Maximum Likelihood Estimation

Maximum Likelihood

Tools for Macroeconomists: The essentials

Petr Sedláček

Maximum Likelihood

Estimating structural parameters

- up to now, we assumed that model parameters are known
- we can also estimate them with Maximum Likelihood (ML)
  - i.e. given data on y_t and initial conditions
  - estimate $\mathbf{\Psi} = [H, F, Q, R]$
- the Kalman filter is particularly convenient for this task

Maximum Likelihood

Main Idea

Preliminaries

- if $\zeta_{1|0}$ is Gaussian and $\{w_t, v_t\}_{t=1}^{T}$ are Gaussian

Preliminaries

- if $\zeta_{1|0}$ is Gaussian and $\{w_t, v_t\}_{t=1}^{T}$ are Gaussian
- $\rightarrow$ distribution of y_t conditional on $\mathcal{Y}_{t-1}$ is also Gaussian

$$
\begin{aligned}
\widetilde{y}_{t|t-1}|\mathcal{Y}_{t-1} \sim & N(0, H'P_{t|t-1}H + R) \\
y_t|\mathcal{Y}_{t-1} \sim & N(H'\widehat{\zeta}_{t|t-1}, H'P_{t|t-1}H + R)
\end{aligned}
$$

Preliminaries

- given values of $\Psi \rightarrow$ calculate mean and variance of y
- we know the distribution of y
- $\rightarrow$ calculate the probability (likelihood) of $(y_1, ..., y_T)$

Likelihood function

the likelihood of a given (Gaussian) observation is

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

Likelihood function

the likelihood of a given (Gaussian) observation is

$$\frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

in our case this is

Likelihood function

the likelihood of a given (Gaussian) observation is

$$\frac{1}{\sigma\sqrt{2\pi}}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}$$

in our case this is

$$\begin{aligned} f(y_t|\mathcal{Y}_{t-1};\Psi) =& (2\pi)^{-1/2}(H'P_{t|t-1}H+R)^{-1/2}x \\ & \exp\left\{-1/2(y_t-\widehat{y}_{t|t-1})'(H'P_{t|t-1}H+R)^{-1}(y_t-\widehat{y}_{t|t-1})\right\} \\ & \text{for } t=1,...,T \end{aligned} \tag{1}$$

Likelihood function

we are interested in the likelihood of the entire sample

we are interested in the likelihood of the entire sample

- because forecast errors are orthogonal to each other

we are interested in the likelihood of the entire sample

- because forecast errors are orthogonal to each other
- $\mathcal{L}(\mathcal{Y}_t|\Psi) = f(y_0;\Psi)\Pi_{t=1}^{T} f(y_t|\mathcal{Y}_{t-1};\Psi)$

Likelihood function

we are interested in the likelihood of the entire sample

- because forecast errors are orthogonal to each other
- $\mathcal{L}(\mathcal{Y}_t|\Psi) = f(y_0;\Psi)\Pi_{t=1}^{T} f(y_t|\mathcal{Y}_{t-1};\Psi)$

it is convenient to work with the sample log-likelihood:

$$\log \mathcal{L}(\mathcal{Y}_t|\Psi) = \log f(y_0) + \sum_{t=1}^{T} \log f(y_t|\mathcal{Y}_{t-1};\Psi)$$

Why is the Kalman filter convenient?

- the Kalman filter produces $\widehat{y}_{t|t-1}$ and $P_{t|t-1}$
- the (log)-likelihood is easy to construct with the Kalman filter
- one can then maximize it with respect to the parameters $\boldsymbol{\Psi}$
- this will be your task in the afternoon session

Maximum Likelihood

Back to DSGE Models

Neoclassical growth model

- representative household maximizing expected lifetime utility
- household owns production technology
- capital is the only factor of production
- resources spent on consumption and investment into capital
- each period existing capital depreciates at certain rate
- production subject to exogenous fluctuations in productivity

