Linear state-space models

A. State-space representation of a dynamic system

Consider following model

State equation:

$$\underset{r\times 1}{\boldsymbol{\xi}_{t+1}} = \underset{r\times r}{\mathbf{F}}\,\underset{r\times 1}{\boldsymbol{\xi}_t} + \underset{r\times 1}{\mathbf{v}_{t+1}}$$

Observation equation:

$$\underset{n\times 1}{\mathbf{y}_t} = \underset{n\times k}{\mathbf{A}'}\,\underset{k\times 1}{\mathbf{x}_t} + \underset{n\times r}{\mathbf{H}'}\,\underset{r\times 1}{\boldsymbol{\xi}_t} + \underset{n\times 1}{\mathbf{w}_t}$$

Observed variables: $\mathbf{y}_t, \mathbf{x}_t$

Unobserved variables: $\boldsymbol{\xi}_t, \mathbf{v}_t, \mathbf{w}_t$

Matrices of parameters: $\mathbf{F}, \mathbf{A}, \mathbf{H}$

$$\begin{bmatrix} \mathbf{v}_t \\ \mathbf{w}_t \end{bmatrix} \sim \text{i.i.d. } N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

$\mathbf{Q} = r \times r$

$\mathbf{R} = n \times n$

Example 1:

$$\xi_{t+1} = \begin{bmatrix} \phi_1 & \phi_2 & \cdots & \phi_{r-1} & \phi_r \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \xi_t + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$\xi_{j,t+1} = L^{j-1}\xi_{1t} \quad \text{for } j = 2, 3, \ldots, r$$

$$\xi_{1,t+1} = \phi_1 \xi_{1t} + \phi_2 L^1 \xi_{1t} + \phi_3 L^2 \xi_{1t} + \cdots + \phi_p L^{p-1} \xi_{1t} + \varepsilon_{t+1}$$

$$\phi(L)\xi_{1,t+1} = \varepsilon_{t+1}$$

Observation equation:

$$y_t = \mu + \begin{bmatrix} 1 & \theta_1 & \theta_2 & \cdots & \theta_{r-1} \end{bmatrix} \xi_t$$

$$y_t - \mu = \theta(L)\xi_{1t}$$

put together with state equation:

$$\phi(L)\xi_{1t} = \varepsilon_t$$

$$\phi(L)(y_t - \mu) = \theta(L)\varepsilon_t$$

Conclusion: any ARMA process can be written as a state-space model.

Example 2:

C_t = state of business cycle

χ_{it} = idiosyncratic component for sector i

C_t, χ_{it} unobserved

y_{it} = growth in sector i (observed)

$$\boldsymbol{\xi}_t = (C_t, \chi_{1t}, \chi_{2t}, \ldots, \chi_{nt})'$$

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}\boldsymbol{\xi}_t + \mathbf{v}_{t+1}$$

$$\mathbf{F} = \begin{bmatrix} \phi_C & 0 & 0 & \cdots & 0 \\ 0 & \phi_1 & 0 & \cdots & 0 \\ 0 & 0 & \phi_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \phi_r \end{bmatrix}$$

Observation equation:

$$\begin{bmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{nt} \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_n \end{bmatrix} + \begin{bmatrix} \gamma_1 & 1 & 0 & \cdots & 0 \\ \gamma_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ \gamma_n & 0 & 0 & \cdots & 1 \end{bmatrix} \boldsymbol{\xi}_t$$

Purpose of state-space representation: state vector ξ_t contains all information about system dynamics and forecasting.

$$\xi_{t+1} = \mathbf{F}\xi_t + \mathbf{v}_{t+1}$$

$$\mathbf{y}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\xi_t + \mathbf{w}_t$$

$$E(\mathbf{y}_{t+j}|\xi_t, \xi_{t-1}, \ldots, \xi_1, \mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{x}_{t+j}, \mathbf{x}_{t+j-1}, \ldots, \mathbf{x}_1) = \mathbf{A}'\mathbf{x}_{t+j} + \mathbf{H}'\mathbf{F}^j\xi_t$$

Linear state-space models

A. State-space representation of a dynamic system
B. Kalman filter

Purpose of Kalman filter: calculate distribution of ξ_t conditional on

$$\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1\}$$

$$\xi_t|\Omega_t \sim N(\hat{\xi}_{t|t}, \mathbf{P}_{t|t})$$

$$\xi_{t+1} = \mathbf{F}\xi_t + \mathbf{v}_{t+1}$$
$$\mathbf{y}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\xi_t + \mathbf{w}_t$$
$$\begin{bmatrix} \mathbf{v}_t \\ \mathbf{w}_t \end{bmatrix} \sim \text{i.i.d. } N\left(\begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix} \right)$$

Begin with the prior:

$\xi_0 \sim N(\hat{\xi}_{0|0}, \mathbf{P}_{0|0})$

$\hat{\xi}_{0|0} =$ prior best guess as to value of ξ_0

$\mathbf{P}_{0|0} =$ uncertainty about this guess

(much uncertainty $=$ large diagonal elements of $\mathbf{P}_{0|0}$)

$$\xi_1 = \mathbf{F}\xi_0 + \mathbf{v}_1$$
$$\xi_1 \sim N(\hat{\xi}_{1|0}, \mathbf{P}_{1|0})$$
$$\hat{\xi}_{1|0} = \mathbf{F}\hat{\xi}_{0|0}$$
$$\mathbf{P}_{1|0} = \mathbf{F}\mathbf{P}_{0|0}\mathbf{F}' + \mathbf{Q}$$

Useful result: suppose that

$$\begin{bmatrix} \mathbf{y}_1|\mathbf{x} \\ \mathbf{y}_2|\mathbf{x} \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$$

where $\boldsymbol{\mu}_i$ and $\boldsymbol{\Sigma}_{ij}$ may depend on $\mathbf{x}$. Then

$$\mathbf{y}_2|\mathbf{y}_1,\mathbf{x} \sim N(\mathbf{m}^*,\mathbf{M}^*)$$

$$\mathbf{m}^* = \boldsymbol{\mu}_2 + \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}(\mathbf{y}_1 - \boldsymbol{\mu}_1)$$

$$\mathbf{M}^* = \boldsymbol{\Sigma}_{22} - \boldsymbol{\Sigma}_{21}\boldsymbol{\Sigma}_{11}^{-1}\boldsymbol{\Sigma}_{12}$$

Here

$$\begin{bmatrix} \mathbf{y}_1|\mathbf{x}_1,\Omega_0 \\ \xi_1|\mathbf{x}_1,\Omega_0 \end{bmatrix} \sim N\left(\begin{bmatrix} \boldsymbol{\mu}_1 \\ \boldsymbol{\mu}_2 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}\right)$$

$$\boldsymbol{\mu}_2 = \hat{\xi}_{1|0} \qquad \boldsymbol{\Sigma}_{22} = \mathbf{P}_{1|0}$$

$$\boldsymbol{\mu}_1 = \mathbf{A}'\mathbf{x}_1 + \mathbf{H}'\hat{\xi}_{1|0} \qquad \boldsymbol{\Sigma}_{11} = \mathbf{H}'\mathbf{P}_{1|0}\mathbf{H} + \mathbf{R}$$

$$\boldsymbol{\Sigma}_{21} = \mathbf{P}_{1|0}\mathbf{H}$$

Hence

$$\xi_1|\mathbf{y}_1,\mathbf{x}_1,\Omega_0 = \xi_1|\Omega_1 \sim N(\hat{\xi}_{1|1},\mathbf{P}_{1|1})$$

$$\hat{\xi}_{1|1} = \hat{\xi}_{1|0} + \mathbf{P}_{1|0}\mathbf{H}(\mathbf{H}'\mathbf{P}_{1|0}\mathbf{H} + \mathbf{R})^{-1} \times \left(\mathbf{y}_1 - \mathbf{A}'\mathbf{x}_1 - \mathbf{H}'\hat{\xi}_{1|0}\right)$$

$$\mathbf{P}_{1|1} = \mathbf{P}_{1|0} - \mathbf{P}_{1|0}\mathbf{H}(\mathbf{H}'\mathbf{P}_{1|0}\mathbf{H} + \mathbf{R})^{-1}\mathbf{H}'\mathbf{P}_{1|0}$$

