PhD Macroeconomics - DSGE methods

Willi Mutschler, M.Sc.

Institute of Econometrics and Economic Statistics
University of Münster
willi.mutschler@wiwi.uni-muenster.de

Summer 2015

Previously...

- Theory and intuition behind baseline DSGE models
- Derivation of the structural form, log-linearization and solution via method of undetermined coefficients

Insight

DSGE model consists of

- set of Euler equations, i.e. first-order optimality conditions,
- transition equations for state and control variables
- transition equations for structural shocks and innovations,
- observable variables and measurement errors

which can be cast into a nonlinear system of expectational difference equations.

Introduction to Dynare

Dynare

- computes the solution of deterministic models (arbitrary accuracy)
- computes first, second and third order approximation to solution of stochastic models
- estimates (maximum likelihood or Bayesian approach) parameters of DSGE models
- computes optimal policy
- performs global sensitivity analysis of a model
- solves problems under partial information
- estimates BVAR and Markov-Switching Bayesian VAR models
- estimates DSGE-VAR models
- estimates Markov-Switching DSGE models
- ...

General DSGE framework

General DSGE model

$$
\begin{aligned}
0 &= E_t f\left(x_{t+1}, y_{t+1}, x_t, y_t | \theta\right), \\
x_{t+1} &= h(x_t, \varepsilon_{t+1} | \theta), \\
y_{t+1} &= g(x_t, \varepsilon_{t+1} | \theta), \\
d_t &= D y_t + \mu_t.
\end{aligned}
$$

with $E(\varepsilon_t) = 0$, $E(\varepsilon_t \varepsilon_t') = \Sigma_\varepsilon$ and $E(\mu_t) = 0$, $E(\mu_t \mu_t') = \Sigma_\mu$.

Some remarks

- DSGE-models can be interpreted as *state-space-models*.
- Driving force of the model are exogenous shocks and innovations ε_t.
- Stability and determinacy: Check Blanchard and Khan conditions
- Flexible framework: you can add auxiliary variables and equations.

Deterministic vs. Stochastic models

- Major distinction: Are future shocks known?
  - Deterministic models: occurrence of all future shocks is known exactly at the time of computing the model's solution
  - Stochastic models: only the distribution of future shocks is known
- If you have a linear model or first order linear approximation of the stochastic model the two cases become practically the same, due to certainty equivalence.

Deterministic vs. Stochastic models

Deterministic models useful for

- Models with full information, perfect foresight and no uncertainty around shocks.
- Focus on impact of a change in regime, e.g. introduction of a new tax
- Shocks can hit today or at any time in the future, in which case they would be expected with perfect foresight. They can also last one or several periods.
- Most often, though, models introduce a positive shock today and zero shocks thereafter (with certainty).
- Solution does not require linearization, but numerical simulation techniques to find the exact paths of endogenous variables that meet the model
- In practice useful to get a first glimpse of model before using the stochastic model for further analysis and estimation

Deterministic vs. Stochastic models

Stochastic models useful for

- More popular in literature: RBC models or new keynesian monetary models.
- Shocks hit today (with a surprise), but thereafter their expected value is zero. Expected future shocks, or permanent changes in the exogenous variables cannot be handled due to the use of Taylor approximations around a steady state.
- Note that when these models are linearized to the first order, agents behave as if future shocks where equal to zero (since their expectation is null), which is the certainty equivalence property.
  - This does NOT mean that model is deterministic
- Useful for estimation!

DSGE solution

One distinguishes between linear and non-linear methods:

- Linear methods: Anderson/Moore (1983), Binder and Pesaran, Blanchard/Khan (1980), (1997), Christiano (2002), King and Watson (1998), Klein (2000), Sims (2001) and Uhlig (1999) (See Anderson (2008) for a comparison).
- Nonlinear methods: Projection methods, iteration procedures or perturbation approach (see DeJong/Dave (2011) for a comparison)

Dynare uses the perturbation approach:

- Perturbation approach finds a local approximation of the policy functions in a neighborhood of a particular point
- Mostly steady-state, since we can solve it analytically or numerically.

DSGE solution

First-order approximation

Pros:

- Simple linear state-space representation of the model, which in many cases is sufficiently exact.
- One can use the Kalman-filter to empirically evaluate the system.

Cons:

- One looses important information during the linearization.
- Higher moments play an important role for analyzing markets, risk, welfare, etc.
- An approximation to, say, the second order can yield different results, because the variance of future shocks matters (risk premium).

Certainty-equivalence

Important theoretical result for stochastic models:

- Even though agents take the effect of future shocks into account when optimizing, in a linearization to the first-order they don't matter for the decision rule.

Certainty-equivalence-property

- In a first-order approximation the constant term needs not to be corrected for uncertainty (i.e. variance of shocks)
- Expectation of x_t and y_t is equal to its non-stochastic steady-state

- This is problematic when uncertainty does matter: risk premia, welfare evaluation, ...

Toy example: Clarida-Gali-Gertler

Household

Household preferences are given by

$$E_0 \sum_{t=0}^{\infty} \left(\log C_t - \exp(\tau_t) \frac{N_t^{1+\varphi}}{1+\varphi} \right), \qquad \tau_t = \lambda \tau_{t-1} + \varepsilon_t^{\tau}, \qquad \varepsilon_t^{\tau} \sim iid.$$

- C_t denotes consumption,
- τ_t denotes a time preference shock,
- N_t denotes employment,
- φ denotes a labor supply parameter.

