Macroeconomics II

Lecture Notes II

The main reference for this session is Blake and Mumtaz (2012). All the materials are available in https://github.com/joaoblf0/Macroeconomics-II.

1 Model Representation

After log-linearizing the model we obtain the following representation

$$\Gamma_0(\boldsymbol{\theta})\boldsymbol{Z}_t = \Gamma_1(\boldsymbol{\theta})\boldsymbol{Z}_{t-1} + \Psi(\boldsymbol{\theta})\boldsymbol{u}_t + \Pi(\boldsymbol{\theta})\boldsymbol{\eta}_t \tag{1}$$

As we've shown in previous sessions, we can get a VAR representation using the procedure by Sims (2001).

$$\boldsymbol{Z}_t = G_1(\boldsymbol{\theta})\boldsymbol{Z}_{t-1} + M(\boldsymbol{\theta})\boldsymbol{u}_t \tag{2}$$

where $\boldsymbol{\theta}$ is the vector of structural parameters in the model, $\boldsymbol{Z_t}$ are the endogenous variables, and $\boldsymbol{u_t}$ is the vector of idyosincratic errors.

Standard MLE estimation (suppose distribution for $\boldsymbol{u}_t$ and calculate a likelihood function $\mathcal{L}(Z_t|\boldsymbol{\theta})$ to maximize) has some problems. First, $\boldsymbol{Z}_t$ may not be observable and we may have data only for a subsample $\boldsymbol{Z}_t^* \subset Z_t$. Second, the available time-series T is usually short, while $\boldsymbol{\theta}$ dimension may be very high. The most efficient solution is the Bayesian estimation.

2 Bayesian Estimation

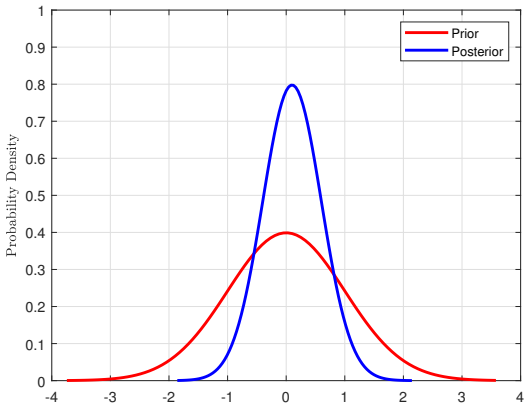
The basic idea is to assume an initial *prior* distribution for the parameters and update it with the data in order to achieve a *posterior* distribution. The final result will not be a single vector of estimators $\hat{\theta}$ for θ but a joint probability distribution for θ. With this distribution we can obtain ex-post IRFs with confidence intervals and calculate distributions for model moments.

2.1 Priors Distributions

First you should define which parameters you want to estimate. You can choose to estimate all the parameters or estimate some of them and calibrate other ones. It is convenient to calibrate the parameters that are well established in the literature (e.g. intertemporal discount factor β, inverse Frisch elasticity of labor supply φ) or the ones have clear data or real-world counterpart (e.g. tax rates, inflation target).

Once the parameters to be estimated are defined, it is necessary to establish a prior distribution for each of them. Priors are initial guesses about the true value of the structural parameters. They can have different levels of informativeness, e.g. priors with low variance are more informative, while priors with more variance are less informative. Figure 1 below shows an example of a prior based on $N(0, 1)$ and an estimated posterior $N(0.1, 0.5)$.

Figure 1: Example of Prior and Posterior Distributions



There is a tradeoff on priors decision:

1. More informative priors may lead to more precise posterior distributions. Defining informative priors for some parameters may also help to obtain better posteriors to other parameters, for which we assumed less informative priors. The main problem, however, arises when a high informative prior is wrong leading into distorted posteriors or biased distributions for other parameters.

2. Less informative parameters are useful when there are parameters that have few and/or sparse results in the literature. Defining larger range estimations may help avoid possible errors that can impact estimations. The problem is that, due to the small data size and the large number of parameters, there can be several (linear or non-linear) different combinations of parameters that lead to similar outcomes in the data. This can result in unpredictable, imprecise, and possible incorrect final estimations.

The choice must rely on the researcher, that must justify the chosen priors for the parameters. The most usual choice is to look in the literature for papers within the same country and use the same priors, or use their posteriors as your priors. Usually, Central Banks' research departments are good sources for a first search.

