Intensive Courses in the context of the Jean Monnet Chair:

Big data in official statistics

Block 2: Structural time series models

14 DECEMBER 2018,

UNIVERSITY OF PISA

Jan van den Brakel

Introduction

Time series models:

1. Box & Jenkens ARIMA models

2. Structural time series models

Ad. 1:

The approach followed by Box and Jenkins (1989) for modelling time series starts by making an observed series stationary. Informally spoken, this implies that the trend in an observed series is removed by taking differences between subsequent periods. Seasonal patterns are removed in a similar way by taking differences between the observations of the same quarters or months of two successive years. Once an observed series is made stationary, it is modelled with autoregressive and moving average components.

Ad. 2:

Structural time series modelling follows a more direct and intuitive approach (subjective opinion). An observed series is directly modelled without attempting to remove non-stationarity through differencing of the observed series. This is the approach followed by authors like Harvey (1989), and Durbin and Koopman (2012).

Structural Time Series Models

Observed series $y_t, \quad t = 1, \ldots, T.$

Structural Time Series (STS) models decompose an observed series in:

1. Trend (L_t)
2. Seasonal (S_t)
3. Cycles (γ_t)
4. Regression component ($\boldsymbol{\beta}_t'\mathbf{x}_t$)
5. White noise (I_t)

Additive model:

$$y_t = L_t + S_t + \gamma_t + \boldsymbol{\beta}_t'\mathbf{x}_t + I_t, \quad t = 1, \ldots, T.$$

Multiplicative model:

$$y_t = L_t \times S_t \times \gamma_t \times \boldsymbol{\beta}_t'\mathbf{x}_t \times I_t, \quad t = 1, \ldots, T.$$

(Additive again after taking logs)

Ad. 1: The trend models the low frequency variation in the observed series. As will follows, it is modelled with a dynamic model that has the flexibility to adapt to gradually changes over time. Depending on its flexibility it can both capture trend and business cycles. If no separate component for the cycle is included, the trend will be estimated more flexible to capture also cyclic movements.

Ad. 2: The seasonal component describes the periodic fluctuation within a period of one year.

Ad. 3: Besides seasonal fluctuations there are also other periodic fluctuations, e.g. business cycles, with periods longer than a year, than can be modelled with separate components. In that case it is separated from the trend. We do not use separate cycle components in the models in this course.

Ad. 4: Related auxiliary series can be included in the model as a regression component. The regression coefficients can be made time-dependent (by modelling them as a random walk).

Ad. 5: The unexplained variation in the series is modelled as white noise.

Local Level Model

Very simple trend model: L_t is a random walk:

$$
\begin{aligned}
y_t &= L_t + I_t & I_t &\simeq \mathcal{N}(0, \sigma_I^2) \\
L_t &= L_{t-1} + \zeta_t & \zeta_t &\simeq \mathcal{N}(0, \sigma_\zeta^2)
\end{aligned}
$$

Note:

$$y_t = L_0 + \sum_{t=1}^{t} \zeta_t + I_t$$

- Serial correlation between observations y_t. This makes routine computations from normal regression theory inefficient
- Filtering and smoothing algorithms developed as an alternative
- Express STS model as a state space model
- Local level model is already in state space representation
- Kalman filter to obtain optimal estimates for L_t

Local Linear Trend Model

Popular trend model for economic time series:

$$
\begin{aligned}
y_t &= L_t + I_t & I_t &\simeq \mathcal{N}(0, \sigma_I^2) \\
L_t &= L_{t-1} + R_{t-1} + \zeta_t & \zeta_t &\simeq \mathcal{N}(0, \sigma_\zeta^2) \\
R_t &= R_{t-1} + \tau_t & \tau_t &\simeq \mathcal{N}(0, \sigma_\tau^2)
\end{aligned}
$$

- L_t often referred to as the level
- R_t interpreted as a slope parameter
- Trend models with random levels are often volatile

Exercise:

- What happens if for the local level model $\sigma_\zeta^2 = 0$? (Illustrate with a graph.)
- What happens if for the local linear model $\sigma_\zeta^2 = 0$ and $\sigma_\tau^2 = 0$? (Illustrate with a graph.)

