

Stefano Grassi

DSGE Estimation (Based on Filippo Ferroni Slides)

DSGE models

- What is a DSGE model ?

A system of non linear equations with expectations

$$E_t f(x_{t+1}, x_t, x_{t-1}, \epsilon_t; \theta) = 0 \quad \epsilon_t \sim (0, \Sigma)$$

where x_t is a vector of endogenous variables and ϵ_t is a vector of exogenous innovation θ recollects the 'deep' or structural parameters of the model.

- We want to express the endogenous variables as a function of exogenous ones, i.e. solve the model $x_t = h(x_{t-1}, \epsilon_t; \theta)$
- Analytical solution of h are not implementable. Taylor expansions around a point, the steady state, and solve numerically.

$$x_t = A(\theta)x_{t-1} + B(\theta)\epsilon_t$$

State Space representation

- How do we bridge the model to the data ? State space representation

$$y_t = Cx_t \quad (1)$$

$$x_t = A(\theta)x_{t-1} + B(\theta)\epsilon_t \quad \epsilon_t \sim N(0, \Sigma) \quad (2)$$

- The first equation is called the *measurement* equation and links the observables to the recursive model dynamics
- The second equation is called the *transition* equation and embeds the recursive dynamics of the DSGE model.
- For estimation proposes make sure that $n_\epsilon \geq n_y$. Else Stochastic singularity. [▶ Example](#)

Bayesian Likelihood

- Recall that we are interested in this object, $p(\theta|y)$, the posterior distribution of the parameters and $p(\theta|y) \propto p(\theta)L(y|\theta)$
- We need a device to compute $L(y|\theta)$ of the state space system
- If we have a dataset of time series data $y^T = \{y_1, y_2, \dots, y_T\}$, then using Bayes Theorem it is straightforward to show that

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

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(1)

- The Kalman filter is a recursive algorithm for computing the mathematical expectation $x_{t|t} = E[x_t | y_t, \dots, y_0]$ of a hidden state vector x_t , conditional on observing a history y_t, \dots, y_0 of a vector of (noisy) signals on the hidden state. It is a signal extraction or filtering problem.
- In a linear and gaussian world, the Kalman filter gives the best estimates for the unobserved states, $x_{t|t}$, and the likelihood of the system **given** A, B, C, Σ and a series of observed variables y_1, \dots, y_T .

Calculation of the Likelihood Function(2)

- The log-likelihood is then given by

$$\ln L(y^T|\theta) = -\frac{Tr}{2}\ln(2\pi) - \frac{1}{2}\sum_{t=1}^T(\ln \det(F_t) + e_t'F_t^{-1}e_t)$$

where r is the number of measurements at each period, and e_t, F_t are obtained from the Kalman Filter recursions [▶ More Details on KF](#)

$$e_t = y_t - Cx_{t|t-1}$$

$$F_t = CP_tC'$$

$$x_{t+1|t} = Ax_{t|t-1} + AP_tC'F_t^{-1}e_t$$

$$P_{t+1} = AP_tA' - AP_tC'F_t^{-1}CP_tA' + B\Sigma B'$$

subject to the initial conditions $x_{1|0} = 0$, and P_1 being the solution of the Lyapunov equation $P_1 = AP_1A' + B\Sigma B'$.

First Stage Estimation - Maximizing the Likelihood

- The first thing done by Dynare in the estimation stage is to maximize the Bayesian likelihood, i.e.

$$\max_{\theta} p(\theta)L(y|\theta)$$

- 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 $\int L(y^T|\theta, M_i)d\theta$
- Different models M_i may have some parameters fixed at 0, or estimated under different information sets. The econometrician will prefer the model with highest marginal likelihood
- The Laplace approximation to the log marginal likelihood is given by $\frac{N}{2}\ln(2\pi) + \ln L(y^T|\theta^*, M_i) - \frac{1}{2}\ln(\det(I_T))$ where I_T is the information matrix evaluated at the maximum θ^* .

Problems with Maximizing the Likelihood

- 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 RW MH algorithm described earlier

MH algorithm for DSGE models in a nutshell

Given θ_0 from the maximization step, Ω from the information matrix and the size of the jump c , for $\ell = 1, \dots, L$

- ➊ Draw a candidate draw from $\theta^* \sim N(\theta_{\ell-1}, c\Omega)$.
- ➋ Plug it in the DSGE model, $E_t F(x_{t+1}, x_t, x_{t-1}, \epsilon_t; \theta^*) = 0$.
- ➌ solve the DSGE, $x_t = A(\theta^*)x_{t-1} + B(\theta^*)\epsilon_t$ $y_t = Cx_t$.
- ➍ Compute the likelihood using the Kalman filter, i.e. $L(y|\theta^*)$
- ➎ Contrast the kernels of the candidate draw and previous accepted draws,

$$R = \frac{p(\theta^*)L(\theta^*|y)}{p(\theta_{\ell-1})L(\theta_{\ell-1}|y)}$$

- ➏ Keep the draw with certain probability. Draw $u \sim U(0, 1)$

$$\text{if } R > u, \quad \theta_\ell = \theta^*$$

$$\text{if } R \leq u, \quad \theta_\ell = \theta_{\ell-1}$$

- ➐ go back to 1.