The budget constraint of the household is

$$P_t C_t + B_{t+1} \leq W_t N_t + R_{t-1} B_t + T_t,$$

- T_t denotes (lump-sum) taxes and profits,
- P_t denotes price level,
- W_t denotes nominal wage rate
- B_{t+1} denotes bonds purchased at time t, which deliver a non-state-contingent rate of return, R_t, in period $t+1$.

Toy example: Clarida-Gali-Gertler

Competitive firms

Competitive firms produce a homogeneous output good, Y_t, using the following technology:

$$Y_t = \left[\int_0^1 Y_{i,t}^{\frac{\varepsilon-1}{\varepsilon}} \right]^{\frac{\varepsilon}{\varepsilon-1}} di, \ \varepsilon > 1,$$

- $Y_{i,t}$ denotes the i^{th} intermediate good, $i \in (0,1)$.

Competitive firms take the price of the final output good, P_t, and the prices of the intermediate goods, $P_{i,t}$, as given and choose Y_t and Y_{it} to maximize profits;

$$Y_{i,t} = Y_t \left(\frac{P_t}{P_{i,t}} \right)^{\varepsilon}.$$

- This is the demand curve for the producer of Y_{it} in the intermediate sector.

Toy example: Clarida-Gali-Gertler

Intermediate firms

The i^{th} intermediate good firm is a monopolist for Y_{it} and uses labor, $N_{i,t}$, to produce output using the following production function:

$$Y_{i,t} = \exp(a_t) N_{i,t}, \ \Delta a_t = \rho \Delta a_{t-1} + \varepsilon_t^a,$$

- Δ is the first difference operator,
- ε_t^a is an iid shock $\rightarrow$ a_t has a unit root representation

The i^{th} firm sets prices subject to Calvo frictions. In particular,

$$P_{i,t} = \begin{cases} \tilde{P}_t & \text{with probability } 1-\theta \\ P_{i,t} & \text{with probability } \theta \end{cases},$$

- $\tilde{P}_t$ denotes the price chosen by the $1-\theta$ firms that can reoptimize their price at time t.

Toy example: Clarida-Gali-Gertler

Intermediate firms

- The i^{th} producer is competitive in labor markets, pays $W_t (1-\nu)$ for one unit of labor.
- ν represents a subsidy which eliminates the monopoly distortion on labor in the steady state:

$$1-\nu = (\varepsilon - 1)/\varepsilon.$$

Toy example: Clarida-Gali-Gertler

Ramsey equilibrium

The Ramsey equilibrium for the model is the equilibrium associated with the optimal monetary policy, it is characterized by

- zero inflation, $\pi_t = 0$, at each date and for each realization of a_t & τ_t
- consumption and employment correspond to their first best levels
- C_t and N_t satisfy the resource constraint

$$C_t = \exp(a_t) N_t,$$

- marginal rate of substitution between consumption and labor equals the marginal product of labor

$$\frac{\text{marginal utility of leisure}}{\text{marginal utility of consumption}} = C_t \exp(\tau_t) N_t^{\varphi} = \exp(a_t).$$

Toy example: Clarida-Gali-Gertler

Ramsey equilibrium

Solving yields:

$$\pi_t^* = 0,$$
$$n_t^* := \log(N_t^*) = -\frac{\tau_t}{1+\varphi},$$
$$c_t^* := \log(C_t^*) = a_t - \frac{\tau_t}{1+\varphi} = \log(Y_t^*) =: y_t^*,$$
$$r_t^* := log(R_t^*\beta) = E_t\Delta a_{t+1} - E_t\frac{\tau_{t+1}-\tau_t}{1+\varphi},$$

- $*$ indicates a variable corresponding to the Ramsey equilibrium, i.e. natural rates,
- lower case letters are log of the corresponding variable,
- r_t^* is log deviation from non-stochastic steady-state.

Toy example: Clarida-Gali-Gertler

The model equations

Linearizing about the steady-state the model equations are given by

$$
\begin{aligned}
\pi_t &= \beta E_t \pi_{t+1} + \kappa x_t \text{ (Calvo pricing equation)} \\
x_t &= -\left[r_t - E_t \pi_{t+1} - r_t^*\right] + E_t x_{t+1} \text{ (intertemporal equation)} \\
r_t &= \alpha r_{t-1} + (1-\alpha)\left[\phi_\pi \pi_t + \phi_x x_t\right] \text{ (policy rule)} \\
r_t^* &= \rho \Delta a_t + \frac{1}{1+\varphi}(1-\lambda)\tau_t \text{ (natural rate)} \\
\Delta y_t &= x_t - x_{t-1} + \Delta a_t - \frac{\tau_t - \tau_{t-1}}{1+\varphi} \text{ (output growth)} \\
\Delta a_t &= \rho \Delta a_{t-1} + \varepsilon_t^a \text{ (technological shock)} \\
\tau_t &= \lambda \tau_{t-1} + \varepsilon_t^\tau \text{ (preference shock)}
\end{aligned}
$$

and

$$
\begin{aligned}
y_t^* &= a_t - \frac{1}{1+\varphi}\tau_t \text{ (natural output)} \\
x_t &= y_t - y_t^* \text{ (output gap)} \\
\kappa &= \frac{(1-\theta)(1-\beta\theta)(1+\varphi)}{\theta}
\end{aligned}
$$

Toy example: Clarida-Gali-Gertler

Practicing Dynare

1. Install Matlab and Dynare, open cgg.mod, try to understand the code, run it. Interpret the Dynare output.
2. Compute the impulse response function of the model to a technology shock for the next 7 periods in both the deterministic as well as stochastic model. Explain the difference or equivalence.
3. Given the IRF, indicate whether the economy over- or under-responds due to the shocks, relative to their natural response. What is the economic intuition in each case?
4. Do the calculations with $\phi_\pi = 0.99$. Explain the error message and give economic intuition behind this.