Smooth Trend Model

Special case of the local linear trend model:

$$y_t = L_t + I_t \qquad\qquad I_t \simeq \mathcal{N}(0, \sigma_I^2)$$

$$L_t = L_{t-1} + R_{t-1}$$

$$R_t = R_{t-1} + \tau_t \qquad\qquad \eta_t \simeq \mathcal{N}(0, \sigma_\tau^2)$$

- Only the slope is random
- Results in more stable trend patterns

State Space Representation

State space representation STS model:

1. Measurement equation: $y_t = \mathbf{Z}\boldsymbol{\alpha}_t + I_t$
   - $\boldsymbol{\alpha}_t$: vector with state variables (trend, seasonal, etc)
   - $\mathbf{Z}$: Design matrix measurement equation
   - $I_t \simeq \mathcal{N}(0, \sigma_I^2)$
2. Transition equation: $\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t$
   - $\mathbf{T}$: Design matrix transition equation
   - $\boldsymbol{\eta}_t$: vector disturbances of the state variables (trend, seasonal, etc)
   - $\boldsymbol{\eta}_t \simeq \mathcal{N}(\mathbf{0}, \mathbf{H})$

Ad. 1: The measurement equation describes how the observed series depends on unobserved state variables that describe trend, seasonal components, regression components, etc.

Ad.2: The transition equation describes how the state variables evolve over time. More precisely; how they change from one period to the next.

Exercise

Give the state space representation for the local linear trend model:

$$
\begin{aligned}
y_t &= L_t + I_t & I_t &\simeq \mathcal{N}(0, \sigma_I^2) \\
L_t &= L_{t-1} + R_{t-1} + \zeta_t & \zeta_t &\simeq \mathcal{N}(0, \sigma_\zeta^2) \\
R_t &= R_{t-1} + \tau_t & \tau_t &\simeq \mathcal{N}(0, \sigma_\tau^2)
\end{aligned}
$$

State space representation local linear trend model

- Measurement equation: $y_t = \mathbf{Z}\boldsymbol{\alpha}_t + I_t$

$$\mathbf{Z} = (1 \;\; 0)$$

$$\boldsymbol{\alpha}_t = (L_t \;\; R_t)'$$

$$\Rightarrow y_t = (1 \;\; 0)\begin{pmatrix} L_t \\ R_t \end{pmatrix} + I_t$$

$$I_t \simeq \mathcal{N}(0, \sigma_I^2)$$

- Transition equation: $\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t$

$$\mathbf{T} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{\eta}_t = (\zeta_t \;\; \tau_t)'$$

$$\Rightarrow \begin{pmatrix} L_t \\ R_t \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} L_{t-1} \\ R_{t-1} \end{pmatrix} + \begin{pmatrix} \zeta_t \\ \tau_t \end{pmatrix}$$

$$\boldsymbol{\eta}_t \simeq \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_\zeta^2 & 0 \\ 0 & \sigma_\tau^2 \end{pmatrix}\right)$$

Kalman filter

- Structural time series models in state space form
- Kalman filter to obtain optimal estimates for state variables (and signal)
- Recursive algorithm that gives optimal estimates for $\boldsymbol{\alpha}_t$ based on the information available at time t
- Assumes that covariance matrices of the measurement and system equation are known, i.e. σ_I^2 and $\mathbf{H}$
- $\Rightarrow$ Kalman filter gives Best Linear Unbiased Predictions (BLUP) for state variables
- Let $\mathbf{a}_t$ denote the BLUP for $\boldsymbol{\alpha}_t$ based on information available at time t (i.e. the filtered estimate)
- Let $\mathbf{P}_t$ denote the covariance matrix of the estimation errors of $\mathbf{a}_t$
- Assume that the values for $\mathbf{a}_0$ and $\mathbf{P}_0$ are known

Kalman filter recursion

- Prediction equations:

$$\mathbf{a}_{t|t-1} = \mathbf{T}\mathbf{a}_{t-1}$$

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

The prediction equations follow directly from the transition equation.