Identical calculations: if $\xi_t|\Omega_t \sim N(\hat{\xi}_{t|t}, \mathbf{P}_{t|t})$,

then $\xi_{t+1}|\Omega_{t+1} \sim N(\hat{\xi}_{t+1|t+1}, \mathbf{P}_{t+1|t+1})$

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q}$$

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|t} - \mathbf{P}_{t+1|t}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t+1|t}\mathbf{H} + \mathbf{R})^{-1}\mathbf{H}'\mathbf{P}_{t+1|t}$$

$$\hat{\xi}_{t+1|t} = \mathbf{F}\hat{\xi}_{t|t}$$

$$\hat{\boldsymbol{\varepsilon}}_{t+1|t} = \mathbf{y}_{t+1} - \mathbf{A}'\mathbf{x}_{t+1} - \mathbf{H}'\hat{\xi}_{t+1|t}$$

$$\hat{\xi}_{t+1|t+1} = \hat{\xi}_{t+1|t} + \mathbf{P}_{t+1|t}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t+1|t}\mathbf{H} + \mathbf{R})^{-1}\hat{\boldsymbol{\varepsilon}}_{t+1|t}$$

Iterating on these calculations for $t = 1, 2, \ldots, T$ to produce the sequences $\{\mathbf{P}_{t|t}\}_{t=1}^{T}$ and $\left\{\hat{\xi}_{t|t}\right\}_{t=1}^{T}$ is called the Kalman filter.

$\hat{\xi}_{t|t}$ is the expectation of ξ_t given observation of $\Omega_t = \{\mathbf{y}_t, \mathbf{y}_{t-1}, \ldots, \mathbf{y}_1, \mathbf{x}_t, \mathbf{x}_{t-1}, \ldots, \mathbf{x}_1\}$.

$$\mathbf{P}_{t|t} = E\left(\hat{\xi}_{t|t} - \xi_t\right)\left(\hat{\xi}_{t|t} - \xi_t\right)'$$

where these expectations condition on the values of $\mathbf{F}, \mathbf{Q}, \mathbf{A}, \mathbf{H}, \mathbf{R}$.

Forecasting:

$\mathbf{y}_t = \mathbf{A}'\mathbf{x}_t + \mathbf{H}'\xi_t + \mathbf{w}_t$

$E(\mathbf{y}_{t+j}|\Omega_t, \mathbf{x}_{t+j}, \mathbf{F}, \mathbf{Q}, \mathbf{A}, \mathbf{H}, \mathbf{R})$

$= \mathbf{A}'\mathbf{x}_{t+j} + \mathbf{H}'\mathbf{F}^j\hat{\xi}_{t|t}$

MSE for $j = 1$:

$E(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})(\mathbf{y}_{t+1} - \hat{\mathbf{y}}_{t+1|t})'$

$= \mathbf{H}'\mathbf{P}_{t+1|t}\mathbf{H} + \mathbf{R}$

Smoothed inference: might also want to form inference about ξ_t using all the data Ω_T:

$\xi_t|\Omega_T \sim N(\hat{\xi}_{t|T}, \mathbf{P}_{t|T})$

To derive formula, consider instead

$\xi_t|\xi_{t+1}, \Omega_t \sim N(\xi^*_{t|t}, \mathbf{P}^*_{t|t})$

Same kind of derivation as for Kalman filter establishes that

$\xi^*_{t|t} = \hat{\xi}_{t|t} + \mathbf{J}_t(\xi_{t+1} - \hat{\xi}_{t+1|t})$

$\mathbf{J}_t = \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}^{-1}_{t+1|t}$

$\mathbf{P}^*_{t|t} = \mathbf{P}_{t|t} - \mathbf{J}_t\mathbf{F}\mathbf{P}_{t|t}$

Generalization: what if $\mathbf{P}_{t+1|t}$ is singular?

If $\mathbf{P}_{t+1|t}$ is singular, then some linear combinations of ξ_{t+1} can be forecast perfectly from Ω_t, implying inference about ξ_t given Ω_t and these linear combinations of ξ_{t+1} is identical to inference about ξ_t given Ω_t alone.

Let ξ_t be $(r \times 1)$ and let the rank of $\mathbf{P}_{t+1|t}$ be $s \leq r$. Define the $(s \times 1)$ vector $\xi_t^{**} = \mathbf{H}^{**}\xi_t$ for an arbitrary $(s \times r)$ matrix $\mathbf{H}^{**}$ such that $\mathbf{P}_{t+1|t}^{**} \equiv \mathbf{H}^{**}\mathbf{P}_{t+1|t}\mathbf{H}^{**\prime}$ has rank s. For example, ξ_t^{**} might be the first s elements of ξ_t in which case $\mathbf{P}_{t+1|t}^{**}$ would be first s rows and columns of $\mathbf{P}_{t+1|t}$. Then $\xi_t|\xi_{t+1}, \Omega_t$ has same distribution as $\xi_t|\xi_{t+1}^{**}, \Omega_t$.

Generalization of previous results for singular $\mathbf{P}_{t+1|t}$:

$$\xi_{t|t}^{*} = \hat{\xi}_{t|t} + \mathbf{J}_t^{**}(\xi_{t+1}^{**} - \hat{\xi}_{t+1|t}^{**})$$
$$\mathbf{J}_t^{**} = \mathbf{P}_{t|t}(\mathbf{H}^{**}\mathbf{F})'\mathbf{P}_{t+1|t}^{**-1}$$
$$\mathbf{P}_{t|t}^{*} = \mathbf{P}_{t|t} - \mathbf{J}_t^{**}\mathbf{H}^{**}\mathbf{F}\mathbf{P}_{t|t}$$

Next suppose that, in addition to ξ_{t+1}, we had also observed $\mathbf{y}_{t+1}, \mathbf{y}_{t+2}, \ldots, \mathbf{y}_T$. This would contain no more information about ξ_t than was provided by ξ_{t+1} and Ω_t alone:

$$\xi_t | \xi_{t+1}, \Omega_T \sim N(\xi^*_{t|t}, \mathbf{P}^*_{t|t})$$

for the same $\xi^*_{t|t}, \mathbf{P}^*_{t|t}$.

And since

$$E(\xi_t | \xi_{t+1}, \Omega_T) = \hat{\xi}_{t|t} + \mathbf{J}_t^{**}(\xi^{**}_{t+1} - \hat{\xi}^{**}_{t+1|t}),$$

it follows from law of iterated expectations that

$$E(\xi_t | \Omega_T) = \hat{\xi}_{t|t} + \mathbf{J}_t^{**}(\hat{\xi}^{**}_{t+1|T} - \hat{\xi}^{**}_{t+1|t})$$

which we can calculate by iterating backwards for $t = T-1, T-2, \ldots$

The MSE's of these smoothed inferences are given by

$$E(\xi_t - \hat{\xi}_{t|T})(\xi_t - \hat{\xi}_{t|T})' = \mathbf{P}_{t|T}$$

where $\mathbf{P}_{t|T}$ can be found by iterating on

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t^{**}\mathbf{H}^{**}(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{H}^{**\prime}\mathbf{J}_t^{**\prime}$$

backward starting from $t = T-1$.

Procedure to calculate smoothed inferences $\{\hat{\xi}_{t|T}\}_{t=1}^{T}$.

(1) Perform Kalman filter recursion and save the values of $\{\hat{\xi}_{t|t}, \hat{\xi}_{t+1|t}, \mathbf{P}_{t|t}, \mathbf{P}_{t+1|t}\}_{t=1}^{T}$.

(2) Calculate

$$\mathbf{J}_t^{**} = \mathbf{P}_{t|t}(\mathbf{H}^{**}\mathbf{F})'(\mathbf{H}^{**}\mathbf{P}_{t+1|t}\mathbf{H}^{**\prime})^{-1}$$

for $t = 1, 2, \ldots, T-1$, where $\mathbf{H}^{**}$ is an $(s \times r)$ matrix selecting the nonredundant elements of ξ_t.

(3) Calculate

$$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} + \mathbf{J}_t^{**}\mathbf{H}^{**}(\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t})$$

for $t = T-1$ where $\hat{\xi}_{T-1|T-1}, \hat{\xi}_{T|T}$, and $\hat{\xi}_{T|T-1}$ are all known from step (1).