Toy example: Clarida-Gali-Gertler

Practicing Dynare

Return to $\phi_\pi = 1.5$.

5. Explain why it is that when the monetary policy rule is replaced by the natural equilibrium, i.e. $r_t = r_t^*$, the solution is indeterminate.
6. Now replace the monetary policy rule by

$$r_t = r_t^* + \alpha(r_{t-1} - r_{t-1}^*) + (1-\alpha)\left[\phi_\pi \pi_t + \phi_x x_t\right]$$

   Explain why this Taylor rule uniquely supports the natural equilibrium.

Return to the original Taylor rule. Calibrate the model to a more empirically relevant parametrization: $\phi_x = 0.15, \alpha = 0.8, \rho = 0.9$

7. Simulate the model for 1000 periods. Save the middle 100 observations of dy_t and π_t into an Excel-file as well as into a mat-file. Plot the path of output growth.

Exercise: Write your own MOD file

Consider the following simplified RBC-model (social planer problem);

$$\max_{\{c_{t+j},\ell_{t+j},k_{t+j}\}_{j=0}^{\infty}} W_t = \sum_{j=0}^{\infty} \beta^j u(c_{t+j}, \ell_{t+j})$$

$$\begin{aligned} s.t. \quad & y_t = c_t + i_t, && A_t = \bar{A} e^{a_t}, \\ & y_t = A_t f(k_{t-1}, \ell_t), && a_t = \rho a_{t-1} + \varepsilon_t, \\ & k_t = i_t + (1-\delta) k_{t-1}, && \varepsilon_t \sim N(0, \sigma_\varepsilon^2), \end{aligned}$$

where preferences and technology follow:

$$u(c_t, \ell_t) = \frac{\left[c_t^\theta (1-\ell_t)^{1-\theta}\right]^{1-\tau}}{1-\tau}, \qquad f(k_{t-1}, \ell_t) = \left[\alpha k_{t-1}^\psi + (1-\alpha)\ell_t^\psi\right]^{1/\psi}.$$

Optimality is given by:

$$u_c(c_t, \ell_t) = \beta E_t \left\{ u_c(c_{t+1}, \ell_{t+1}) \left[A_{t+1} f_k(k_t, \ell_{t+1}) + 1 - \delta \right] \right\},$$

$$\frac{u_\ell(c_t, \ell_t)}{u_c(c_t, \ell_t)} + A_t f_\ell(k_{t-1}, \ell_t) = 0,$$

$$c_t + k_t = A_t f(k_{t-1}, \ell_t) + (1-\delta) k_{t-1}.$$

Exercise: Write your own MOD file

Steady-state

Note that we could either compute the steady-state analytically and then use steady_state_model in our mod-file or good guesses for an initval block. If you want to use intival, you can use these values

```
k     = 20;
y     = 1;
l     = 0.5;
c     = 1.5;
A     = 1;
a     = 0;
i     = 0.5;
uc    = 0.5;
ul    = -1;
fk    = 0.1;
fl    = 3;
f     = 2;
```

Exercise: Write your own MOD file

Calibration

Weight of consumption in utility	θ	0.357
Risk aversion	τ	2.0
Share of capital in production	α	0.45
Elasticity of substitution capital/labor	ψ	-0.1
Discount factor	β	0.99
Depreciation rate	δ	0.02
Autocorrelation of productivity	ρ	0.8
Steady state level of productivity	A^*	1

Exercise: Write your own MOD file

(a) Write a mod-file for this model. If your residuals are not 0, you can try different optimizers to find the steady-sate, e.g. steady(solve_algo=3).

(b) Assume that the economy starts from $k_0 = 0.5\bar{k}$. Use a deterministic simulation to show how consumption and capital return to equilibrium. Hint: Use histval block.

(c) Assume that the economy starts from steady state. Use a deterministic simulation to show the effects on consumption and capital of an unexpected 1% negative shock at the beginning of period 1. Hint: Use shocks block.

(d) Assume that the economy starts from steady state. Use a deterministic simulation to show the effects consumption and capital of favourable pre-announced shocks in the future, i.e. a sequence of positive shocks to A_t: 4% in period 5 and an additional 1% during the following periods. Hint: Use shocks block.

(e) Assume that the economy starts from steady state. In period 1, TFP increases by 5% permanently (and this was unexpected). Use a deterministic simulation to show the transition path of consumption and capital in response to this permanent shock. Hint: Use endval block.

(f) Assume that the economy starts from steady state. In period 6, TFP increases by 5% permanently (and this was expected). Use a deterministic simulation to show the transition path of consumption and capital in response to this pre-announced permanent shock. Hint: Use endval and shock block.

(g) Simulate a sample of 10000 observations for c_t, ℓ_t and y_t using `stoch_simul` and save it in a mat-file.

Overview Estimation-Methods

- Econometrically, a DSGE-Model is a state-space model of which one has to determine the parameters.
- Three concepts:
  1. **Calibration**: The parameters are set in such a way, that they closely correspond to some theoretical moment or stylized fact of data.
  2. **Methods of limited information [not covered]** or weak econometric interpretation: Minimize the distance between theoretical and empirical moments, i.e. *General-Method-of-Moments* or *Indirect Inference*.
  3. **Methods of full information** or strong econometric interpretation: The goal is an exact characterization of observed data, i.e. *Maximum-Likelihood* or *Bayesian methods*.