Production

$$y_t = Z_t k_t^{\alpha}$$

Production

$$y_t = Z_t k_t^{\alpha}$$

$$Z_t = 1 - \rho + \rho Z_{t-1} + \epsilon_t$$

$$\mathbb{E}\epsilon_t = 0$$

$$\mathbb{E}\epsilon_t^2 = \sigma_z^2$$

Household decision

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \log(c_t)$$

Household decision

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \log(c_t)$$

$$\text{s.t.} \quad c_t + k_{t+1} = y_t + (1-\delta)k_t$$

Household decision

$$\max_{\{c_t, k_{t+1}\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \log(c_t)$$

$$\text{s.t.} \quad c_t + k_{t+1} = y_t + (1-\delta)k_t$$

$$k_0 \quad \text{given}$$

$$Z_0 \quad \text{given}$$

Equilibrium conditions

$$c_t^{-1} = \mathbb{E}_t\left[\beta c_{t+1}^{-1}(\alpha z_{t+1} k_t^{\alpha-1} + 1 - \delta)\right]$$

$$c_t + k_t = z_t k_{t-1}^{\alpha} + (1-\delta)k_{t-1}$$

$$z_t = 1 - \rho + \rho z_{t-1} + \epsilon_t$$

$$\epsilon_t \sim N(0, \sigma^2)$$

Solution

What is the solution?

Solution

What is the solution?

- a sequence $\{c_t, k_{t+1}\}_{t=0}^{\infty}$
- maximizing the expected discounted sum of per-period utilities

Sounds like a tough problem!

Solution

What is the solution?

- a sequence $\{c_t, k_{t+1}\}_{t=0}^{\infty}$
- maximizing the expected discounted sum of per-period utilities

Sounds like a tough problem!

- different $k_0 \rightarrow$ optimal sequences different!
- different realizations of $Z_t \rightarrow$ optimal sequences different!

Linearized version

$$
\begin{aligned}
k_t =& \bar{k} + a_{kk}(k_{t-1} - \bar{k}) + a_{kz}(z_t - \bar{z}) \\
z_t =& 1 - \rho + \rho z_{t-1} + \epsilon_t \\
& \epsilon_t \sim N(0, \sigma^2) \\
& k_0, z_0 \text{ given}
\end{aligned}
$$

Linearized version

$$
\begin{aligned}
k_t =& \bar{k} + a_{kk}(k_{t-1} - \bar{k}) + a_{kz}(z_t - \bar{z}) \\
z_t =& 1 - \rho + \rho z_{t-1} + \epsilon_t \\
& \epsilon_t \sim N(0, \sigma^2) \\
& k_0, z_0 \text{ given}
\end{aligned}
$$

- a_{kk}, a_{kz} and $\bar{k}$ depend on structural parameters $\boldsymbol{\Psi}$

Linearized version

$$
\begin{aligned}
k_t =& \bar{k} + a_{kk}(k_{t-1} - \bar{k}) + a_{kz}(z_t - \bar{z}) \\
z_t =& 1 - \rho + \rho z_{t-1} + \epsilon_t \\
& \epsilon_t \sim N(0, \sigma^2) \\
& k_0, z_0 \text{ given}
\end{aligned}
$$

- a_{kk}, a_{kz} and $\bar{k}$ depend on structural parameters Ψ
- $\Psi = [\alpha, \beta, \delta, \rho, \sigma, z_0]$

Maximum Likelihood

Estimating Structural Parameters

Estimating structural parameters

consider estimating the structural parameters using ML

Estimating structural parameters

consider estimating the structural parameters using ML

- how to write down the likelihood of the model?

Estimating structural parameters

consider estimating the structural parameters using ML

- how to write down the likelihood of the model?
  - for Kalman filter, we still need to figure out H, F

Estimating structural parameters

consider estimating the structural parameters using ML

- how to write down the likelihood of the model?
  - for Kalman filter, we still need to figure out H, F
  - can we do something else (simpler) instead?