Important: Distribution densities must be compatible with parameters' range values, e.g. the distribution density for the Calvo re-optimization probability λ must necessarily be in the range $[0, 1]$. The most usual prior choices are

- Beta: for probabilities, discount factor, shocks persistence. The distribution has $[0, 1]$ support.
- Inverse Gamma: for the variance of idiosyncratic shocks. The distribution is asymmetric and concentrated in small values, but allows for high shocks.

- Gamma: for usual positive parameters. The support is positive, but it has an almost symmetric distribution.

Important: Despite the rule-of-thumb above, all priors' choice must be justified.

2.2 Computing Likelihood

Given an arbitrary array of parameters $\boldsymbol{\theta}$ we can write the model in the form (1). Numerically we can apply Sims (2001) method and obtain the VAR representation as in (2). Thus, we can define the the error vector $\bar{\boldsymbol{u}}_t = M(\theta)\boldsymbol{u}_t$ as jointly normal such that $\bar{\boldsymbol{u}}_t \sim N(0, \Sigma(\theta))$.

Using this distribution we can easily find the likelihood function for some sample Z_T: $P(Z_T|\theta)$

$$P(Z_T|\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi|\Sigma(\theta)|}} \exp\left[\frac{(Z_t - \hat{Z}_t)^T \Sigma(\theta)^{-1}(Z_t - \hat{Z}_t)}{2}\right] \tag{3}$$

where $\hat{Z}_t = G_1(\theta)Z_{t-1}$ with initial $Z_1 = \mathbf{1}$.

However, data about all variables in Z_t may not be available. We must then define a vector $Z_t^* \subset Z_t$ of observable variables, and define a matrix H such that

$$Z_t^* = HZ_t \tag{4}$$

In this case we must run the **Kalman Filter**, which, in practical terms, compute the following likelihood

$$P(Z_T^*|\theta) = \prod_{t=1}^{T} \frac{1}{\sqrt{2\pi|H\Sigma(\theta)H^T|}} \exp\left[\frac{(Z_t^* - \hat{Z}_t^*)^T (H\Sigma(\theta)H^T)^{-1}(Z_t^* - \hat{Z}_t^*)}{2}\right] \tag{5}$$

where $\hat{Z}_t^* = HZ_t$ and $Z_t = G_1(\theta)(Z_{t-1} + \Sigma H^T(H\Sigma(\theta)H^T)^{-1}(Z_{t-1} - Z_{t-1}^*))$ with initial $Z_0 = \mathbf{1}$ (see the example code in page 189 of Blake and Mumtaz (2012)).

Once obtained the likelihood $P(Z_T^*|\theta)$, we can find the density of the posterior distribution for θ ($P(\theta|Z_t^*)$) by Bayes rule

$$P(\theta|Z_T^*) = \frac{P(Z_T^*|\theta)P(\theta)}{P(Z_T^*)} = \frac{P(Z_T^*|\theta)P(\theta)}{\int P(Z_T^*|\tilde{\theta})P(\tilde{\theta})d\tilde{\theta}} \tag{6}$$

where $P(\theta)$ is our prior evaluated at θ. The main problem here is that denominator may not be easily calculated, thus we need to use another strategy.

3 Random Walk - Metropolis Hastings

Consider the likelihood for two different parameters vectors $\boldsymbol{\theta}'$ and $\boldsymbol{\theta}''$

$$P(\theta'|Z_T^*) = \frac{P(Z_T^*|\theta')P(\theta')}{P(Z_T^*)} \qquad P(\theta''|Z_T^*) = \frac{P(Z_T^*|\theta'')P(\theta'')}{P(Z_T^*)} \tag{7}$$

Despite $P(\theta'|Z_T^*)$ and $P(\theta''|Z_T^*)$ probably being hard to compute due to the denominator, we can derive the ratio

$$\frac{P(\theta'|Z_T^*)}{P(\theta''|Z_T^*)} = \frac{P(Z_T^*|\theta')P(\theta')}{P(Z_T^*|\theta'')P(\theta'')} \tag{8}$$

Intuitively, if we could calculate this ratio for a very wide range of θ values we will likely be able to find a distribution for $P(\theta|Z_T^*)$. Formally we will do the following steps

3.1 Numerical Maximization

Using numerical methods find the vector θ_0 that maximizes $P(Z_T^*|\theta)P(\theta)$. Once calculated we can also obtain the numerical inverse Hessian evaluated at θ^*: $\mathcal{H}^{-1}(\theta_0)$. It will be our starting point.