Calibration

- Goal: To answer a specific quantitative research question using a structural model.
- Construct and parameterize the model such, that it corresponds to certain properties of the true economy.
- Use steady-state-characteristics for choosing the parameters in accordance with observed data.
- Often: stable long-run averages (wages, working-hours, interest rates, inflation, consumption-shares, government-spending-ratios, etc.).
- You can use micro-studies as well, however, one has to be careful about the aggregation!

Calibration

Hints for calibrating a model

- Use long-term averages of interest rates, inflation, average growth of productivity, etc. for *steady-state* values.
- BUT: Weil (1989) shows, that in models with representative agents there is an overestimation of *steady-state* interest rates (*risk-free rate puzzle*). It is possible that you get absurd constellation of parameters, like a discount-factor of $\beta > 1$.
- Usual mark-up on prices is around 1.15 (Corsetti et al (2012)).
- Intertemporal elasticity of substitution $1/\sigma$ somewhere between $\sigma = 1$ and $\sigma = 3$ (King, Plosser and Rebelo (1988), Rotemberg and Woodford (1992), Lucas (2003)).

Calibration

Hints for calibrating a model

- Rigidity of prices: For an average price adjustment of 12-15 months see Keen and Wang (2007).
- Coefficients of monetary policy: Often Taylor-Rule, you can use the relative coefficients to put more emphasize/weight on the stability of prices or on smoothing the business cycle.
- Parameters of stochastic processes: Often persistent, small standard-deviations, otherwise you get high oscillations. You can also estimate the production function via OLS (Solow-residual).
- How to choose shocks: Look at similar studies: Christiano, Eichenbaum and Evans (2005), Smets and Wouters (2003), etc..
- Ultimately: Try & Error!

Calibration

Pros

- Calibration is commonly used in the literature. It gives a first impression, a flavor of the strengths and weaknesses of a model.
- A good calibration can provide a valuable and precise image of data.
- Using different calibrations, one can asses interesting implications of different policies:
  - How does the economy react, if the central bank focuses more on smoothing the business cycle than on price stability?
  - What happens to consumption, if the households have a strong intertemporal elasticity of substitution? What if it is low?

Calibration

Cons

- This Ad-hoc-approach is at the center of criticism of DSGE-models.
- There is no statistical foundation, it is based upon subjective views, assessments and opinions.
- Many parameter, such as those of the exogenous processes, leave room for different values and interpretations (intertemporal elasticity of substitution, monetary and fiscal parameters, coefficients of rigidity, standard deviations, etc.).

Prescott (1986, S. 10) regarding RBC-models:

The models constructed within this theoretical framework are necessarily **highly abstract**. Consequently, they are necessarily false, and statistical hypothesis testing will reject them. This does not imply, however, that nothing can be learned from such a **quantitative theoretical exercise**.

Full information estimation

Idea

- Full information estimation requires a complete characterization of the data-generating-process (not only specific moments).
- Consider the linear first-order *state-space* representation of the model:

$$z_t = A(\theta)z_{t-1} + B(\theta)\varepsilon_t, \quad \text{with } E[\varepsilon_t] = 0, \quad E[\varepsilon_t\varepsilon_t'] = \Sigma_\varepsilon \tag{1}$$

$$d_t = Dz_t + \mu_t, \quad \text{with } E[\mu_t] = 0, \quad E[\mu_t\mu_t'] = \Sigma_\mu. \tag{2}$$

- $z_t = (\widehat{x}_t', \widehat{y}_t')'$ contains all model variables as deviations from steady-state, and A and B are functions of g_x and h_x.
- Matrix D combines the model variables z_t with observable data variables d_t
- Equation (1): *state-* or *transition-equation*
  - Corresponds to the solution of the model.
  - ε_t are the stochastic innovations.
- Equation (2): *observation-equation*
  - Corresponds to the measurement equations,
  - subject to possible measurement errors μ_t in the data.

Full information estimation

Idea

- Given distributional assumptions about ε_t and μ_t, one can derive the log-likelihood-function, $\log L(d|\theta)$, analytically or numerically.
- In the log-linear case and considering normally distributed variables, the Kalman-filter is used to calculate the likelihood analytically.
- In the nonlinear case the policy functions are functions of the vector of parameters θ. The particle-filter or the *efficient importance sampling* is then used to derive the likelihood numerically.
- There are two approaches for analyzing and evaluating the log-likelihood:
  1. **the classic (frequentist) Maximum-Likelihood-method**,
  2. **the bayesian method**.

Kalman-filter

Notation

We simplify and consider only the linear case and ignore possible measurement errors in the data:

- $d_t = Dz_t$
- $z_{t+1} = Az_t + B\varepsilon_{t+1}$
- $\varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \Sigma_\varepsilon),\ \Sigma_\varepsilon = E(\varepsilon_i \varepsilon_i'),\ E(\varepsilon_i \varepsilon_j') = 0$

Notation for the linear projection

$$
\begin{aligned}
\widehat{z}_{t|t-j} &= E(z_t | d_{t-j}, d_{t-j-1}, \ldots d_1) \\
\Sigma_{t|t-j} &= E(z_t - \widehat{z}_{t|t-j})(z_t - \widehat{z}_{t|t-j})' \\
\widehat{d}_{t|t-j} &= E(d_t | d_{t-j}, d_{t-j-1}, \ldots, d_1) \\
u_t &= d_t - \widehat{d}_{t|t-1} = D(z_t - \widehat{z}_{t|t-1}) \\
E(u_t u_t') &= D\Sigma_{t|t-1} D'
\end{aligned}
$$

for $t = 1, 2, \ldots, T$ and $j = 0, 1, \ldots T$.