MCMC Metropolis-Hastings algorithm Practical Issues.

- How many draws ?

Test for convergence. Brooks and Gelman (1998). Idea is to launch different chains and verify that the parameter distribution across chains is statistically identical.

[▶ More Details on Brooks and Gelman](#)

- The acceptance rate ?

Seek to obtain 'reasonable' acceptance rate by adjusting the covariance matrix of $c\Omega$

- adjust the 'scale' c
- Ideally acceptance rate is 20-40% \Rightarrow each move goes a reasonable distance in parameter space, but is not rejected too frequently

MCMC Metropolis-Hastings algorithm Practical Issues.

Priors ?

- 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 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 parameters.

Stochastic Singularity

Let A be $n_x \times n_x$, B be $n_x \times n_\epsilon$ and C be $n_y \times n_x$. the state space becomes

$$x_t - Ax_{t-1} = B\epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

$$(I - AL)x_t = B\epsilon_t$$

$$y_t = C(I - AL)^{-1}B\epsilon_t = W\epsilon_t$$

where $W = C(I - AL)^{-1}B$ is a $n_y \times n_\epsilon$ matrix. If $n_y > n_\epsilon$, then W is not invertible. If $n_\epsilon > n_y$, you can take a combination of shocks, i.e. $y_t = W_1\eta_t$ where $\eta_t = W_2\epsilon_t$ and $W = W_1W_2$ with W_1 $n_y \times n_y$ and W_2 $n_y \times n_\epsilon$.

Assume $n_y = n_x = 2$, $n_\epsilon = 1$, $C = I$. We obtain

$$y_t - Ay_{t-1} = B\epsilon_t \quad \epsilon_t \sim N(0, \sigma^2)$$

B is 2×1 , clearly not invertible, and hence we cannot compute the likelihood.

► return

Testing for MCMC Convergence

- Dynare uses some indicative statistics, summarized by diagrams, as recommended by Brooks and Gelman (1998). These are made up of
 - 3 multivariate figures, representing convergence indicators for all parameters considered together
 - 3 figures for each parameter, representing univariate convergence indicators
- Basic univariate 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} , $j = 1, \dots, m$, $k = 1, \dots, n$, 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 j th chain, and $\psi_{..}$ is the mean over all chains

Testing for MCMC Convergence (cont)

- One measure of convergence is that the $\psi_{j\cdot}$ are all equal to $\psi_{\cdot\cdot}$ i.e. that the initial value of the draw in each chain has not affected the mean. Another test is whether 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

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

- Brook and Gelman recommend that V and W are plotted sequentially for $k = 1, \dots, n$; 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 .
 - If the posterior distribution is unimodal, this is essentially a check that both means and variances of all chains' estimates of ψ are tending to the same limit.
 - If the posterior distribution is not unimodal, then it makes sense to extend this to other moments, and Dynare does a similar calculation for third moments as well.

Testing for MCMC Convergence - Interval Measures

- Based on the intuitive notion that R_2 also represents a squared ratio of the proportion of draws within a certain confidence interval. To perform this explicitly, Brook and Gelman suggest a measure $R_{interval}$ that uses, as before, the last n of the $2n$ draws of each chain, and then
- From each chain find the empirical $100(1-\alpha)\%$ interval i.e. the number of draws within the empirical $100\frac{\alpha}{2}\%$ and $100(1 - \frac{\alpha}{2})\%$ points; Dynare sets $\alpha = 0.2$.
- Do the same for all the mn draws from all the m chains
- Evaluate $R_{interval} \equiv V_{interval} / W_{interval}$ where $V_{interval}$ =length of total-sequence interval, $W_{interval}$ =mean length of within-sequence intervals. As before, it is insightful to plot both $V_{interval}$ and $W_{interval}$.

Testing for MCMC Convergence - Multivariate Measures

- 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 $W^{-\frac{1}{2}} V W^{-\frac{1}{2}}$, 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
- Interval measure: count the number of draws for which each of the elements θ_i of the vector θ lie within their individual empirical $100(1-\alpha)\%$ intervals. Find the average, and compare with the whole sample taken together.