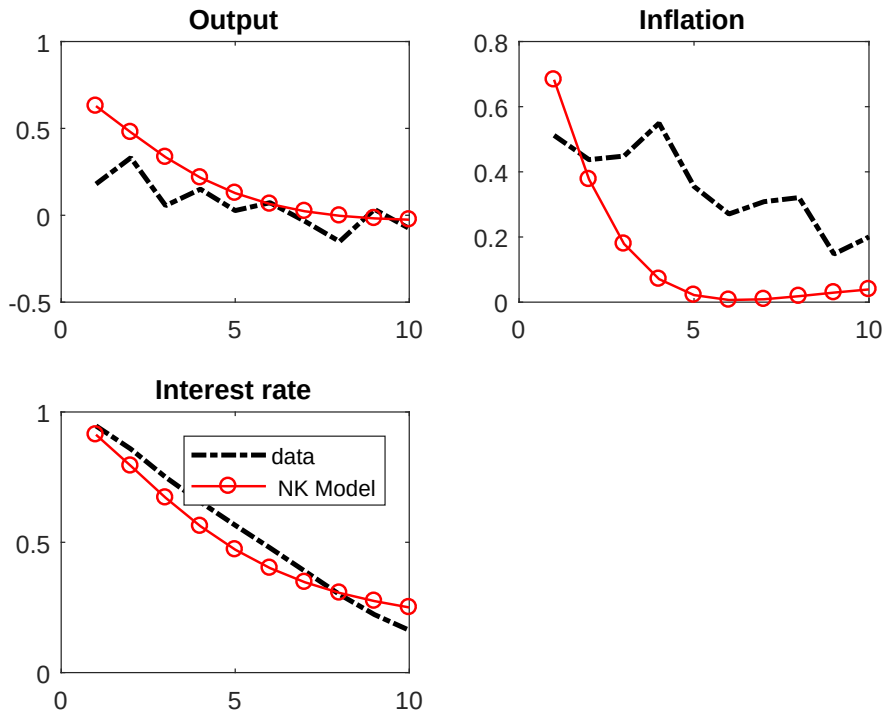


Figure: Auto-correlations of Observables in the Actual Data and in the Estimated NK Model

Summary of Model Estimation and Validation

- The **choice of filter** to make the data stationary - either a two-step procedure filtering the series separately of a one-step procedure as on the Course.
- The **measurement equation**: this links the data with the output of the model
- the choice of **priors**: depends on the range of possible values for the parameter. General guidance: *gamma* or *inverse gamma* distributions for non-negativity constraints, *beta* distributions for fractions or probabilities, *normal* distributions when more informative priors are necessary (*uniform* or 'flat' priors if there is little information about the parameter)
- Computation of **Posterior**: Bayes theorem, mode computation and MCMC.
- **Model validation**: compare second moments with those of the data.
- **Next: Model comparison**

Exercise 1 on Dynare: Estimation

1. Using **us_data.mat**, **NKlinear_Est.mod** has been estimated using mode-compute 6 which is designed to give a negative definite Hessian. The mode file saved as **NKlinear_Est_mode_saved** and this can be re-cycled to use as an initial estimate of the mode in subsequent estimations of closely related models.
2. The estimated model up to now assumed a *conventional* Taylor rule, With the three observed variables chosen in the notes, first perform the identification and then estimate the NK model with an *implementable* Taylor rule. Use only 10,000 mcmc draws for now and re-run the results later with at least 100,000.
3. Use the graph plotter from Days 1 and 2 provided to compare the impulse responses to the four shocks and **acfs_plot.m** to compare the ACFs. What do you notice about your results?

Exercise 2 on Dynare: Estimation

1. Recall the problem arising in the identification from having four shocks and three observables. Address this by adding consumption growth *dc* as an observable.
2. Now perform the identification and re-run the estimation (with an *implementable rule*). Compare IRFs and ACFs with the original set-up. What do you notice about your results?