Kalman-filter

Initialization

- Since z_t is covariance-stationary, the variance is given by:

$$\begin{aligned}
\underbrace{E(z_t z_t')}_{\equiv \Sigma_z} &= E\left[(Az_{t-1} + B\varepsilon_t)(Az_{t-1} + B\varepsilon_t)'\right] \\
&= A\underbrace{E(z_{t-1}z_{t-1}')}_{\equiv \Sigma_z}A' + B\underbrace{E(\varepsilon_t\varepsilon_t')}_{=\Sigma_\varepsilon}B' \\
\Leftrightarrow \Sigma_z &= A\Sigma_z A' + B\Sigma_\varepsilon B' \\
\Leftrightarrow vec(\Sigma_z) &= (I - A \otimes A)^{-1} vec(B\Sigma_\varepsilon B')
\end{aligned}$$

Vectorization

The *vec*-operation stacks the rows of a $m \times n$ Matrix M into a $mn \times 1$ vector $vec(M)$. Then for arbitrary Matrices $\underset{m\times n}{A}$, $\underset{n\times p}{B}$ and $\underset{p\times k}{C}$:

$$vec(ABC) = (C' \otimes A)vec(B), \quad \text{with } \otimes : \text{Kronecker-product.}$$

Kalman-filter

Initialization

The unconditional expectation of z_1 is used for the initialization of the Kalman-filter, since there is no additional information yet:

$$\widehat{z}_1 = \underbrace{E(z_1)}_{=E(z)} = A\underbrace{E(z_0)}_{=E(z)} + B\underbrace{E(\varepsilon_1)}_{=0} \Leftrightarrow \widehat{z}_1 = 0,$$

$$vec(\Sigma_{1|0}) = E(z_1 - 0)(z_1 - 0)' = vec(\Sigma_z) = (I - A \otimes A)^{-1} vec(B\Sigma_\varepsilon B').$$

Kalman-filter

Recursion

The recursion is then given by:

$$\widehat{z}_{t+1|t} = A\widehat{z}_{t|t}$$

Formula for updating a linear projection (Hamilton (1994, S.99 und S.379))

$$\begin{aligned}
\widehat{z}_{t|t} &= \widehat{z}_{t|t-1} + \left[E(z_t - \widehat{z}_{t|t-1})(d_t - \widehat{d}_{t|t-1})'\right]\left[E(d_t - \widehat{d}_{t|t-1})(d_t - \widehat{d}_{t|t-1})'\right]^{-1} u_t \\
\Leftrightarrow \widehat{z}_{t|t} &= \widehat{z}_{t|t-1} + \Sigma_{t|t-1}D'\left(D\Sigma_{t|t-1}D'\right)^{-1} u_t \\
\Rightarrow \widehat{z}_{t+1|t} = A\widehat{z}_{t|t} &= A\widehat{z}_{t|t-1} + A\Sigma_{t|t-1}D'\left(D\Sigma_{t|t-1}D'\right)^{-1} u_t,
\end{aligned}$$

$$\text{with } u_t = d_t - \widehat{d}_{t|t-1} = (d_t - D\widehat{z}_{t|t-1}).$$

Kalman-filter

Recursion

- $z_{t+1} - \widehat{z}_{t+1|t} = A\left(z_t - \widehat{z}_{t|t-1}\right) + B\varepsilon_{t+1} - A\Sigma_{t|t-1}D'\left(D\Sigma_{t|t-1}D'\right)^{-1} u_t$
- The *MSE:* $\Sigma_{t+1|t} = E\left(z_{t+1} - \widehat{z}_{t+1|t}\right)\left(z_{t+1} - \widehat{z}_{t+1|t}\right)'$ is given by:

$$\Sigma_{t+1|t} =$$
$$A\Sigma_{t|t-1}A' + B\Sigma_\varepsilon B' - A\Sigma_{t|t-1}D'\underbrace{\left(D\Sigma_{t|t-1}D'\right)^{-1}\underbrace{E(u_t u_t')}_{=D\Sigma_{t|t-1}D'}\left(D\Sigma_{t|t-1}D'\right)^{-1}}_{=I} D\Sigma_{t|t-1}A'$$

Mean-Sqared-Error (MSE)

$$\Sigma_{t+1} \equiv \Sigma_{t+1|t} = A\Sigma_{t|t-1}A' + B\Sigma_\varepsilon B' - A\Sigma_{t|t-1}D'\left(D\Sigma_{t|t-1}D'\right)^{-1} D\Sigma_{t|t-1}A'$$

Kalman-filter

Summary

The Kalman-filter can be summarized as follows:

1. Initialization with
   - $\widehat{z}_1 = 0$,
   - $vec(\Sigma_{1|0}) = (I - A \otimes A)^{-1} vec(B\Sigma_\varepsilon B')$.
2. Period-t likelihood function
   - $u_t = (d_t - D\widehat{z}_{t|t-1})$
   - $d_{t|t-1} = D\widehat{z}_{t|t-1}$
   - $\Omega_{t|t-1} := E(u_t u_t') = D\Sigma_{t|t-1}D$
3. Period-t filtering density
   - $\widehat{z}_{t|t} = \widehat{z}_{t|t-1} + \Sigma_{t|t-1}D' \left(D\Sigma_{t|t-1}D'\right)^{-1} u_t$
   - $\Sigma_{t|t} = \Sigma_{t|t-1} - \Sigma_{t|t-1}D' \left(D\Sigma_{t|t-1}D'\right)^{-1} D\Sigma_{t|t-1}$
4. Period-t predictive density
   - $\widehat{z}_{t+1|t} = A\widehat{z}_{t|t-1} + A\Sigma_{t|t-1}D' \left(D\Sigma_{t|t-1}D'\right)^{-1} u_t$
   - $\Sigma_{t+1} = A\Sigma_t A' + B\Sigma_\varepsilon B' - K_t D\Sigma_t A'$.