Simple case of evaluating the likelihood

Simple case of evaluating the likelihood

$$
\begin{aligned}
k_t =& \overline{k} + a_{kk}(k_{t-1} - \overline{k}) + a_{kz}(z_t - \bar{z}) \\
z_t =& (1-\rho)\bar{z} + \rho z_{t-1} + \epsilon_t
\end{aligned}
$$

Simple case of evaluating the likelihood

$$
\begin{aligned}
k_t =& \bar{k} + a_{kk}(k_{t-1} - \bar{k}) + a_{kz}(z_t - \bar{z}) \\
z_t =& (1-\rho)\bar{z} + \rho z_{t-1} + \epsilon_t
\end{aligned}
$$

$$
\begin{aligned}
z_t =& \bar{z} + \frac{k_t - \bar{k} - a_{kk}(k_{t-1} - \bar{k})}{a_{kz}} \\
\epsilon_t =& \bar{z}(\rho - 1) + z_t - \rho z_{t-1}
\end{aligned}
$$

Simple case of evaluating the likelihood

$$\begin{aligned} k_t =& \bar{k} + a_{kk}(k_{t-1} - \bar{k}) + a_{kz}(z_t - \bar{z}) \\ z_t =& (1-\rho)\bar{z} + \rho z_{t-1} + \epsilon_t \end{aligned}$$

$$\begin{aligned} z_t =& \bar{z} + \frac{k_t - \bar{k} - a_{kk}(k_{t-1} - \bar{k})}{a_{kz}} \\ \epsilon_t =& \bar{z}(\rho - 1) + z_t - \rho z_{t-1} \end{aligned}$$

- given a guess of Ψ and given k_0, k_1 and z_0
- $\rightarrow$ calculate z_1 $\rightarrow$ calculate ϵ_1 $\rightarrow$ calculate z_2 etc.

Log-likelihood

in this case, the log-likelihood is simply

Log-likelihood

in this case, the log-likelihood is simply

$$\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{\epsilon_t(\Psi)^2}{2\sigma^2}$$

Maximum Likelihood

Model in State-Space Form

Putting model into state-space form

- the neoclassical growth model is relatively simple
- for more complex models, policy function inversion is tough

Putting model into state-space form

- the neoclassical growth model is relatively simple
- for more complex models, policy function inversion is tough
- but we know that
    - the Kalman filter is convenient for likelihood construction
    - because it produces $y_t - \widehat{y}_{t|t-1}$ and $P_{t|t-1}$

Putting model into state-space form

- the neoclassical growth model is relatively simple
- for more complex models, policy function inversion is tough
- but we know that
  - the Kalman filter is convenient for likelihood construction
  - because it produces $y_t - \widehat{y}_{t|t-1}$ and $P_{t|t-1}$
- the question is how to cast DSGE model into state-space form

DSGE mode in state-space form

$$
\begin{aligned}
\zeta_{t+1} =& F\zeta_t + v_{t+1}, & \mathbb{E}(v_t, v_t') = Q \quad \forall t \\
y_t =& H'\zeta_t + w_t, & \mathbb{E}(w_t, w_t') = R \quad \forall t
\end{aligned}
$$

- what is the observable and what are the unobserved states?

DSGE mode in state-space form

DSGE MODE IN STATE-SPACE FORM

$$\begin{bmatrix} k_t - \bar{k} \\ z_{t+1} - \bar{z} \end{bmatrix} = \begin{bmatrix} a_{kk} & a_{kz} \\ 0 & \rho \end{bmatrix} \begin{bmatrix} k_{t-1} - \bar{k} \\ z_t - \bar{z} \end{bmatrix} + \begin{bmatrix} 0 \\ \epsilon_{t+1} \end{bmatrix}$$

$$k_{t-1} - \bar{k} = [1 \;\; 0] \begin{bmatrix} k_{t-1} - \bar{k} \\ z_t - \bar{z} \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix}$$

Different observables?

What if we don't have data on capital, but only on output (p_t)?

Different observables?