- From the measurement equation it follows:

$$\hat{y}_{t|t-1} = \mathbf{Z}\mathbf{a}_{t|t-1}$$

- Innovation (new information if y_t becomes available):

$$\nu_t = y_t - \hat{y}_{t|t-1} = \mathbf{Z}(\boldsymbol{\alpha}_t - \mathbf{a}_{t|t-1}) + I_t$$

- Variance innovations

$$f_t = \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}' + \sigma_I^2$$

The variance of the innovations follows directly from the measurement equation.

- Updating equations (BLUP for $\boldsymbol{\alpha}_t$):

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + \frac{\nu_t}{f_t}\mathbf{P}_{t|t-1}\mathbf{Z}'$$
$$\mathbf{P}_t = \mathbf{P}_{t|t-1} - \frac{1}{f_t}\mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{Z}\mathbf{P}_{t|t-1}$$

The updating equations follow from the assumption that $\boldsymbol{\alpha}_0$, I_t, and $\boldsymbol{\eta}_t$ are multivariate normally distributed and subsequently the conditional distribution of $\boldsymbol{\alpha}_t$ given y_t. For a proof see Harvey (1989), Ch. 3.

- To start the filter recursion it is assumed that $\sigma_I^2, \mathbf{H}, \mathbf{a}_0$, and $\mathbf{P}_0$ are known

- In practise:

  - σ_I^2 and $\mathbf{H}$ are replaced by their ML estimates

  - Diffuse initialization of the Kalman filter, i.e.

$$\mathbf{a}_0 = \mathbf{0}$$

$$\mathbf{P}_0 = \kappa\mathbf{I}$$

  with e.g. $\kappa = 10^7$

Smoothing

- Kalman filter recursion runs from $t = 1, ..., T$ and gives BLUP's for $\boldsymbol{\alpha}_t$ given information obtained until period t
- Smoothing improves $\mathbf{a}_t$ using information obtained after period t
- Widely applied smoothing algorithm: fixed interval smoother
- Recursive algorithm that starts with the final quantities $\mathbf{a}_T$ and $\mathbf{P}_T$ and runs back from $t = T - 1, ..., 1$
- Smoothed BLUP's of $\boldsymbol{\alpha}_t$:

$$\mathbf{a}_{t|T} = \mathbf{a}_t + \mathbf{P}_t \mathbf{T}' \mathbf{P}_{t+1|t}^{-1} (\mathbf{a}_{t+1|T} - \mathbf{T}\mathbf{a}_t)$$

- Covariance matrix of prediction errors of $\mathbf{a}_{t|T}$:

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

Seasonal components

- Model a cycle with a period of one year
- Models:
  - Dummy seasonal model
  - Trigonometric seasonal models

Dummy seasonal model

- Seasonal pattern constant in time:

$$\sum_{j=1}^{J} S_j^* = 0$$

- Monthly data: $J = 12$

- Seasonal effect for period t corresponding to month j:

  $S_t = S_j^*$

- Time dependent seasonal pattern:

$$\sum_{j=0}^{J-1} S_{t-j} = \omega_t$$

$$\omega_t \simeq \mathcal{N}(0, \sigma_\omega^2)$$

$$\text{so : } \; S_t = -\sum_{j=1}^{J-1} S_{t-j} + \omega_t$$

- Basic Structural Time series Model (BSM):

$$y_t = L_t + S_t + I_t$$

  (Series = trend + dummy seasonal + white noise)