Log-Likelihood

Given the gaussian assumption about the forecast error u_t one can derive the distribution of the data $\underset{n\times 1}{d_t}$ conditional on $(z_t, d_{t-1}, d_{t-2}, \dots)$ and set up the log-likelihood function:

Log-likelihood

$$\begin{aligned}\log \mathcal{L}(d|\theta) &= \sum_{t=1}^{T} \log \mathcal{L}(d_t|\theta) \\ &= -\frac{nT}{2}\log(2\pi) - \frac{1}{2}\sum_{t=1}^{T}\log|\Omega_t| - \frac{1}{2}\sum_{t=1}^{T} u_t{}'\Omega_t{}^{-1}u_t.\end{aligned}$$

Maximum-Likelihood

Idea

- **Approach:** The parameters θ are fixed and the data is a random realization of this specific parametrization.
- The *Maximum-Likelihood*-estimator $\widehat{\mu}_{ML}$ is then defined as

$$\widehat{\theta}_{ML} = \underset{\theta}{\operatorname{argmax}} \left\{ \sum_{t=1}^{T} \log \mathcal{L}(d_t|\theta) \right\}.$$

- Given some regularity conditions the *ML*-estimator is consistent, asymptotically efficient and asymptotically gaussian.
- Uncertainty and inference are based upon the assumptions that **to each realization of data there corresponds a different vector of parameters that maximizes the likelihood**.
- Hint for the estimation of the parameters of a DSGE-model:
  - The dimension of d_t must be greater or equal to the dimension of the structural shocks ε_t, or otherwise the residual term has a singular variance-covariance-matrix.
  - If not: Add measurement errors or additional shocks.

Maximum-Likelihood

Discussion

- Experience shows that it can be pretty hard and tricky to estimate a DSGE model via Maximum-Likelihood.
- Data is often not sufficiently informative, i.e. the likelihood is flat in some directions (identification).
- DSGE-models are always misspecified. This can lead to absurd parameter values.

Example: CGG with Maximum Likelihood

Consider the CGG model and estimate the coefficients of the stochastic process (λ, ρ)

1. via maximum likelihood: Use 1000 observations and start far from the true values.
2. via maximum likelihood: Use only 100 observations and start at the true values. Do the results change?

Exercise: RBC with Maximum Likelihood

Now consider your own RBC model and estimate α, θ and τ

1. via maximum likelihood: Use 1000 observations and start far from the true values.
2. via maximum likelihood: Use only 100 observations and start at the true values. Do the results change?

Bayesian methods

Idea

- Based upon the likelihood as well: the complete characterization of the data generating process.
- **Approach:** The parameters θ are random and data d is fixed.
- The idea is to combine known information (data) with additional believes (*prior-believes*) about the parameters and to get an expression for the conditional probability of the parameters.
- Hence, one is able to put more weight on a suspected span of the parameter space.
- Bayesian methods are a bridge between calibration and the *Maximum-Likelihood*-method:

"Bayesian Inference is a Way of Thinking, Not a Basket of Methods (Christopher Sims)"

Bayesian methods

Idea

- Likelihood-function $\mathcal{L}(d|\theta)$ is a conditional density of observed data given the parameters: $\wp(d|\theta) = \mathcal{L}(d|\theta)$.
- Denote $\wp(\theta)$ as the known prior density of the vector of parameters, then using Bayes-rule:

$$\wp(\theta|d) = \frac{\mathcal{L}(d|\theta)\wp(\theta)}{\wp(d)} = \frac{\mathcal{L}(d|\theta)\wp(\theta)}{\int \wp(\theta)\mathcal{L}(d|\theta)\ d\theta} \propto \mathcal{L}(d|\theta)\wp(\theta),$$

  with $\propto$ meaning "proportional to".
- $\wp(d)$ is the *marginal likelihood* of the data and ultimately only a constant that normalizes the expression to unity. It is independent of the parameters.
- Removing it doesn't change the form of the posterior density $\wp(\theta|d)$, it merely doesn't integrate to one.
- This non-normalized density is called *posterior-kernel* or, in logs, *log-posterior-kernel*.

Bayesian methods

Idea

- The mode is the Bayesian estimator $\widehat{\theta}_B$ of the true parameter vector:

$$\widehat{\theta}_B = \underset{\theta}{\text{argmax}} \{\log \wp(\theta|d)\} = \underset{\theta}{\text{argmax}} \{\log \mathcal{L}(d|\theta) + \log \wp(\theta)\}$$

- Procedure: Calculate the log-likelihood with the Kalman-filter and simulate the *log-posterior-kernel* through *sampling*- or *Monte-Carlo*-methods.
- In the literature – and in Dynare – the *Metropolis-Hastings-algorithm* is commonly used.
- Inference can then be conducted via the properties of the posterior-distribution.

An and Schorfheide (2007, S. 132)

The algorithm constructs a Gaussian approximation around the posterior mode and uses a scaled version of the asymptotic covariance matrix as the covariance matrix for the proposal distribution. This allows for an efficient exploration of the posterior distribution at least in the neighborhood of the mode.