(4) Evaluate

$\hat{\xi}_{t|T} = \hat{\xi}_{t|t} + \mathbf{J}_t^{**}\mathbf{H}^{**}(\hat{\xi}_{t+1|T} - \hat{\xi}_{t+1|t})$

for $t = T - 2$ where right-hand variables are all known from step (3). Iterate for $t = T - 3, T - 4, \ldots$

Linear state-space models

A. State-space representation of a dynamic system
B. Kalman filter
C. Maximum likelihood estimation

Starting value for $t = 0$:

$$\xi_{t+1} = \mathbf{F}\xi_t + \mathbf{v}_{t+1}$$

If eigenvalues of $\mathbf{F}$ are all inside unit circle, set

$$\hat{\xi}_{0|0} = E(\xi_0) = \mathbf{0}$$

$$\mathbf{P}_{0|0} = E(\xi_0\xi_0')$$

$$\mathrm{vec}(\mathbf{P}_{0|0}) = [\mathbf{I}_{r^2} - (\mathbf{F} \otimes \mathbf{F})]^{-1}\,\mathrm{vec}(\mathbf{Q})$$

Alternatively, can use any distribution

$\xi_0 \sim N(\hat{\xi}_{0|0}, \mathbf{P}_{0|0})$

Frequentist perspective:

this is unconditional distribution of observation 0

Bayesian perspective:

this is prior beliefs about ξ_0

If diagonal elements of $\mathbf{P}_{0|0}$ are large (e.g. $\mathbf{P}_{0|0} = 10^4\mathbf{I}_r$), has little influence on any results

How to estimate unknown parameters?

Let θ be vector containing unknown elements of $\mathbf{F}, \mathbf{Q}, \mathbf{A}, \mathbf{H}, \mathbf{R}$

frequentist principle:

chose θ so as to max likelihood function

(1) pick an arbitrary initial guess for θ

(2) run through the KF for this fixed θ calculating

$\hat{\xi}_{t|t-1}(\theta)$

$\mathbf{P}_{t|t-1}(\theta)$

$\hat{\mathbf{y}}_{t|t-1}(\theta) = \mathbf{A}(\theta)'\mathbf{x}_t + \mathbf{H}(\theta)'\hat{\xi}_{t|t-1}(\theta)$

$\mathbf{C}_{t|t-1}(\theta) = \mathbf{H}(\theta)'\mathbf{P}_{t|t-1}(\theta)\mathbf{H}(\theta) + \mathbf{R}(\theta)$

(3) since this $\boldsymbol{\theta}$ implies

$\mathbf{y}_t|\Omega_{t-1},\mathbf{x}_t;\boldsymbol{\theta} \sim N(\hat{\mathbf{y}}_{t|t-1}(\boldsymbol{\theta}),\mathbf{C}_{t|t-1}(\boldsymbol{\theta}))$,

choose $\hat{\boldsymbol{\theta}}$ so as to maximize log likelihood:

$$\mathcal{L}(\boldsymbol{\theta}) = -\tfrac{Tn}{2}\log 2\pi - \tfrac{1}{2}\sum_{t=1}^{T}\log|\mathbf{C}_{t|t-1}(\boldsymbol{\theta})|$$
$$-\tfrac{1}{2}\sum_{t=1}^{T}[\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}(\boldsymbol{\theta})]'[\mathbf{C}_{t|t-1}(\boldsymbol{\theta})]^{-1}[\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}(\boldsymbol{\theta})]$$

by numerical search

(given guess $\boldsymbol{\theta}_j$, find a $\boldsymbol{\theta}_{j+1}$

associated with bigger value for $\mathcal{L}(\boldsymbol{\theta})$)

Asymptotic standard errors from

$\hat{\boldsymbol{\theta}} \approx N(\boldsymbol{\theta}_0,\widehat{\mathbf{C}})$

$$\widehat{\mathbf{C}} = \left[-\frac{\partial^2 \log p(\mathbf{Y}|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}\,\partial\boldsymbol{\theta}'}\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}\right]^{-1}$$

)

EM algorithm: convenient numerical algorithm for finding value of $\boldsymbol{\theta}$ that maximizes $\mathcal{L}(\boldsymbol{\theta})$.

Let $p(\mathbf{y};\boldsymbol{\theta})$ denote likelihood

(joint density of $\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_T$)

$$\log p(\mathbf{Y};\boldsymbol{\theta}) = \sum_{t=1}^{T}\log p(\mathbf{y}_t|\Omega_{t-1};\boldsymbol{\theta})$$

Consider $p(\mathbf{Y},\boldsymbol{\Xi};\boldsymbol{\theta})$ = joint density

of $\mathbf{y}_1,\mathbf{y}_2,\ldots,\mathbf{y}_T,\boldsymbol{\xi}_1,\boldsymbol{\xi}_2,\ldots,\boldsymbol{\xi}_T$

if $\boldsymbol{\xi}_t$ were observed.

$$\log p(\mathbf{Y},\Xi;\boldsymbol{\theta}) = -\tfrac{Tr}{2}\log(2\pi) - \tfrac{T}{2}\log|\mathbf{Q}|$$
$$-\tfrac{1}{2}\text{trace}\Big[\mathbf{Q}^{-1}\textstyle\sum_{t=1}^{T}(\boldsymbol{\xi}_t - \mathbf{F}\boldsymbol{\xi}_{t-1})(\boldsymbol{\xi}_t - \mathbf{F}\boldsymbol{\xi}_{t-1})'\Big]$$
$$-\tfrac{Tr}{2}\log(2\pi) - \tfrac{T}{2}\log|\mathbf{R}|$$
$$-\tfrac{1}{2}\text{trace}[\mathbf{R}^{-1}\textstyle\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\boldsymbol{\xi}_t)$$
$$\times(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\boldsymbol{\xi}_t)']$$

The EM algorithm is a sequence of values $\{\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots\}$ such that given $\boldsymbol{\theta}_\ell$, the value of $\boldsymbol{\theta}_{\ell+1}$ maximizes

$$\int \log p(\mathbf{Y},\Xi;\boldsymbol{\theta}_{\ell+1})p(\mathbf{Y},\Xi;\boldsymbol{\theta}_\ell)d\Xi$$

1) Why does EM algorithm work?

FOC will satisfy

$$\int \frac{\partial p(\mathbf{Y},\Xi;\boldsymbol{\theta}_{\ell+1})}{\partial\boldsymbol{\theta}_{\ell+1}} \frac{1}{p(\mathbf{Y},\Xi;\boldsymbol{\theta}_{\ell+1})} p(\mathbf{Y},\Xi;\boldsymbol{\theta}_\ell)d\Xi = \mathbf{0}$$

If we had a fixed point ($\boldsymbol{\theta}_{\ell+1} = \boldsymbol{\theta}_\ell$),

$$\int \frac{\partial p(\mathbf{Y},\boldsymbol{\Xi};\boldsymbol{\theta}_{\ell+1})}{\partial \boldsymbol{\theta}_{\ell+1}} \frac{1}{p(\mathbf{Y},\boldsymbol{\Xi};\boldsymbol{\theta}_{\ell+1})} p(\mathbf{Y},\boldsymbol{\Xi};\boldsymbol{\theta}_{\ell+1}) d\boldsymbol{\Xi}$$

$$= \int \frac{\partial p(\mathbf{Y},\boldsymbol{\Xi};\boldsymbol{\theta}_{\ell+1})}{\partial \boldsymbol{\theta}_{\ell+1}} d\boldsymbol{\Xi}$$

$$= \frac{\partial}{\partial \boldsymbol{\theta}_{\ell+1}} \int p(\mathbf{Y},\boldsymbol{\Xi};\boldsymbol{\theta}_{\ell+1}) d\boldsymbol{\Xi}$$

$$= \frac{\partial p(\mathbf{Y};\boldsymbol{\theta}_{\ell+1})}{\partial \boldsymbol{\theta}_{\ell+1}}$$

$$= \mathbf{0}$$

so a fixed point is the MLE

Furthermore, it can be shown that $\mathcal{L}(\boldsymbol{\theta}_{\ell+1}) > \mathcal{L}(\boldsymbol{\theta}_\ell)$

that is, each step of EM algorithm increases the log likelihood

(2) How do we implement EM algorithm?