What if we don't have data on capital, but only on output (p_t)?

$$\begin{bmatrix} k_t - \overline{k} \\ z_{t+1} - \overline{z} \\ p_t - \overline{p} \end{bmatrix} = \begin{bmatrix} a_{kk} & a_{kz} & 0 \\ 0 & \rho & 0 \\ \alpha\overline{z}\overline{k}^{\alpha-1} & \overline{k}^{\alpha} & 0 \end{bmatrix} \begin{bmatrix} k_{t-1} - \overline{k} \\ z_t - \overline{z} \\ p_{t-1} - \overline{p} \end{bmatrix} + \begin{bmatrix} 0 \\ \epsilon_{t+1} \\ 0 \end{bmatrix}$$

$$[p_{t-1} - \overline{p}] = [0 \;\; 0 \;\; 1] \begin{bmatrix} k_{t-1} - \overline{k} \\ z_t - \overline{z} \\ p_{t-1} - \overline{p} \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix}$$

Different observables?

An alternative state-space for the same setup would be

Different observables?

An alternative state-space for the same setup would be

$$\begin{bmatrix} k_t - \bar{k} \\ z_{t+1} - \bar{z} \end{bmatrix} = \begin{bmatrix} a_{kk} & a_{kz} \\ 0 & \rho \end{bmatrix} \begin{bmatrix} k_{t-1} - \bar{k} \\ z_t - \bar{z} \end{bmatrix} + \begin{bmatrix} 0 \\ \epsilon_{t+1} \end{bmatrix}$$

$$[p_t - \bar{p}] = \begin{bmatrix} \alpha \bar{z} \bar{k}^{\alpha-1} & \bar{k}^{\alpha} \end{bmatrix} \begin{bmatrix} k_{t-1} - \bar{k} \\ z_t - \bar{z} \end{bmatrix} + \begin{bmatrix} 0 \end{bmatrix}$$

Maximum Likelihood

Observables and Singularities

What if we also observe productivity?

What if we observe capital and also productivity (z_t)?

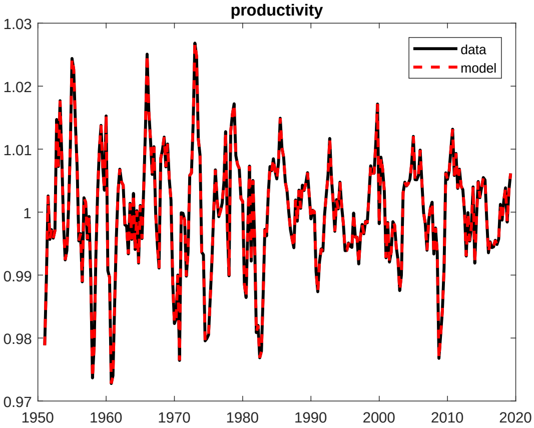
What if we also observe productivity?

What if we observe capital and also productivity (z_t)?

- if our model is the true data-generating process
  - $\rightarrow$ likelihood = 1 for true Ψ and 0 otherwise
- if our model is not the true data-generating process
  - $\rightarrow$ likelihood = 0 for any values of Ψ

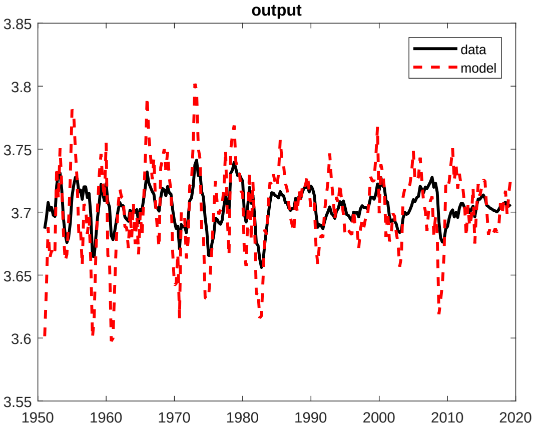
Ways out?

Neoclassical growth model estimated on labor productivity



Ways out?

Neoclassical growth model estimated on labor productivity



What if we also observe productivity?

To understand the above notice that with productivity data

What if we also observe productivity?

To understand the above notice that with productivity data

- 4 periods are enough to pin down $\bar{k}$, ρ, a_{kk}, a_{kz}

What about the other periods?

What if we also observe productivity?

To understand the above notice that with productivity data

- 4 periods are enough to pin down $\bar{k}$, ρ, a_{kk}, a_{kz}

What about the other periods?

- shocks adjust such that productivity is matched perfectly

$$z_t = 1 - \rho + \rho z_{t-1} + \epsilon_t$$

What if we also observe productivity?