- The algorithm uses the fact that under very general regularity conditions the moments of a distribution are asymptotically normal.
- It constructs a sequence of draws (Markov-chains) from a proposal density.
- This does not need to be identical with the posterior density. It is only required that the algorithm can draw samples from the whole range of the posterior density.

Bayesian methods

Metropolis-Hastings-algorithm

- The current candidate (draw) θ^* is dependent on the previous candidate $\theta^{(s-1)}$.
- Weights for all candidates are the same, however, they are only accepted with a certain probability α, calculated as the ratio of the *posterior-kernel* of the current to the one of the previous candidate.
- Due to this construct the algorithm tends to shift the draws from areas of low posterior probability to areas of high probability.
  - If $\theta^{(s-1)}$ is in an area of high posterior probability, it is likely that only candidates in the same area are accepted.
  - If $\theta^{(s-1)}$ is in an area of low posterior probability, it is very likely that new candidates are accepted.
- The covariance-matrix of the proposal distribution plays a major role, since it is important to set α neither too large nor to small.
- Current practice uses the covariance matrix of the mode $\widehat{\theta}_B$ and scales it with a factor c such that the average acceptance probability is between 20% and 30%.

Bayesian methods

Metropolis-Hastings-algorithm

1. Specify c_0, c and S.
2. Maximize $\log \mathcal{L}(d|\theta) + \log \wp(\theta)$ using numerical methods. $\widehat{\theta}_B$ denotes the mode.
3. Calculate the inverse of the Hessian evaluated at the mode, denote it with Σ_B.
4. Specify an initial value $\theta^{(0)}$ or draw it from $\mathcal{N}(\widehat{\theta}_B, c_0^2 \Sigma_B)$.

5. For $s = 1, \ldots, S$:
   - Draw θ^* from the candidate-generating distribution (proposal density) $\mathcal{N}(\mu^{(s-1)}, c^2 \Sigma_B)$.
   - Calculate the acceptance probability α:

$$\alpha \equiv \alpha\left(\theta^{(s-1)}, \theta^*\right) = \frac{\mathcal{L}\left(\theta^* | d\right) \ \wp\left(\theta^*\right)}{\mathcal{L}\left(\theta^{(s-1)} | d\right) \ \wp\left(\theta^{(s-1)}\right)}$$

   - With probability $\min\{\alpha, 1\}$ accept the jump from $\theta^{(s-1)}$ to θ^*. In other words: If $\alpha \geq 1$, set $\theta^{(s)} = \theta^*$.
   - With complementary probability don't accept the jump, i.e. draw a uniformly distributed random variable r between 0 and 1:
     - If $r \leq \alpha$ set $\theta^{(s)} = \theta^*$ (jump).
     - If $r > \alpha$ set $\theta^{(s)} = \theta^{(s-1)}$ (don't jump).

6. Estimate the posterior expectation of a function $\hbar(\theta)$ with $\frac{1}{S}\sum_{s=1}^{S}\hbar\left(\theta^{(s)}\right)$.
7. If the average acceptance probability does not yield a desirable value (typically between $20\% - 30\%$) or the algorithm does not converge, change c_0, c or S.

Bayesian methods

Remarks

- Bayesian estimation of a DSGE-model requires that the number of shocks is equivalent to the numbers of observable variables.
- Common choice for priors: gaussian, (normal, shifted or inverse) Gamma, Beta or the uniform distribution.
- Choosing a proper prior one has to consider lower and upper bounds as well as the skewness and kurtosis of the distribution.
- The results can vary due to the choice of priors and their parametrization.
- Therefore one has to check the robustness of the results:
  - Try a different parametrization.
  - Try more general priors.
  - Noninformative priors.
  - Sensitivity analysis.

Properties of the Posterior-distribution

- The posterior density combines all information about θ: information after the data is observed as well as information prior to the data.
- Bayesian estimation works for every sample size, however, it has also the following asymptotic properties:
  1. The priors become irrelevant for the determination of the posterior.
  2. The posterior converges to a degenerate distribution around the true value.
  3. The posterior is approximately gaussian.
- Using the posterior distribution one can
  - set up Bayesian confidence intervals (credibility sets),
  - calculate forecasts using the predictive-density:
    $\mathcal{L}(d_f|d) = \int \mathcal{L}(d_f|(\theta|d))d\theta = \int \mathcal{L}(d_f|\theta, d)\wp(\theta|d)d\theta$
  - compare models.

Example: Bayesian Estimation of CGG

Estimate the coefficients of the stochastic process (λ, ρ) of the CGG model via Bayesian methods. Use the beta distribution (mean set to true values, standard deviation set to 0.4) as the prior on the two autocorrelations. Use 100 observations for the estimation, 1000 MCMC replications, one MCMC chain, and 2.5 for the scale parameter.

Exercise: Bayesian Estimation of RBC

Now let's look again at your own RBC model

1. Define priors for α, θ and τ (or a different set of parameters).
2. Estimate the posterior mode using the `estimation` command and a limited sample with 200 observations. How man observable variables do you need? Check the posterior mode using `mode_check`. If you get errors due to a non-positive definite Hessian, try a different optimization algorithm or change the initial values.
3. If you are satisfied with the posterior mode, approximate the posterior distribution using the the Metropolis- Hastings-Algorithmus with 3×5000 iterations. If it does not converge to the (ergodic) posterior-distribution, repeat the algorithm without discarding the previous draws.
4. How robust are the results regarding the specification of the priors? Repeat the estimation of the posterior-mode for different priors.