Suppose that $\boldsymbol{\xi}_t$ was observed directly and we want to choose $\boldsymbol{\theta}$ to max

$$\log p(\mathbf{Y},\boldsymbol{\Xi};\boldsymbol{\theta}) = -\tfrac{Tr}{2}\log(2\pi) - \tfrac{T}{2}\log|\mathbf{Q}|$$
$$-\tfrac{1}{2}\text{trace}\left[\mathbf{Q}^{-1}\textstyle\sum_{t=1}^{T}(\boldsymbol{\xi}_t - \mathbf{F}\boldsymbol{\xi}_{t-1})(\boldsymbol{\xi}_t - \mathbf{F}\boldsymbol{\xi}_{t-1})'\right]$$
$$-\tfrac{Tr}{2}\log(2\pi) - \tfrac{T}{2}\log|\mathbf{R}|$$
$$-\tfrac{1}{2}\text{trace}[\mathbf{R}^{-1}\textstyle\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\boldsymbol{\xi}_t)$$
$$\times(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\boldsymbol{\xi}_t)']$$

Consider first parameters of $\mathbf{F}$.
If we observed ξ_t these would be found from OLS regression of ξ_t on ξ_{t-1}:
$\left(\sum_{t=1}^{T} \xi_{t-1}\xi_{t-1}'\right)\mathbf{F} = \left(\sum_{t=1}^{T} \xi_t \xi_{t-1}'\right)$.

$\left(\sum_{t=1}^{T} \xi_{t-1}\xi_{t-1}'\right)\mathbf{F} = \left(\sum_{t=1}^{T} \xi_t \xi_{t-1}'\right)$.
In step $\ell + 1$ of the EM algorithm we don't observe ξ_t so don't maximize $\log p(\mathbf{Y}, \Xi; \boldsymbol{\theta})$ but instead max the expectation of $\log p(\mathbf{Y}, \Xi; \boldsymbol{\theta})$.
In other words, we integrate out Ξ conditioning on data $\mathbf{Y}$ and assuming the previous iteration's value for $\boldsymbol{\theta}_\ell$.

$$E(\xi_{t-1}\xi_{t-1}' | \mathbf{Y}, \boldsymbol{\theta}_\ell)$$
$$= \mathbf{P}_{t-1|T}(\boldsymbol{\theta}_\ell) + \left[\hat{\xi}_{t-1|T}(\boldsymbol{\theta}_\ell)\right]\left[\hat{\xi}_{t-1|T}(\boldsymbol{\theta}_\ell)\right]'$$
$$\equiv \mathbf{S}_{t-1}(\boldsymbol{\theta}_\ell)$$
where $\hat{\xi}_{t-1|T}(\boldsymbol{\theta}_\ell)$ and $\mathbf{P}_{t-1|T}(\boldsymbol{\theta}_\ell)$ denote the estimates coming from the Kalman smoother evaluated at previous iteration's value for $\boldsymbol{\theta}_\ell$.

Similarly

$$E(\xi_t \xi_{t-1}' | \mathbf{Y}, \theta_\ell)$$
$$= \mathbf{P}_{t,t-1|T}(\theta_\ell) + \left[\hat{\xi}_{t|T}(\theta_\ell)\right]\left[\hat{\xi}_{t-1|T}(\theta_\ell)\right]'$$
$$\equiv \mathbf{S}_{t,t-1|T}(\theta_\ell)$$

for

$$\mathbf{P}_{t,t-1|T}(\theta_\ell) = \mathbf{J}_{t-1}(\theta_\ell)\mathbf{P}_{t|T}(\theta_\ell)$$
$$\mathbf{J}_{t-1}(\theta_\ell) = \mathbf{P}_{t-1|t-1}(\theta_\ell)\mathbf{F}_\ell'[\mathbf{P}_{t|t-1}(\theta_\ell)]^{-1}$$

FOC if ξ_t observed:

$$\left(\sum_{t=1}^T \xi_{t-1}\xi_{t-1}'\right)\mathbf{F} = \left(\sum_{t=1}^T \xi_t \xi_{t-1}'\right).$$

Value of $\mathbf{F}$ chosen by step $\ell + 1$ of EM algorithm:

$$\left[T^{-1}\sum_{t=1}^T \mathbf{S}_{t-1|T}(\theta_\ell)\right]\mathbf{F}_{\ell+1}$$
$$= \left[T^{-1}\sum_{t=1}^T \mathbf{S}_{t,t-1|T}(\theta_\ell)\right]$$
$$\mathbf{F}_{\ell+1} = \left[T^{-1}\sum_{t=1}^T \mathbf{S}_{t-1|T}(\theta_\ell)\right]^{-1}$$
$$\times \left[T^{-1}\sum_{t=1}^T \mathbf{S}_{t,t-1|T}(\theta_\ell)\right].$$

Likewise, if we were estimating $\mathbf{Q}$ with ξ_t observed, we would max

$$-\tfrac{Tr}{2}\log(2\pi) - \tfrac{T}{2}\log|\mathbf{Q}|$$
$$-\tfrac{1}{2}\text{trace}\left[\mathbf{Q}^{-1}\sum_{t=1}^T (\xi_t - \mathbf{F}\xi_{t-1})(\xi_t - \mathbf{F}\xi_{t-1})'\right]$$
$$\Rightarrow \mathbf{Q} = T^{-1}\sum_{t=1}^T (\xi_t - \hat{\mathbf{F}}\xi_{t-1})(\xi_t - \hat{\mathbf{F}}\xi_{t-1})'.$$

Step $\ell + 1$ of EM chooses

$$\mathbf{Q}_{\ell+1} = T^{-1}\sum_{t=1}^T \{\mathbf{S}_{t|T}(\theta_\ell) - \mathbf{F}_{\ell+1}\mathbf{S}_{t-1,t}(\theta_\ell)$$
$$- \mathbf{S}_{t-1,t}(\theta_\ell)'\mathbf{F}_{\ell+1}' + \mathbf{F}_{\ell+1}\mathbf{S}_{t-1|T}(\theta_\ell)\mathbf{F}_{\ell+1}'\}.$$

Analogously, with ξ_t observed we would choose $\mathbf{A}, \mathbf{H}$ to max

$$-\tfrac{Tr}{2}\log(2\pi) - \tfrac{T}{2}\log|\mathbf{R}|$$
$$-\tfrac{1}{2}\text{trace}[\mathbf{R}^{-1}\textstyle\sum_{t=1}^{T}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\xi_t)$$
$$\times (\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\xi_t)']$$

If ξ_t were observed we would do OLS regression of $\mathbf{y}_t$ on $\mathbf{z}_t$:

$$\mathbf{z}_t = \begin{bmatrix} \mathbf{x}_t \\ \xi_t \end{bmatrix}$$
$$\mathbf{\Pi} = \begin{bmatrix} \mathbf{A}' & \mathbf{H}' \end{bmatrix}$$
$$\left(\textstyle\sum_{t=1}^{T}\mathbf{z}_t\mathbf{z}_t'\right)\mathbf{\Pi} = \left(\textstyle\sum_{t=1}^{T}\mathbf{y}_t\mathbf{z}_t'\right).$$

Step $\ell + 1$ of EM algorithm thus uses

$$\begin{bmatrix} \mathbf{A}_{\ell+1}' & \mathbf{H}_{\ell+1}' \end{bmatrix}$$
$$= \begin{bmatrix} \sum_{t=1}^{T}\mathbf{y}_t\mathbf{x}_t' & \sum_{t=1}^{T}\mathbf{y}_t\hat{\xi}_{t|T}(\boldsymbol{\theta}_\ell)' \end{bmatrix}$$
$$\times \begin{bmatrix} \sum_{t=1}^{T}\mathbf{x}_t\mathbf{x}_t' & \sum_{t=1}^{T}\mathbf{x}_t\xi_{t|T}(\boldsymbol{\theta}_\ell)' \\ \sum_{t=1}^{T}\xi_{t|T}(\boldsymbol{\theta}_\ell)\mathbf{x}_t' & \sum_{t=1}^{T}\mathbf{S}_{t|T}(\boldsymbol{\theta}_\ell) \end{bmatrix}^{-1}$$

Linear state-space models

A. State-space representation of a dynamic system
B. Kalman filter
C. Maximum likelihood estimation
D. Applications
   1. Time-varying parameter models and missing observations