Dummy seasonal model - state space form

- Measurement equation: $y_t = \mathbf{Z}\boldsymbol{\alpha}_t + I_t$

$$\mathbf{Z} = (1\ 0\ \ 1\ \mathbf{0}'_{[10]}) \equiv (\mathbf{Z}^{[L]}\ \ \mathbf{Z}^{[S]})$$

$$\boldsymbol{\alpha}_t = (L_t\ R_t\ \ S_t\ S_{t-1}\ S_{t-2} \ldots\ S_{t-10})' \equiv \begin{pmatrix} \boldsymbol{\alpha}_t^{[L]} \\ \boldsymbol{\alpha}_t^{[S]} \end{pmatrix}$$

$$I_t \simeq \mathcal{N}(0, \sigma_I^2)$$

- Transition equation: $\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \Leftrightarrow$

$$\begin{pmatrix} L_t \\ R_t \\ S_t \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \\ \vdots \\ S_{t-9} \\ S_{t-10} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & -1 & -1 & -1 & \ldots & -1 & -1 \\ 0 & 0 & 1 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & \ldots & 0 & 0 \\ \vdots & & & & & \ldots & & \vdots \\ 0 & 0 & 0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \ldots & 1 & 0 \end{pmatrix} \begin{pmatrix} L_{t-1} \\ R_{t-1} \\ S_{t-1} \\ S_{t-2} \\ S_{t-3} \\ S_{t-4} \\ \vdots \\ S_{t-10} \\ S_{t-11} \end{pmatrix} + \begin{pmatrix} 0 \\ \eta_t \\ \omega_t \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix}$$

$$\Leftrightarrow \begin{pmatrix} \boldsymbol{\alpha}_t^{[L]} \\ \boldsymbol{\alpha}_t^{[S]} \end{pmatrix} = \begin{pmatrix} \mathbf{T}^{[L]} & \mathbf{O}_{[2\times 11]} \\ \mathbf{O}_{[11\times 2]} & \mathbf{T}^{[S]} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_{t-1}^{[L]} \\ \boldsymbol{\alpha}_{t-1}^{[S]} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_t^{[L]} \\ \boldsymbol{\eta}_t^{[S]} \end{pmatrix}$$

$$\mathbf{T}^{[S]} = \begin{pmatrix} -\mathbf{j}'_{[10]} & -1 \\ \mathbf{I}_{[10\times 10]} & \mathbf{0}_{[10]} \end{pmatrix}$$

$$\boldsymbol{\eta}_t^{[S]} = (\omega_t\ \ \mathbf{0}'_{[10]})'$$

$$\boldsymbol{\eta}_t \simeq \mathcal{N}\left(\mathbf{0}_{[13]}, \mathbf{Diag}(0\ \sigma_\eta^2\ \ \sigma_\omega^2\ \mathbf{0}'_{[10]})\right)$$

Notation:

- $\mathbf{O}_{[p\times q]}$: $p \times q$ matrix with elements equal to 0
- $\mathbf{I}_{[p\times p]}$: $p \times p$ identity matrix
- $\mathbf{0}_{[p]}$: p column vector with elements equal to 0
- $\mathbf{j}_{[p]}$: p column vector with elements equal to 1

Trigonometric seasonal model

- More flexibility but also more complicated
- Describes yearly cycle with a set of harmonics:

$$
\begin{aligned}
S_t &= \sum_{j=1}^{J/2} \gamma_{j,t} \\
\gamma_{j,t} &= \gamma_{j,t-1} \cos\left(\frac{\pi\, j}{J/2}\right) + \gamma^*_{j,t-1} \sin\left(\frac{\pi\, j}{J/2}\right) + \omega_{j,t} \\
\gamma^*_{j,t} &= \gamma^*_{j,t-1} \cos\left(\frac{\pi\, j}{J/2}\right) - \gamma_{j,t-1} \sin\left(\frac{\pi\, j}{J/2}\right) + \omega^*_{j,t} \\
\omega_{j,t} &\simeq \mathcal{N}(0, \sigma^2_\omega) \quad \omega^*_{j,t} \simeq \mathcal{N}(0, \sigma^2_\omega) \\
& \qquad j = 1, \ldots, J/2
\end{aligned}
$$