3.2 Random Walk

Define a positive parameter k. The vector θ_1 is defined as θ_0 (initial) plus and adjustment rate e. We need to create a "basket" or a "list" to include the arrays for θ being found in the iteration.

Now we proceed with a random-walk in the following way: $\theta_1 = \theta_0 + e$ where $e \sim N(0, k\mathcal{H}^{-1}(\theta_0))$. Once we obtain θ_1, we compute $P(Z_T^*|\theta_1)P(\theta_1)$. Then, we compare this value with $P(Z_T^*|\theta_0)P(\theta_0)$. The loop process follows

- If $\frac{P(Z_T^*|\theta_1)P(\theta_1)}{P(Z_T^*|\theta_0)P(\theta_0)} > 1$ then the posterior likelihood of θ_1 is higher than θ_0. In this case include the array θ_1 in the basket.
- If $\frac{P(Z_T^*|\theta_1)P(\theta')}{P(Z_T^*|\theta_0)P(\theta_0)} = \alpha < 1$ then the posterior likelihood of θ_1 is lower than θ_0. In this case, include θ_1 in the basket with probability α, i.e. draw a number based on a distribution $U(0,1)$, if it is lower than α, then include θ_1 in the basket, if it is higher, do not include.

This process is iterated forward using $\theta_{i+1} = \theta_i + e$ where $e \sim N(0, k\mathcal{H}^{-1}(\theta^*))$. The vector θ_i is the last one included in the basket, thus if θ_i was in fact included we use it to find θ_{i+1}, however, if it was not included, we use the last θ that was included in the basket. To decide if θ_{i+1} will be included in the basket we follow the same rule defined above, always comparing with the previous vector included in the basket.

Repeat this process several times (at least some thousand samples). At the end of the process you should have a basket with some thousand of samples for θ. Note that the number of included vectors will always be lower than the total sample, as some θ may not be included in the basket.

3.3 Control Acceptance

After thousands of iterations, check the aceptance rate (included/ total sample). This rate should be within $[20\%, 40\%]$ during all the iteration process. If the acceptance rate is higher than 40%, most samples were included and the random-walk is moving too slow. From θ_i to θ_{i+1} the likelihood difference is too small, thus the likelihood rate is close to one resulting in an inclusion probability of one. You

should increase the parameter k to improve random-walk variance. Analogously, if the acceptance rate is lower than 20% almost all samples are not being included in the basket, thus the random-walk is moving too "fast". You should reduce the parameter k to improve the random-walk variance.

3.4 Check Convergence

We need to define a stopping time for sampling process, however there is not a general rule. Even models with few variables and equations, good priors and other characteristics, may need more than a million iterations to converge to a stable posterior. We need to define a stopping-rule.

The stopping rule is based on the comparison of different chains. The process described in steps 2 and 3 results in a random-walk chain, thus we can run in parallel between 2 or 5 independent chains, each one with its own basket. In lower steps, the baskets will likely be different as each chain goes to a different direction, however, as the iteraction increases, the baskets and, therefore the posterior distributions, will converge. Once the different baskets converged, we can stop the process.

We use Brooks and Gelman (1998) method to check convergence. This method consist in comparing the moments from the different chains and the aggregate. Suppose we call Θ^j the set θ in the chain j basket and there are J baskets, and also define a pooled basket Θ which joins all J baskets. We define 3 measures:

- Interval: difference between 0.1 and 0.9 percentiles. The within-sample interval measure is a simple average of the difference in the J baskets. The pooled interval measure is this difference between these percentiles in pooled basket. Note that if chains converged, this difference should be about the same in each chain and the same in the pooled chain. On the other hand, if chains did not converge yet the pooled interval will be higher than within sample.
- m2 and m3: same procedures but instead of percentile difference it uses the 2nd and the 3rd central moments difference within each chain and in the pooled chain. The convergence is achieved when these values are close.

It is useful to construct plots compating how these measures evolves over the iterations. Red line: within-sample average. Blue line: pooled sample.

Figure 2: Interval, m2 and m3 measures.