Suppose that $\mathbf{F}, \mathbf{Q}, \mathbf{A}, \mathbf{H}, \mathbf{R}$ are known functions of t (or more generally, known functions of $\mathbf{x}_t$):

$$\boldsymbol{\xi}_{t+1} = \mathbf{F}_t \boldsymbol{\xi}_t + \mathbf{v}_{t+1} \qquad E(\mathbf{v}_{t+1}\mathbf{v}_{t+1}') = \mathbf{Q}_t$$

$$\mathbf{y}_t = \mathbf{A}_t'\mathbf{x}_t + \mathbf{H}_t'\boldsymbol{\xi}_t + \mathbf{w}_t \qquad E(\mathbf{w}_t\mathbf{w}_t') = \mathbf{R}_t$$

Then Kalman filter recursion immediately generalizes to:

$$\mathbf{P}_{t+1|t} = \mathbf{F}_t \mathbf{P}_{t|t} \mathbf{F}_t' + \mathbf{Q}_t$$

$$\mathbf{P}_{t+1|t+1} = \mathbf{P}_{t+1|t} - \mathbf{P}_{t+1|t}\mathbf{H}_{t+1}(\mathbf{H}_{t+1}'\mathbf{P}_{t+1|t}\mathbf{H}_{t+1} + \mathbf{R}_{t+1})^{-1}\mathbf{H}_{t+1}'\mathbf{P}_{t+1|t}$$

$$\hat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}_t \hat{\boldsymbol{\xi}}_{t|t}$$

$$\hat{\boldsymbol{\varepsilon}}_{t+1|t} = \mathbf{y}_{t+1} - \mathbf{A}_{t+1}'\mathbf{x}_{t+1} - \mathbf{H}_{t+1}'\hat{\boldsymbol{\xi}}_{t+1|t}$$

$$\hat{\boldsymbol{\xi}}_{t+1|t+1} = \hat{\boldsymbol{\xi}}_{t+1|t} + \mathbf{P}_{t+1|t}\mathbf{H}_{t+1}(\mathbf{H}_{t+1}'\mathbf{P}_{t+1|t}\mathbf{H}_{t+1} + \mathbf{R}_{t+1})^{-1}\hat{\boldsymbol{\varepsilon}}_{t+1|t}$$

One simple trick for handling missing observations: if observation y_{it} is missing for date $t,$ set ith rows of $\mathbf{A}_t'$ and $\mathbf{H}_t'$ to zero, take $y_{it} = 0,$ set row $i,$ col i of $\mathbf{R}_t$ to 1 and all other elements of row i or col i of $\mathbf{R}_t$ to zero.

Why it works: suppose for illustration the first r elements of $\mathbf{y}_{t+1}$ are missing.

$$\mathbf{A}_{t+1}' = \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{A}}' \end{bmatrix} \qquad \mathbf{H}_{t+1}' = \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{H}}' \end{bmatrix}$$

$$\mathbf{R}_{t+1} = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{R}} \end{bmatrix}$$

Then

$$\mathbf{H}_{t+1} = \begin{bmatrix} \mathbf{0} & \tilde{\mathbf{H}} \end{bmatrix}$$

$$\mathbf{P}_{t+1|t}\mathbf{H}_{t+1} = \begin{bmatrix} \mathbf{0} & \mathbf{P}_{t+1|t}\tilde{\mathbf{H}} \end{bmatrix}$$

$$\mathbf{H}_{t+1}'\mathbf{P}_{t+1|t}\mathbf{H}_{t+1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{H}}'\mathbf{P}_{t+1|t}\tilde{\mathbf{H}} \end{bmatrix}$$

$$\mathbf{P}_{t+1|t}\mathbf{H}_{t+1}(\mathbf{H}_{t+1}'\mathbf{P}_{t+1|t}\mathbf{H}_{t+1}+\tilde{\mathbf{R}})^{-1}$$

$$=\begin{bmatrix}\mathbf{0} & \mathbf{P}_{t+1|t}\tilde{\mathbf{H}}\end{bmatrix}\times\begin{bmatrix}\mathbf{I}_r & \mathbf{0}\\ \mathbf{0} & \left(\tilde{\mathbf{H}}'\mathbf{P}_{t+1|t}\tilde{\mathbf{H}}+\tilde{\mathbf{R}}\right)^{-1}\end{bmatrix}$$

$$=\begin{bmatrix}\mathbf{0} & \mathbf{P}_{t+1|t}\tilde{\mathbf{H}}(\tilde{\mathbf{H}}'\mathbf{P}_{t+1|t}\tilde{\mathbf{H}}+\tilde{\mathbf{R}})^{-1}\end{bmatrix}$$

$$\mathbf{P}_{t+1|t}\mathbf{H}_{t+1}(\mathbf{H}_{t+1}'\mathbf{P}_{t+1|t}\mathbf{H}_{t+1}+\tilde{\mathbf{R}})^{-1}$$

$$=\begin{bmatrix}\mathbf{0} & \mathbf{P}_{t+1|t}\tilde{\mathbf{H}}(\tilde{\mathbf{H}}'\mathbf{P}_{t+1|t}\tilde{\mathbf{H}}+\tilde{\mathbf{R}})^{-1}\end{bmatrix}$$

$$\hat{\xi}_{t+1|t+1}=\hat{\xi}_{t+1|t}+$$
$$\mathbf{P}_{t+1|t}\mathbf{H}_{t+1}(\mathbf{H}_{t+1}'\mathbf{P}_{t+1|t}\mathbf{H}_{t+1}+\mathbf{R}_{t+1})^{-1}\hat{\boldsymbol{\varepsilon}}_{t+1|t}$$

acts as if first r elements of $\mathbf{y}_t$ weren't there

Linear state-space models

D. Applications

1. Time-varying parameter models and missing observations
2. Using mixed-frequency data as they arrive in real time

Practical problem for economic forecasters:

Different data are of different, asynchronous frequencies and are subsequently revised

Example: “Introducing the Euro-Sting: Short Term Indicator of Euro Area Growth”, Maximo Camacho and Gabriel Perez-Quiros

Assumption: there is an unobserved scalar f_t representing the monthly growth rate of real economic activity.

$\mathbf{z}_t^h = (4 \times 1)$ vector of “hard” indicators of f_t

$z_{1t}^h =$ industrial production growth

$z_{2t}^h =$ retail sales growth

$z_{3t}^h =$ new industrial orders growth

$z_{4t}^h =$ Euro area export growth

$$z_{it}^h = k_i^h + \beta_i^h f_t + u_{it}^h$$

$$
\begin{aligned}
z_{it}^{h} &= k_i^{h} + \beta_i^{h} f_t + u_{it}^{h} \\
f_t &= a_1 f_{t-1} + a_2 f_{t-2} + \cdots + a_6 f_{t-6} + \varepsilon_t^{f} \\
\varepsilon_t^{f} &\sim N(0,1) \\
u_{it}^{h} &= c_{i1}^{h} u_{i,t-1}^{h} + c_{i2}^{h} u_{i,t-2}^{h} + \cdots + c_{i,6}^{h} u_{i,t-6}^{h} + \varepsilon_{it}^{h} \\
\varepsilon_{it}^{h} &\sim N(0,\sigma_{hi}^{2}) \\
\mathbf{z}_t^{h} &= \mathbf{k}^{h} + \boldsymbol{\beta}^{h} f_t + \mathbf{u}_t^{h} \\
\mathbf{u}_t^{h} &= \mathbf{C}_1^{h}\mathbf{u}_{t-1}^{h} + \mathbf{C}_2^{h}\mathbf{u}_{t-2}^{h} + \cdots + \mathbf{C}_6^{h}\mathbf{u}_{t-6}^{h} + \boldsymbol{\varepsilon}_t^{h} \\
\boldsymbol{\xi}_t &= (f_t, f_{t-1}, \ldots, f_{t-5}, \mathbf{u}_t^{h}, \mathbf{u}_{t-1}^{h}, \ldots, \mathbf{u}_{t-5}^{h})'
\end{aligned}
$$

Also have some “soft” survey measures intended to reflect year-over-year growth

z_{1t}^{s} = Belgium overall business indicator

z_{2t}^{s} = Euro-zone economic sentiment

z_{3t}^{s} = German IFO business climate

z_{4t}^{s} = Euro manufacturing purchasing managers index

z_{5t}^{s} = services PMI

$$
\begin{aligned}
z_{it}^{s} &= k_i^{s} + \beta_i^{s} \sum_{j=0}^{11} f_{t-j} + u_{it}^{s} \\
u_{it}^{s} &= c_{i1}^{s} u_{i,t-1}^{s} + c_{i2}^{s} u_{i,t-2}^{s} + \cdots + c_{i,6}^{s} u_{i,t-6}^{s} + \varepsilon_{it}^{s}
\end{aligned}
$$

q_t = true monthly growth rate of real GDP in deviation from mean (not observed)

$$q_t = \tfrac{1}{3}\beta^q f_t + u_t^q$$

$$u_t^q = c_1^q u_{t-1}^q + c_2^q u_{t-2}^q + \cdots + c_6^q u_{t-6}^q + \varepsilon_t^q$$