What if we now add data on capital?

What if we also observe productivity?

What if we now add data on capital?

- same logic applies:
  - 4 productivity observations enough to pin down $\overline{k}$, ρ, a_{kk}, a_{kz}
  - rest of productivity time-series matched by shocks

What if we also observe productivity?

What if we now add data on capital?

- same logic applies:
    - 4 productivity observations enough to pin down $\overline{k}$, ρ, a_{kk}, a_{kz}
    - rest of productivity time-series matched by shocks
- capital data consistent with model only if model is true DGP!

$$k_t = \overline{k} + a_{kk}(k_{t-1} - \overline{k}) + a_{kz}(z_t - \bar{z})$$

Ways out?

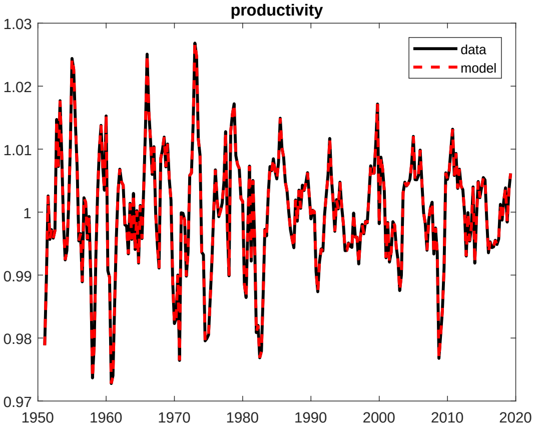
Ways out?

can we simply add an error term?

$$k_t = \overline{k} + a_{kk}(k_{t-1} - \overline{k}) + a_{kz}(z_t - \bar{z}) + u_t$$

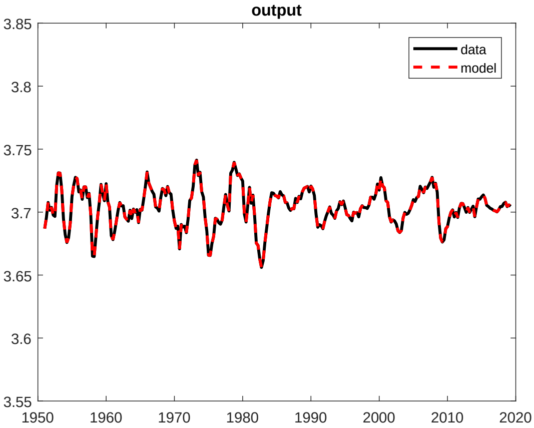
Ways out?

Neoclassical growth model estimated on labor productivity & GDP



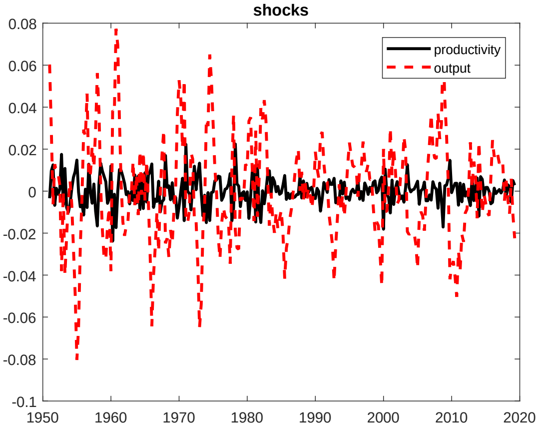
Ways out?

Neoclassical growth model estimated on labor productivity & GDP



Ways out?

Neoclassical growth model estimated on labor productivity & GDP



WAYS OUT?

- if u_t is a structural shock (e.g. preferences)
  - $\rightarrow$ its law of motion influences policy function ($\bar{k}$, a_{kk}, a_{kz})
- if u_t is measurement error
  - OK from an econometric point of view
  - but is it truly measurement error?