Discussion of full information estimators

- More restrictive assumptions are needed compared to the limited information estimation: specification of the distribution of the schocks, i.e. the likelihood.
- Advantages of a *Maximum-Likelihood*-estimation lie in the full characterization of the data-generating-process and the exact, consistent and efficient estimation of the parameters.
- "Dilemma of absurd parameter estimates": Problem of the ML-estimation due to wrong distributional assumptions, problems in the optimization algorithm or non-separable identifiable parameters.
- Even transformations, upper and lower bounds, etc. are only limited to help overcome this problem, when the likelihood is flat.

Discussion of full information estimators

- This is where Bayesian methods come in and bridge the gap between calibration and the *ML-principle*.
- Considering priors one can incorporate additional information into a model.
- "Dilemma of absurd parameter estimates": Even with Bayesian means it is not possible to estimate these parameters (the posterior looks almost the same as the prior), but one can assign probability such that these parameters are very unlikely.

$\Rightarrow$ Using priors one can exclude these absurd parameter estimates.

- Nevertheless the point of robustness and identification of the parameters remains a critical topic.

Discussion of full information estimators

An und Schorfheide (2006, S.124)

Once one acknowledges that the DSGE model provides merely an approximation to the law of motion of the time series (...), then it seems reasonable to assume that there need not exist a single parameter vector (...), that delivers, say, the "true" intertemporal substitution elasticity or price adjustment costs and, simultaneously, the most precise impulse responses to a technology or monetary policy shock. Each estimation method is associated with a particular measure of discrepancy between the "true" law of motion and the class of approximating models.

Identification

General problem

For a mathematical expression with many conditions and parameters, but only a limited sample, there can exist different combinations of parameters that yield the same result and a similar dataset.

- Consider two vectors of parameters θ_1 and θ_2 for which

$$\mathcal{L}(d|\theta_1) = \mathcal{L}(d|\theta_2).$$

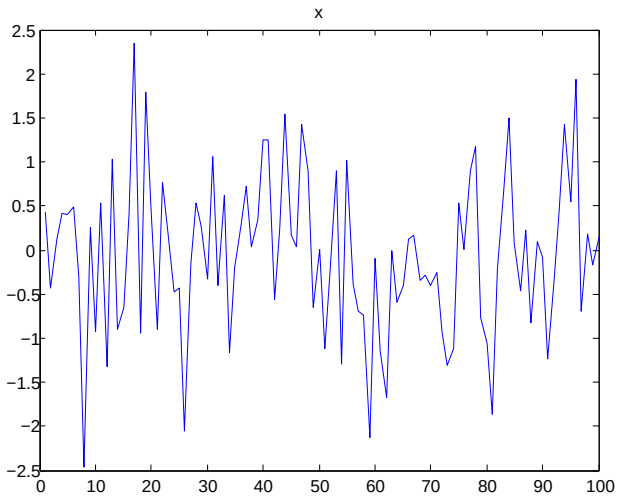
- If $\theta_1 = \theta_2$, then there is identification. If, however, $\theta_1 \neq \theta_2$, then one does not know which vector of parameters has generated the data.

Identification

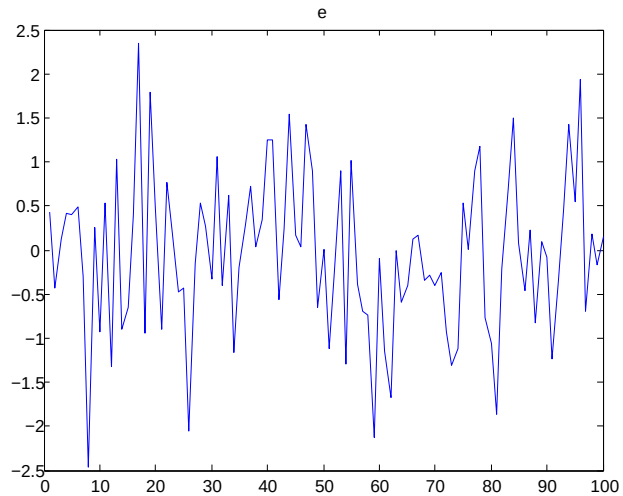
- Simple example: Consider the following ARMA(1,1)-process

$$x_t = \phi_1 x_{t-1} + \varepsilon_t - \phi_2 \varepsilon_{t-1}, \text{ with } \varepsilon \sim N(0, \sigma^2)$$

with parameter vector $\theta = (\phi_1, \phi_2, \sigma)'$:



(a) $\theta_0 = (0.4, 0.4, 1)'$



(b) $\theta_1 = (0, 0, 1)'$

- Obviously both models generate the same data (as long as the shocks ε_t are the same). θ_0 and θ_1 are observationally equivalent.
- Note $\sigma = 1$ is **partially identifiable**.

Identification

- Identification is a mathematical problem (injectivity).
- Identification problems arise if distinct parameter values do not lead to distinct probability distributions of the data.
- Drawing inferences from the probability distribution leads thus to wrong conclusions from estimation and inference.
- When identification fails, properties of estimators change.
- Even with an infinite sample it is not possible to pin down some parameters, no matter what estimation procedure one uses.
- Identification tests: Order and rank conditions, via autocovariances, spectral densities, information matrix, imposing restrictions, Bayesian methods . . . can be studied prior to estimation

References

Recommended Readings

- DYNARE USER GUIDE
- DeJong/Dave (2011) - Structural Macroeconometrics
- An and Schorfheide (2007) - Bayesian Analysis of DSGE Models