Every three months we do observe a second revision of quarterly GDP growth

$$y_t^2 = k^2 + \tfrac{1}{3}q_t + \tfrac{2}{3}q_{t-1} + q_{t-2} + \tfrac{2}{3}q_{t-3} + \tfrac{1}{3}q_{t-4}$$

40 days earlier a more preliminary first revision was available

$$y_t^1 = y_t^2 + e_{2t}$$

20 days before that the initial “flash” estimate of GDP was released

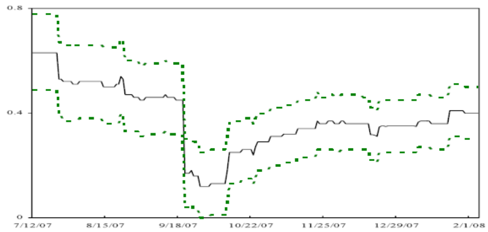
$$y_t^0 = y_t^1 + e_{1t}$$

Model also uses quarterly employment growth ℓ_t.
Potential observation vector:
$\mathbf{y}_t = (y_t^2, \mathbf{z}_t^{h\prime}, \mathbf{z}_t^{s\prime}, \ell_t, y_t^1, y_t^0)'$.

Potential observation vector:
$\mathbf{y}_t = (y_t^2, \mathbf{z}_t^{h\prime}, \mathbf{z}_t^{s\prime}, \ell_t, y_t^1, y_t^0)'$.
In every month, some of these (e.g., $y_t^2, \ell_t,$ and y_t^0) are treated as missing observations
On any given day before the end of the month, a smaller subset is observed.

$$\boldsymbol{\xi}_t = (f_t, f_{t-1}, \ldots, f_{t-11}, u_t^q, u_{t-1}^q, \ldots, u_{t-5}^q, \ldots \mathbf{u}_t^{h\prime}, \ldots, \mathbf{u}_{t-5}^{h\prime}, \mathbf{u}_t^{s\prime}, \ldots, \mathbf{u}_{t-5}^{s\prime}, u_t^\ell, \ldots, u_{t-5}^\ell)'$$

Model allows forecast of any variable using all information available as of any day



Real-time forecasts of 2007:Q4 real GDP growth from release of second revision on 2007/07/12 until 2008/02/13

Linear state-space models

D. Applications

1. Time-varying parameter models and missing observations
2. Using mixed-frequency data as they arrive in real time
3. Estimation of dynamic stochastic general equilibrium models

Basic approach:

(1) Find a state-space model that approximates solution to DSGE

(2) Estimate parameters of DSGE by maximizing implied likelihood

Example:

$$\max_{\{c_t,k_{t+1}\}_{t=0}^{\infty}} E_0 \sum_{t=0}^{\infty} \beta^t \log c_t$$

s.t. $c_t + k_{t+1} = e^{z_t} k_t^{\alpha} + (1-\delta)k_t \quad t = 1,2,\ldots$

$z_t = \rho z_{t-1} + \sigma\varepsilon_t \quad t = 1,2,\ldots$

k_0, z_0 given

$\varepsilon_t \sim N(0,1)$

If $\beta,\alpha,\rho \in (0,1)$, solution takes the form

$c_t = c(k_t, z_t; \sigma)$

$k_{t+1} = k(k_t, z_t; \sigma)$

Problem: The functions $c(.)$ and $k(.)$ cannot be found analytically.

(1) Take a fixed numerical value for $\boldsymbol{\theta} = (\beta,\alpha,\delta,\rho,\sigma)'$.

(2) Find values c, c_k, c_z, k, k_k, k_z as functions of $\boldsymbol{\theta}$ for which

$c(k_t, z_t; \sigma) \simeq c + c_k(k_t - k) + c_z z_t$

$k(k_t, z_t; \sigma) \simeq k + k_k(k_t - k) + k_z z_t$.

(3) If we think of observed variables $(\tilde{c}_t, \tilde{k}_t)$ as differing from model analogs (c_t, k_t) by measurement or specification error,

$\tilde{c}_t = c_t + \varepsilon_{ct}$

$\tilde{k}_t = k_t + \varepsilon_{kt}$,

then resulting system has state-space representation with state vector $\xi_t = (\hat{k}_t, z_t)'$ for $\hat{k}_t = k_t - k$ (deviations from mean).

state equation:

$\hat{k}_{t+1} = k_k \hat{k}_t + k_z z_t$

$z_{t+1} = \rho z_t + \sigma \varepsilon_{t+1}$

observation equation:

$\tilde{c}_t = c + c_k \hat{k}_t + c_z z_t + \varepsilon_{ct}$

$\tilde{k}_t = k + \hat{k}_t + \varepsilon_{kt}$

How do we achieve step (2)?
One approach: perturbation methods.
Consider a continuum of economies indexed by σ and use Taylor's Theorem to find approximation in neighborhood of $\sigma = 0$ (that is, as economy becomes deterministic).

First-order conditions:

$$\frac{1}{c_t} = \beta E_t\left[\frac{1-\delta+\alpha k_{t+1}^{\alpha-1}\exp(z_{t+1})}{c_{t+1}}\right]$$

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

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

(2a) Find steady-state (solutions for the case $\sigma = 0$) :

$$c_t = c_{t+1} = c$$

$$k_t = k_{t+1} = k$$

$$z_t = 0$$

$$\sigma = 0$$

$$\frac{1}{c} = \beta\left[\frac{1-\delta+\alpha k^{\alpha-1}}{c}\right]$$

$$c + k = k^{\alpha} + (1-\delta)k$$

In this case c and k can be found analytically:

$$1 = \beta[1-\delta+\alpha k^{\alpha-1}]$$

$$c = k^{\alpha} - \delta k.$$

More generally, could be found numerically.
(a) Make arbitrary initial guess ($\ell = 0$)
for (c_0, k_0).
(b) Calculate

$$\left\{\frac{1}{c_\ell} - \beta\left[\frac{1-\delta+\alpha k_\ell^{\alpha-1}}{c_\ell}\right]\right\}^2 +$$

$$\{c_\ell + k_\ell - [k_\ell^\alpha + (1-\delta)k_\ell]\}^2.$$

(c) Find better guess $(c_{\ell+1}, k_{\ell+1})$
until objective function acceptably small.

What if data are nonstationary?
One approach: let $\tilde{c}_t$ denote some measure of detrended consumption, and assume

$$\tilde{c}_t = c_t + \varepsilon_{ct}$$

for c_t the magnitude described by model.
Alternative approach: explicitly model trend in z_t, find transformation in model that induces a stationary magnitude c_t, and apply same transformation to data.