Ways out?

what if we also observe consumption (but not productivity)?

$$k_t = \bar{k} + a_{kk}(k_{t-1} - \bar{k}) + a_{kz}(z_t - \bar{z})$$

$$c_t = \bar{c} + a_{ck}(k_{t-1} - \bar{k}) + a_{cz}(z_t - \bar{z})$$

Ways out?

what if we also observe consumption (but not productivity)?

$$k_t = \overline{k} + a_{kk}(k_{t-1} - \overline{k}) + a_{kz}(z_t - \overline{z})$$

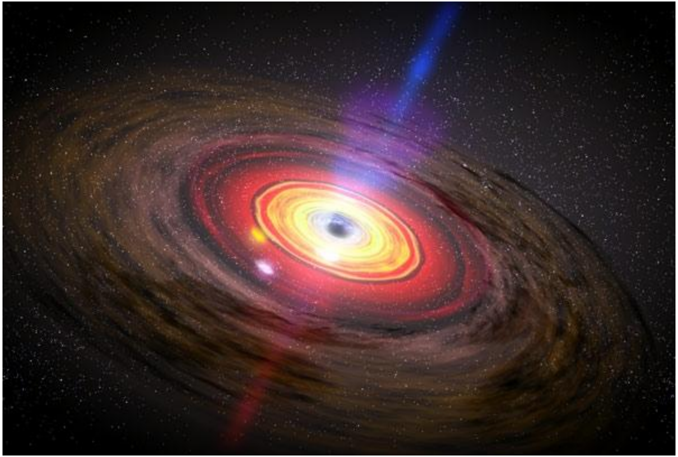
$$c_t = \overline{c} + a_{ck}(k_{t-1} - \overline{k}) + a_{cz}(z_t - \overline{z})$$

won't work!

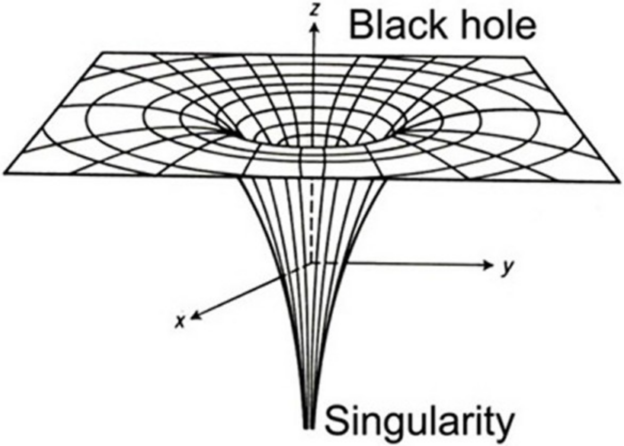
- given Ψ, you can back out z_t from both equations
- with actual data this will give inconsistent answers

Singularity problem

Singularity problem



Singularity problem



Singularity problem

- (stochastic) singularity:
  - many endogenous variables ...
  - driven by a smaller number of structural shocks

Singularity problem

- (stochastic) singularity:
  - many endogenous variables ...
  - driven by a smaller number of structural shocks
- $\rightarrow$ some observables are linear combinations of others

Singularity problem

- (stochastic) singularity:
    - many endogenous variables ...
    - driven by a smaller number of structural shocks
- $\rightarrow$ some observables are linear combinations of others
- $\rightarrow$ the var-covar matrix of observables is *singular*
- what is the problem mathematically?

General rule

- for every observable, you need at least one unobservable shock
- (letting them be measurement error is hard to defend)

General rule

- for every observable, you need at least one unobservable shock
- (letting them be measurement error is hard to defend)

Note that:

- more shocks (measurement errors) than observables is OK
- the choice of observables for estimation is not innocent
- there are ways to choose observables carefully
  - see e.g. Canova, Ferroni, Matthes (2012)

Maximum Likelihood

Taking Stock

Taking stock

Estimating DSGE models with Maximum Likelihood

- Kalman filter very convenient
  - delivers objects needed to construct likelihood function
  - allows for estimation of underlying structural shocks

Taking stock

Estimating DSGE models with Maximum Likelihood

- Kalman filter very convenient
  - delivers objects needed to construct likelihood function
  - allows for estimation of underlying structural shocks
- beware of stochastic singularity ("one shock per observable")


That's all Folks!