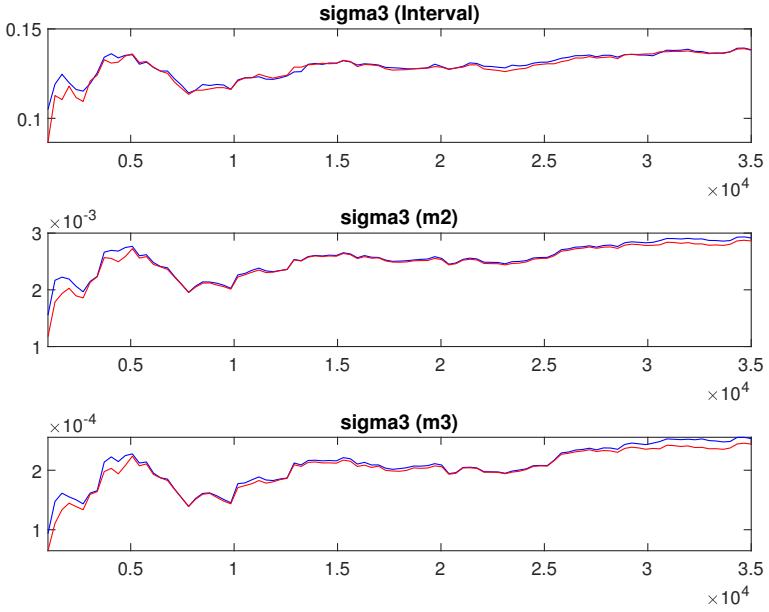
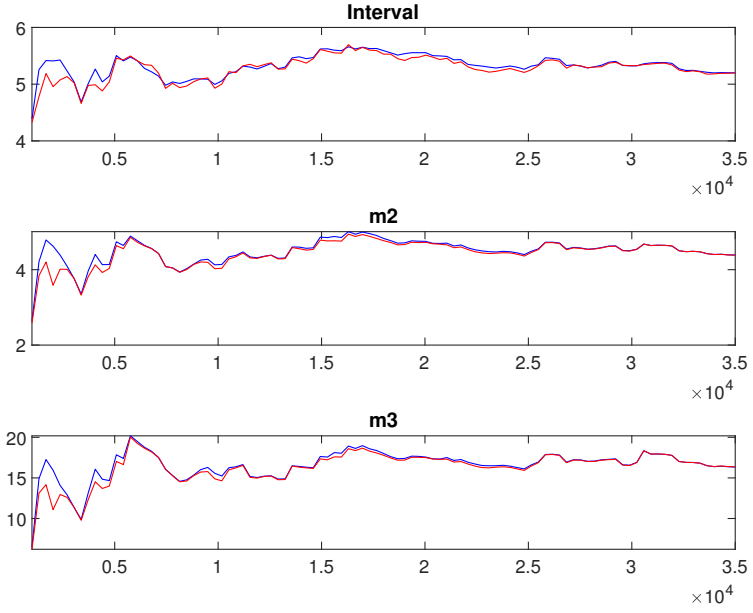


Figure 3: Interval, m2 and m3 measures (multivariate).

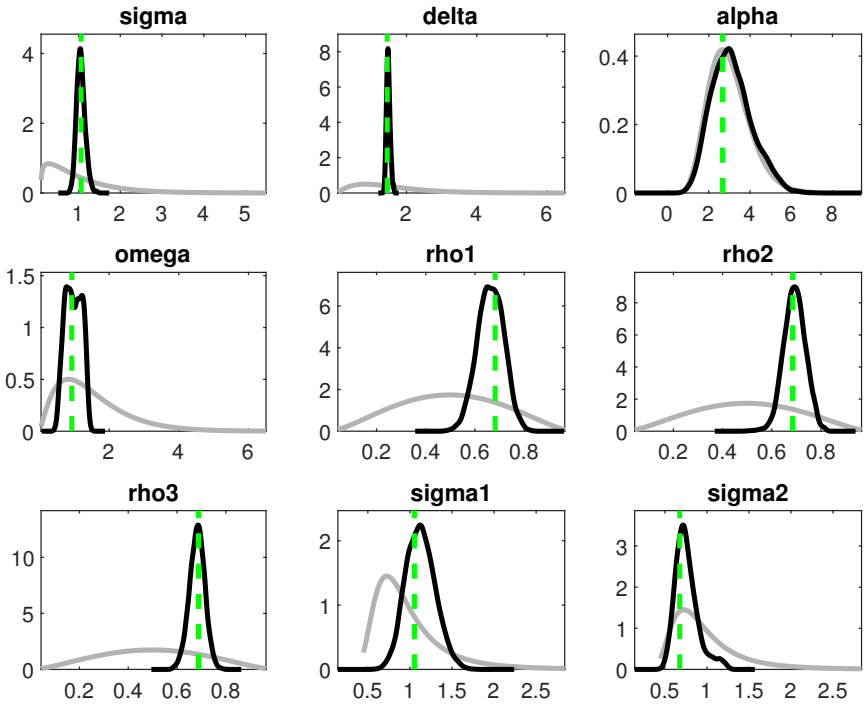


By these charts the convergence is achieved when both lines are close and near stable.