(2b) Use Taylor's Theorem to find approx linear coefficients.
(Take $\delta = 1$ case for illustration)
Write F.O.C. as $E_t \mathbf{a}(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$

$$a_1(k_t, z_t; \sigma, \varepsilon_{t+1}) = \frac{1}{c(k_t, z_t; \sigma)} - \beta\frac{\alpha k(k_t, z_t; \sigma)^{\alpha-1}\exp(\rho z_t + \sigma\varepsilon_{t+1})}{c(k(k_t, z_t; \sigma), \rho z_t + \sigma\varepsilon_{t+1}; \sigma)}$$

$$a_2(k_t, z_t; \sigma, \varepsilon_{t+1}) = c(k_t, z_t; \sigma) + k(k_t, z_t; \sigma) - e^{z_t}k_t^\alpha - (1-\delta)k_t$$

First-order approximation:
Since $E_t \mathbf{a}(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$ for all $k_t, z_t; \sigma$, it follows that
$E_t \mathbf{a}_k(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$
for $\mathbf{a}_k(k_t, z_t; \sigma, \varepsilon_{t+1}) = \frac{\partial \mathbf{a}(k_t, z_t; \sigma, \varepsilon_{t+1})}{\partial k_t}$
likewise
$E_t \mathbf{a}_z(k_t, z_t; \sigma, \varepsilon_{t+1}) = E_t \mathbf{a}_\sigma(k_t, z_t; \sigma, \varepsilon_{t+1}) = \mathbf{0}$

$$E_t \left\{ \frac{\partial a_1(k_t, z_t; \sigma, \varepsilon_{t+1})}{\partial k_t} \bigg|_{k_t = k, z_t = 0, \sigma = 0} \right\} =$$
$$\frac{-1}{c^2} c_k - \frac{\beta\alpha(\alpha-1)k^{\alpha-2}}{c} k_k + \frac{\beta\alpha k^{\alpha-1}}{c^2} c_k k_k$$

Since c and k are known from previous step, setting this to zero gives us an equation in the unknowns c_k and k_k where for example
$$c_k = \frac{\partial c(k_t, z_t; \sigma)}{\partial k_t} \bigg|_{k_t = k, z_t = 0, \sigma = 0}$$

$$\frac{\partial a_2(k_t, z_t; \sigma)}{\partial k_t} \bigg|_{k_t = k, z_t = 0, \sigma = 0} =$$
$$c_k + k_k - \alpha k^{\alpha-1} - (1-\delta)$$

This is a second equation in c_k, k_k, which together with the first can now be solved for c_k, k_k as a function of c and k

$$E_t\left\{\frac{\partial a_1(k_t,z_t;\sigma,\varepsilon_{t+1})}{\partial z_t}\Big|_{k_t=k,z_t=0,\sigma=0}\right\} =$$

$$\frac{-1}{c^2}c_z - \frac{\beta\alpha(\alpha-1)k^{\alpha-2}}{c}k_z - \frac{\beta\alpha k^{\alpha-1}\rho}{c}$$

$$+ \frac{\beta\alpha k^{\alpha-1}}{c^2}(c_k k_z + \rho c_z)$$

$$\frac{\partial a_2(k_t,z_t;\sigma)}{\partial z_t}\Big|_{k_t=k,z_t=0,\sigma=0} =$$

$$c_z + k_z - k^\alpha$$

setting these to zero allows us to solve for c_z, k_z

$$\frac{\partial a_1(k_t,z_t;\sigma,\varepsilon_{t+1})}{\partial \sigma}\Big|_{k_t=k,z_t=0,\sigma=0} =$$

$$\frac{-1}{c^2}c_\sigma - \frac{\beta\alpha(\alpha-1)k^{\alpha-2}}{c}k_\sigma - \frac{\beta\alpha k^{\alpha-1}\varepsilon_{t+1}}{c}$$

$$+ \frac{\beta\alpha k^{\alpha-1}}{c^2}(c_k k_\sigma + \varepsilon_{t+1}c_z + c_\sigma)$$

$$\frac{\partial a_2(k_t,z_t;\sigma)}{\partial \sigma}\Big|_{k_t=k,z_t=0,\sigma=0} =$$

$$c_\sigma + k_\sigma$$

Taking expectations and setting to zero yields

$$\frac{-1}{c^2}c_\sigma - \frac{\beta\alpha(\alpha-1)k^{\alpha-2}}{c}k_\sigma$$

$$+ \frac{\beta\alpha k^{\alpha-1}}{c^2}(c_k k_\sigma + c_\sigma) = 0$$

$$c_\sigma + k_\sigma = 0$$

which has solution $c_\sigma = k_\sigma = 0$

$\Rightarrow$ volatility, risk aversion play no role in first-order approximation

In this example, values for c_k, c_z, k_k, k_z could be found analytically.
More generally, we will have a quadratic system of equations in unknowns like c_k, c_z, k_k, k_z.
Common approach to solving: recognize as linear rational-expectations model and use numerical methods to find stable solution (assuming exists and is unique).

In general, better solutions obtained (particularly for trending data) if linearize in logs rather than levels.

$$\widehat{c}_t = \log c_t, \ \hat{k}_t = \log k_t$$

$$\frac{1}{\exp(\widehat{c}_t)} = \beta E_t\left[\frac{1-\delta+\alpha\exp[(\alpha-1)\hat{k}_{t+1}]\exp(z_{t+1})}{\exp(\widehat{c}_{t+1})}\right]$$

$$\exp(\widehat{c}_t) + \exp(\hat{k}_{t+1}) = e^{z_t}\exp(\alpha\hat{k}_t) + (1-\delta)\exp(\hat{k}_t)$$

$$\widehat{c}_t = \widehat{c}(\hat{k}_t, z_t; \sigma)$$

$$\hat{k}_{t+1} = \hat{k}(\hat{k}_t, z_t; \sigma)$$

(3) Once we have state-space representation for observed data $\mathbf{y}_t = (\tilde{c}_t, \tilde{k}_t)'$ associated with this fixed $\boldsymbol{\theta}$, we can choose $\boldsymbol{\theta}$ to maximize likelihood.

(3) since this θ implies

$\mathbf{y}_t|\Omega_{t-1},\mathbf{x}_t;\theta \sim N(\hat{\mathbf{y}}_{t|t-1}(\theta),\mathbf{C}_{t|t-1}(\theta)),$

choose $\hat{\theta}$ so as to maximize log likelihood:

$$\mathcal{L}(\theta) = -\frac{Tn}{2}\log 2\pi - \frac{1}{2}\sum_{t=1}^{T}\log|\mathbf{C}_{t|t-1}(\theta)|$$
$$-\frac{1}{2}\sum_{t=1}^{T}[\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}(\theta)]'[\mathbf{C}_{t|t-1}(\theta)]^{-1}[\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}(\theta)]$$

by numerical search

(given guess θ_j, find a θ_{j+1}

associated with bigger value for $\mathcal{L}(\theta)$)

Caution: the DSGE is often simultaneously (1) overidentified, (2) underidentified, and (3) weakly identified.

(1) Overidentification: The DSGE implies a state-space model

$$\xi_{t+1} = \mathbf{F}(\theta)\xi_t + \mathbf{v}_{t+1}$$
$$\mathbf{y}_t = \mathbf{a}(\theta) + \mathbf{H}(\theta)'\xi_t + \mathbf{w}_t$$

where $\mathbf{F},\mathbf{a},\mathbf{H}$ satisfy complicated nonlinear restrictions.

A specification with less restrictive $\mathbf{F},\mathbf{a},\mathbf{H}$ may fit data much better.

Empirical applications typically have much richer dynamics than simple theoretical models. Example: Smets and Wouters, Journal of European Economic Association, 2003.

instantaneous utility function:

$$U_t = a_t^b\Big[\tfrac{1}{1-\lambda_c}(C_t - hC_{t-1})^{1-\lambda_c} - \tfrac{a_t^L}{1+\lambda_\ell}(\ell_t)^{1+\lambda_\ell}\Big]$$

$h > 0$ ⇒ habit persistence

$a_t^b = \rho_b a_{t-1}^b + \eta_t^b$ η_t^b ~ i.i.d. $N(0,\sigma_b^2)$

⇒ shock to intertemporal subs

$a_t^L = \rho_L a_{t-1}^L + \eta_t^L$

⇒ shock to intratemporal subs

Let $\widehat{C}_t$ denote deviation of $\log(C_t)$ from its steady-state value

(1) $$\widehat{C}_t = \left(\tfrac{h}{1+h}\right)\widehat{C}_{t-1} + \left(\tfrac{1}{1+h}\right)E_t\widehat{C}_{t+1} - \tfrac{(1-h)}{(1+h)\lambda_c}\left(\hat{R}_t - E_t\hat{\pi}_{t+1}\right) + \tfrac{(1-h)}{(1+h)\lambda_c}(\hat{a}_t^b - E_t\hat{a}_{t+1}^b)$$

capital evolution:

$K_t = K_{t-1}(1-\delta) + [1 - S(a_t^I I_t / I_{t-1})] I_t$

$S(.)$ = adjustment costs

$a_t^I = \rho_I a_{t-1}^I + \eta_t^I$

(2) $\hat{K}_t = (1-\delta)\hat{K}_{t-1} + \delta \hat{I}_{t-1}$

(3) $\hat{I}_t = \left(\frac{1}{1+\beta}\right)\hat{I}_{t-1} + \left(\frac{\beta}{1+\beta}\right)E_t\hat{I}_{t+1} + \frac{\varphi}{1+\beta}\hat{Q}_t - \frac{\beta E_t \hat{a}_{t+1}^I - \hat{a}_t^I}{1+\beta}$