- Last harmonic:
  - $\gamma^*_{6,t}$ is not required for the measurement equation and does not influence other $\gamma' s$
  - $\gamma_{6,t} = -\gamma_{6,t-1}$ since $\cos(\pi) = -1$ and $\sin(\pi) = 0$
  - Therefore $\mathbf{C}_6 = -1$ in the design matrix of the transition equation $\mathbf{T}^{[S]}$ at page 8
- More general: $\omega_{j,t} \simeq \mathcal{N}(0, \sigma^2_{\omega,j})$ and $\omega^*_{j,t} \simeq \mathcal{N}(0, \sigma^2_{\omega,j})$

Trigonometric seasonal model - state space form

- Measurement equation: $y_t = \mathbf{Z}\boldsymbol{\alpha}_t + I_t$

$$\mathbf{Z} = (\mathbf{Z}^{[L]}\ \mathbf{Z}^{[S]}), \qquad \mathbf{Z}^{[S]} = (1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1\ 0\ 1)$$

$$\boldsymbol{\alpha}_t = \begin{pmatrix} \boldsymbol{\alpha}_t^{[L]} \\ \boldsymbol{\alpha}_t^{[S]} \end{pmatrix}, \qquad \boldsymbol{\alpha}_t^{[S]} = \left(\gamma_{1t}\ \gamma_{1t}^*\ \gamma_{2t}\ \gamma_{2t}^*\ \gamma_{3t}\ \gamma_{3t}^*\ \gamma_{4t}\ \gamma_{4t}^*\ \gamma_{5t}\ \gamma_{5t}^*\ \gamma_{6t}\right)'$$

$$I_t \simeq \mathcal{N}(0, \sigma_I^2)$$

- Transition equation: $\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t \Leftrightarrow$

$$\begin{pmatrix} \boldsymbol{\alpha}_t^{[L]} \\ \boldsymbol{\alpha}_t^{[S]} \end{pmatrix} = \begin{pmatrix} \mathbf{T}^{[L]} & \mathbf{O}_{[2\times 11]} \\ \mathbf{O}_{[11\times 2]} & \mathbf{T}^{[S]} \end{pmatrix} \begin{pmatrix} \boldsymbol{\alpha}_{t-1}^{[L]} \\ \boldsymbol{\alpha}_{t-1}^{[S]} \end{pmatrix} + \begin{pmatrix} \boldsymbol{\eta}_t^{[L]} \\ \boldsymbol{\eta}_t^{[S]} \end{pmatrix}$$

$$\mathbf{T}^{[S]} = \mathbf{Blockdiag}(\mathbf{C}_1\ \mathbf{C}_2\ \mathbf{C}_3\ \mathbf{C}_4\ \mathbf{C}_5\ \ -1)$$

$$\mathbf{C}_j = \begin{pmatrix} \cos\left(\frac{\pi\ j}{J/2}\right) & \sin\left(\frac{\pi\ j}{J/2}\right) \\ -\sin\left(\frac{\pi\ j}{J/2}\right) & \cos\left(\frac{\pi\ j}{J/2}\right) \end{pmatrix} \qquad j = 1, \ldots, 5.$$

$$\boldsymbol{\eta}_t^{[S]} = \left(\omega_{1,t}\ \omega_{1,t}^*\ \omega_{2,t}\ \omega_{2,t}^*\ \omega_{3,t}\ \omega_{3,t}^*\ \omega_{4,t}\ \omega_{4,t}^*\ \omega_{5,t}\ \omega_{5,t}^*\ \omega_{6,t}\right)'$$

$$\boldsymbol{\eta}_t \simeq \mathcal{N}\left(\mathbf{0}_{[13]}, \mathbf{H}\right)$$

$$\mathbf{H} = \begin{pmatrix} \mathbf{H}^{[L]} & \mathbf{O}_{[2\times 11]} \\ \mathbf{O}_{[11\times 2]} & \mathbf{H}^{[S]} \end{pmatrix} \qquad \mathbf{H}^{[S]} = \sigma_\omega^2 \mathbf{I}_{[11]}$$

Further reading

More details on modelling seasonal effects in structural time series models:

- Harvey (1989), Section 4.1: dummy seasonal and trigonometric seasonal models
- Durbin and Koopman (2012), Section 3.2: dummy seasonal and trigonometric seasonal models