3.5 Bayesian IRFs

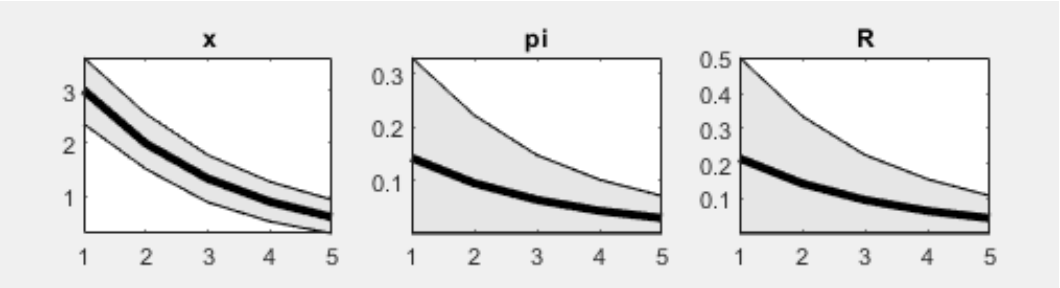
Once concluded the estimation, you will have the parameters' posterior distributions.

Figure 4: Priors and Posteriors.



Using the values in the baskets (or a sub-sample of them as it will have likely million of elements) you can calculate the ex-post Bayesian impulse response functions. You need just to compute the IRFs for each vector θ and then aggregate them in a single fan or range chart.

Figure 5: Priors and Posteriors.



4 Example

Consider the following example:

$$
\begin{aligned}
x_t &= \mathbb{E}_t[x_{t+1}] - \frac{1}{\sigma}\left(i_t - \mathbb{E}_t[\pi_{t+1}]\right) + g_t \\
\pi_t &= \beta\mathbb{E}_t[\pi_{t+1}] + \kappa x_t + u_t \\
i_t &= \delta\pi_t + \nu_t \\
\kappa &= \frac{(1-\bar{\omega})(1-\beta\bar{\omega})}{\alpha\bar{\omega}} \\
g_t &= \rho_1 g_{t-1} + \varepsilon_{1,t}, \qquad \varepsilon_{1,t} \sim N(0, \sigma_1^2) \\
u_t &= \rho_2 u_{t-1} + \varepsilon_{2,t}, \qquad \varepsilon_{2,t} \sim N(0, \sigma_2^2) \\
v_t &= \rho_3 v_{t-1} + \varepsilon_{3,t}, \qquad \varepsilon_{3,t} \sim N(0, \sigma_3^2)
\end{aligned}
$$

We generated a random sample with $T = 200$ and the following parameters: $\sigma = 1$, $\beta = 0.99$, $\delta = 1.5$, $\alpha = 3$, $\bar{\omega} = 1.5$, $\rho_1 = 0.7$, $\rho_2 = 0.7$, $\rho_3 = 0.7$, $\sigma_1 = \sigma_2 = \sigma_3 = 1$. Only the variables x_t, π_t and i_t are observable.

1. Using the true values of the parameters derive the impulse response functions to all shocks in this economy.

2. Suppose you are a researcher interested in this economy. You do not know the true parameters and want to estimate it. First you assume the following prior distributions:

Parameter	Distribution	Mean	Std. Dev
σ	Gamma	1.0	0.8
δ	Gamma	1.5	1.0
α	Gamma	3.0	1.0
ω	Gamma	1.5	1.0
ρ_1	Beta	0.5	0.2
ρ_2	Beta	0.5	0.2
ρ_3	Beta	0.5	0.2
σ_1	Inv. Gamma	1.0	0.5
σ_2	Inv. Gamma	1.0	0.5
σ_3	Inv. Gamma	1.0	0.5

Priors distributions

   Run a Bayesian estimation and find the parameters' posterior distributions. Compare these distributions with the true values.

3. Derive the Bayesian IRFs based in the posteriors found in item b. Compare with true IRFs found in item 1.

References

Blake, A. and Mumtaz, H. (2012). Applied bayesian econometrics for central bankers. *Bank of England*.

Brooks, S. and Gelman, A. (1998). General methods for monitoring convergence of iterative simulations. *Journal of Computational and Graphical Statistics*.

Sims, C. A. (2001). Solving linear rational expectations models. *Computational Economics*.