Appendix

Testing for MCMC Convergence

- Basic uni-variate test motivated by ANOVA considerations. Generate m MCMC chains, each run for $2n$ iterations; first n are discarded to avoid burn-in period. Let ψ represent one of the parameters, with ψ_{jk}, $j = 1, ..., m$, $k = 1, ..., n$, representing the draws. If the ψ_{jk} were normally distributed with variance σ^2, then an unbiased estimator $\hat{\sigma}^2$ of σ^2 is given by

$$(mn-1)\hat{\sigma}^2 = \sum_{j=1}^{m}\sum_{k=1}^{n}(\psi_{jk}-\psi_{..})^2 \equiv \sum_{j=1}^{m}\sum_{k=1}^{n}(\psi_{jk}-\psi_{j.})^2+n\sum_{j=1}^{m}(\psi_{j.}-\psi_{..})^2$$

  where $\psi_{j.}$ represents the mean for the jth chain, and $\psi_{..}$ is the mean over all chains

Testing for MCMC Convergence (cont)

- One measure of convergence is that the $\psi_{j.}$ are all equal to $\psi_{..}$ i.e. that the initial value of the draw in each chain does not affect the mean. Another test is that the variance is equal across all the chains.
- We can test these together by checking whether the *Potential Scale Reduction Factor* $R_2 \equiv V/W$ is approaching 1, where

$$V = \frac{1}{mn-1}\sum_{j=1}^{m}\sum_{k=1}^{n}(\psi_{jk}-\psi_{..})^2 \qquad W = \frac{1}{m(n-1)}\sum_{j=1}^{m}\sum_{k=1}^{n}(\psi_{jk}-\psi_{j.})^2$$

- Brooks and Gelman (1998) recommend that V and W are plotted sequentially for $k = 1, ..., n$; this means that one can check that as n increases, V and W tend individually to a limit, and that this is the same limit as k approaches n.
  - For a unimodal posterior distribution there is a check that both means and variances of all chains' estimates of ψ tend to the same limit.
  - If the posterior distribution is not unimodal, then Dynare does a similar calculation for third moments.

Multivariate Measures for MCMC Convergence

- An unbiased estimate $\hat{\Omega}$ of the covariance matrix of the vector of parameters θ is

$$\begin{aligned}(mn-1)\hat{\Omega} &= \sum_{j=1}^{m}\sum_{k=1}^{n}(\theta_{jk}-\theta_{..})(\theta_{jk}-\theta_{..})^T \\ &\equiv \sum_{j=1}^{m}\sum_{k=1}^{n}(\theta_{jk}-\theta_{j.})(\theta_{jk}-\theta_{j.})^T + n\sum_{j=1}^{m}(\theta_{j.}-\theta_{..})(\theta_{j.}-\theta_{..})^T\end{aligned}$$

  Matrices V and W are then defined analogously to their scalar versions above. One measure closeness is the maximum root statistic - the solution to $max_a(a^T Va)/a^T Wa)$, which is given by the largest eigenvalue of $W^{-\frac{1}{2}}VW^{-\frac{1}{2}}$, which should tend to 1 if the chains are converging to the posterior distribution. The determinants of V and W should also converge.

- A similar approach is taken for third moments

Andrle, M. (2010). A note on identification patterns in DSGE models. Working Paper Series 1235, European Central Bank.

Brooks, S. and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. *Journal of Computational and Graphical Statistics*, **7**(4), 434–455.

Canova, F. and Sala, L. (2009). Back to Square One: Identification Issues in DSGE Models. *Journal of Monetary Economics*, **56**, 431–449.

Christiano, L., Eichenbaum, M., and Evans, C. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. *Journal of Political Economy*, **113**, 1–45.

Fernandez-Villaverde, J., Rubio-Ramirez, J., Sargent, T., and Watson, M. W. (2007). ABC (and Ds) of Understanding VARs. *American Economic Review*, **97**(3), 1021–1026.

Geweke, J. (1999). Using Simulation Methods for Bayesian Econometric Models: Inference, Development and Communication. *Econometric Reviews*, **18**(1), 1–126.

Levine, P., Pearlman, J., Wright, S., and Yang, B. (2019). Information, VARs and DSGE Models. Technical report, School of Economics Discussion Papers 1619, School of Economics, University of Surrey.

Smets, F. and Wouters, R. (2007). Shocks and Frictions in US business cycles: A Bayesian DSGE approach. *American Economic Review*, **97**(3), 586–606.