$\varphi = 1/S''$

Q_t = value of capital stock

(4) $\hat{Q}_t = -\left(\hat{R}_t - \hat{\pi}_{t+1}\right) + \frac{1-\delta}{1-\delta+\overline{r}_k} E_t \hat{Q}_{t+1} + \frac{\overline{r}_k}{1-\delta+\overline{r}_k} E_t \hat{r}_{K,t+1} + \eta_t^Q$

r_{Kt} = rate of return to capital

η_t^Q = tacked on

output from producer of intermediate good of type j

$y_t^j = a_t^a K_{t-1}(j)^\alpha L_t(j)^{1-\alpha} - \Phi$

$\Phi =$ fixed cost

$a_t^a =$ productivity shock

$a_t^a = \rho_a a_{t-1}^a + \eta_t^a$

$L_t(j) =$ aggregate of labor hired from each household τ

$L_t(j) = \left\{\int_0^1 [\ell_t(\tau)]^{1/(1+\lambda_{w,t})} d\tau\right\}^{1+\lambda_{w,t}}$

$\lambda_{w,t} = \lambda_w + \eta_t^w$

$\eta_t^w =$ shock to workers' market power

wage stickiness:

a fraction ξ_w of workers are not allowed to change their wage but instead have their wage increase from the previous value by $(P_{t-1}/P_{t-2})^{\gamma_w}$

$\gamma_w =$ degree of indexing

(5) $\widehat{w}_t = \frac{\beta}{1+\beta} E\widehat{w}_{t+1} + \frac{1}{1+\beta}\widehat{w}_{t-1}$

$+\frac{\beta}{1+\beta} E_t\hat{\pi}_{t+1} - \frac{1+\beta\gamma_w}{1+\beta}\hat{\pi}_t + \frac{\gamma_w}{1+\beta}\hat{\pi}_{t-1}$

$-h_w\Big[\widehat{w}_t - \lambda_L \hat{L}_t - \frac{\lambda_c}{1-h}(\widehat{C}_t - h\widehat{C}_{t-1})$

$-\hat{a}_t^L - \hat{\eta}_t^w]$

$h_w = \frac{1}{1+\beta}\frac{(1-\beta\xi_w)(1-\xi_w)}{[1+(1+\lambda_w)\lambda_L/\lambda_w]\xi_w}$

labor demand

$\frac{W_t L_t(j)}{r_{K,t} z_t K_{t-1}(j)} = \frac{1-\alpha}{\alpha}$

z_t = capital utilization

(6) $\hat{L}_t = -\widehat{w}_t + (1+\psi)\hat{r}_{K,t} + \hat{K}_{t-1}$

ψ = parameter based on cost of utilizing capital

intermediate goods sold to final goods producer with market power of firm j governed by

$\lambda_{p,t} = \lambda_p + \eta_t^p$

ξ_p = fraction allowed to adjust prices

γ_p = indexing parameter

(7) $\hat{\pi}_t = \frac{\beta}{1+\beta\gamma_p} E_t \hat{\pi}_{t+1} + \frac{\gamma_p}{1+\beta\gamma_p} \hat{\pi}_{t-1}$
$+h_p\left[\alpha \hat{r}_{K,t} + (1-\alpha)\widehat{w}_t - \hat{a}_t^a + \eta_t^p\right]$
$h_p = \frac{1}{1+\beta\gamma_p} \frac{(1-\beta\xi_p)(1-\xi_p)}{\xi_p}$

goods market equilibrium
(8) $\widehat{Y}_t = [1 - \delta(\overline{K}/\overline{Y}) - (\overline{G}/\overline{Y})]\widehat{C}_t$
$+\delta(\overline{K}/\overline{Y})\hat{I}_t + (\overline{G}/\overline{Y})\hat{a}_t^G$
production function then
determines $r_{K,t}$
(9) $\widehat{Y}_t = \phi \hat{a}_t^a + \phi\alpha\hat{K}_{t-1} + \phi\alpha\psi\hat{r}_{K,t}$
$+\phi(1-\alpha)\hat{L}_t$
$\phi = 1 + \frac{\Phi}{\text{s.s. costs}}$

monetary policy (Taylor Rule)
(10) $\hat{R}_t = \rho\hat{R}_{t-1} + (1-\rho)\{\overline{\pi}_t +$
$r_\pi(\hat{\pi}_{t-1} - \overline{\pi}_{t-1}) + r_Y(\widehat{Y}_t - \widehat{Y}_t^P)\}$
$+r_{\Delta\pi}(\hat{\pi}_t - \hat{\pi}_{t-1})$
$+\ r_{\Delta Y}\left[\widehat{Y}_t - \widehat{Y}_t^p - (\widehat{Y}_{t-1} - \widehat{Y}_{t-1}^p)\right] + \eta_t^R$
$\overline{\pi}_t =$ inflation target
$\overline{\pi}_t = \rho_\pi \overline{\pi}_{t-1} + \eta_t^\pi$
$\widehat{Y}_t^p =$ output level if prices perfectly flexible

$\mathbf{y}_t = (\widehat{C}_t, \widehat{C}_{t-1}, \hat{R}_t, \hat{R}_{t-1}, \hat{K}_t, \hat{K}_{t-1}, \hat{I}_t, \hat{I}_{t-1},$
$\quad \hat{Q}_t, \widehat{w}_t, \widehat{w}_{t-1}, \hat{L}_t, \hat{\pi}_t, \hat{\pi}_{t-1}, \widehat{Y}_t, \hat{r}_{K,t})'$

$\mathbf{x}_t = (\hat{a}_t^b, \hat{a}_t^I, \eta_t^Q, \hat{a}_t^L, \eta_t^w, \hat{a}_t^a, \eta_t^p, \hat{a}_t^G, \overline{\pi}_t, \eta_t^R)'$

equations (1)-(10) (along with lag definitions) can be written as

$\mathbf{A}E_t\mathbf{y}_{t+1} = \mathbf{B}\mathbf{y}_t + \mathbf{C}\mathbf{x}_t$

while shocks satisfy

$\mathbf{x}_{t+1} = \mathbf{\Phi}\mathbf{x}_t + \boldsymbol{\varepsilon}_{t+1}$

(note also $E_t\mathbf{x}_{t+1} = \mathbf{\Phi}\mathbf{x}_t$)

(2) At the same time that DSGE implies refutable overidentifying restrictions, $\boldsymbol{\theta}$ itself may be unidentified (Komunjer and Ng, Econometrica, 2011).

With Gaussian errors, the observable implications of the state-space representation are entirely summarized by $\mathbf{y} \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega})$

for $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$

$\quad \boldsymbol{\mu} = \mathbf{a}(\boldsymbol{\theta}) \otimes \mathbf{1}_T$

$\quad \boldsymbol{\Omega}$ a known function of $\boldsymbol{\theta}$

e.g., diagonal blocks of $\boldsymbol{\Omega}$ given by

$\mathbf{H}(\boldsymbol{\theta})'\boldsymbol{\Sigma}(\boldsymbol{\theta})\mathbf{H}(\boldsymbol{\theta}) + \mathbf{R}(\boldsymbol{\theta})$

$\mathrm{vec}(\boldsymbol{\Sigma}(\boldsymbol{\theta}) =$

$\quad [\mathbf{I}_{r^2} - (\mathbf{F}(\boldsymbol{\theta}) \otimes \mathbf{F}(\boldsymbol{\theta}))]^{-1} \mathrm{vec}(\mathbf{Q}(\boldsymbol{\theta}))$

Model is unidentified at θ_0 if there exists a $\theta_1 \neq \theta_0$ such that

$$\mathbf{a}(\theta_0) = \mathbf{a}(\theta_1)$$

$$\Omega(\theta_0) = \Omega(\theta_1)$$

Can check this locally by numerically calculating derivatives with respect to θ

(3) Finally, other parameters of θ may only be weakly identified

$$\Omega(\theta_0) \neq \Omega(\theta_1)$$

but $\mathcal{L}(\theta_0) - \mathcal{L}(\theta_1)$ small for $\|\theta_0 - \theta_1\|$ large.

Common approach: fix some parameters such as β using a priori information.

Bayesian estimation with informative priors can help with some of these numerical problems.