Multivariate State Space Models

- Measurement equation: $\mathbf{y}_t = \mathbf{Z}\boldsymbol{\alpha}_t + \mathbf{I}_t$

  with $\mathbf{y}_t = (y_{1,t}, ..., y_{n,t})'$

$$\mathbf{I}_t \simeq \mathcal{N}(\mathbf{0}, \mathbf{G})$$

$$\mathbf{G} = \mathbf{Diag}(\sigma^2_{I_1}, ..., \sigma^2_{I_n})$$

- Transition equation: $\boldsymbol{\alpha}_t = \mathbf{T}\boldsymbol{\alpha}_{t-1} + \boldsymbol{\eta}_t$

$$\boldsymbol{\eta}_t \simeq \mathcal{N}(\mathbf{0}, \mathbf{H})$$

Multivariate State Space Models

- Kalman filter recursion:
  - Prediction equations:

$$\mathbf{a}_{t|t-1} = \mathbf{T}\mathbf{a}_{t-1}$$

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

  - Updating equations:

$$\mathbf{a}_t = \mathbf{a}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{Z}'\mathbf{F}_t^{-1}(\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})$$

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

  - Covariance matrix innovations:

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

- Required:
  - values for hyperparameters $\mathbf{G}$ and $\mathbf{H}$
  - initial values for $\mathbf{a}_0$ and $\mathbf{P}_0$

Initialization Kalman filter

- Starting values for the Kalman filter: $\mathbf{a}_0$ and $\mathbf{P}_0$
- Sometimes a-priori information: exact initialization
- If there is no a-priori information; distinguish between
  - Non-stationary state variables:
    * Diffuse initialization
    * $\mathbf{a}_0 = \mathbf{0}$
    * $\mathbf{P}_0 = \kappa \mathbf{I}$ with $\kappa = 10^7$
  - Stationary state variables:
    * Exact initialization
    * $\mathbf{a}_0 = \mathbf{0}$ (expected value)
    * $\mathbf{P}_0$ derived from its process
- State space model with d non stationary state variables
- First d observations are used to construct a proper distribution for the non-stationary state variables

Hyperparameters

- Kalman filter assumes variance components are known
- Generally unknown
- Therefore replaced by Maximum Likelihood (ML) estimates
- $\mathbf{\Psi} = (\sigma^2_{I_1}, \sigma^2_{\zeta_1}, \sigma^2_{\tau_1}, \sigma^2_{\omega_1}, ..., \sigma^2_{I_n}, \sigma^2_{\zeta_n}, \sigma^2_{\tau_n}, \sigma^2_{\omega_n})$

$\sigma^2_{I_1}, ..., \sigma^2_{I_n}$: variances disturbance terms measurement equation

$\sigma^2_{\zeta_1}, ..., \sigma^2_{\zeta_n}$: variances level disturbance terms of the trend component

$\sigma^2_{\tau_1}, ..., \sigma^2_{\tau_n}$: variances slope disturbance terms of the trend component

$\sigma^2_{\omega_1}, ..., \sigma^2_{\omega_n}$: variances seasonal disturbance terms

- ML assumes independently distributed observations
- Observations $\mathbf{y}_t$, $t = 1, ..., T$ are dependent
- Likelihood function must account for dependency through the so-called prediction error decomposition
- Joint density function time series: $L(\mathbf{y}_T, ..., \mathbf{y}_1; \mathbf{\Psi})$
- Since $p(a, b) = p(a|b)p(b)$, it follows that
  $L(\mathbf{y}_T, ..., \mathbf{y}_1; \mathbf{\Psi}) = L(\mathbf{y}_T|\mathbf{y}_{T-1}, ..., \mathbf{y}_1; \mathbf{\Psi})L(\mathbf{y}_{T-1}, ..., \mathbf{y}_1; \mathbf{\Psi})$

- Repeatedly applying gives an expression for the joint likelihood:

  $L(\mathbf{y}_T, ..., \mathbf{y}_1; \mathbf{\Psi}) = \left[\prod_{t=2}^{T} L(\mathbf{y}_t|\mathbf{y}_{t-1}, ..., \mathbf{y}_1; \mathbf{\Psi})\right] L(\mathbf{y}_1; \mathbf{\Psi})$

- From the assumption that disturbances $\boldsymbol{\eta}_t$ and $\mathbf{I}_t$ and the initial state vector are normally

  distributed and from the derivation of the Kalman filter it follows that conditionally on $(\mathbf{y}_{t-1}, ..., \mathbf{y}_1)$, $\mathbf{y}_t$ is normally distributed with mean $\hat{\mathbf{y}}_{t|t-1} = \mathbf{Z}\boldsymbol{\alpha}_{t|t-1}$ with covariance matrix $\mathbf{F}_t = \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}' + \mathbf{G}$

- This gives the following expression for the log of the likelihood function:

$$Log[L(\mathbf{y}_T, ..., \mathbf{y}_1; \mathbf{\Psi})] = -\frac{nT}{2} Log(2\pi) - 1/2 \sum_{t=1}^{T} Log(|\mathbf{F}_t|)$$

$$-1/2 \sum_{t=1}^{T} (\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1})' \mathbf{F}_t^{-1} (\mathbf{y}_t - \hat{\mathbf{y}}_{t|t-1}) \qquad (1)$$

- Note that the likelihood $L(\mathbf{y}_T, ..., \mathbf{y}_1; \mathbf{\Psi})$ is

  decomposed in the probability density function of the innovations. Therefore this approach is called the prediction error decomposition

- ML estimates for $\mathbf{\Psi}$ is obtained by maximizing (1) with respect to the elements of $\mathbf{\Psi}$
- Generally with a numerical procedure that repeatedly runs the Kalman filter
- For details see for example Harvey (1989), Section 3.2

Initialization Kalman filter:

- State space model with d non-stationary diffuse state variables
- First d observations are used to construct a proper distribution for the non-stationary state variables
- Log likelihood function is evaluated using observations $t = d + 1, ..., T$
- Makes likelihoods incomparable for models with different numbers of non-stationary variables

Model evaluation

Model assumptions:

- Disturbance terms measurement and transition equations are normally and serially independent distributed
- $\Rightarrow$ Innovations or one-step forecast errors are normally and serially independent distributed

  Follows from the prediction error decomposition.
- Model diagnostics are focussed on checking the assumption that standardized innovations are standard normal distributed

Standardized innovations:

$$\tilde{\nu}_t = \frac{\nu_t}{\sqrt{f_t}}$$

with:

$$\nu_t = y_t - \hat{y}_{t|t-1}$$

$$f_t = Var(\nu_t) = \mathbf{Z}\mathbf{P}_{t|t-1}\mathbf{Z}' + \sigma_I^2$$

Recall that the first d time periods are ignored in the evaluation of the likelihood function for d diffuse state variables.

Normality:

- First four moments:
  - Mean:
  $$m_1 = \frac{1}{(T-d)} \sum_{t=d}^{T} \tilde{\nu}_t$$
  - Moments $q = 2$, 3 and 4:
  $$m_q = \frac{1}{(T-d)} \sum_{t=d}^{T} (\tilde{\nu}_t - m_1)^q$$
  - Skewness:
  $$S = \frac{m_3}{\sqrt{m_2^3}} \simeq \mathcal{N}\left(0, \frac{6}{(T-d)}\right)$$
  - Kurtosis:
  $$K = \frac{m_4}{m_2^2} \simeq \mathcal{N}\left(3, \frac{24}{(T-d)}\right)$$
  - Bowman-Shenton test on normality
  $$N = (T-d)\left(\frac{S^2}{6} + \frac{(K-3)^2}{24}\right) \simeq \chi_2^2$$
  - QQ-plots
  - Histogram

– Plot of $\tilde{\nu}_t$ for $t = d, \ldots, T$ with 95% confidence interval

Heteroscedasticity:

F-test:

$$F = \frac{\sum_{t=d}^{h+d} \tilde{\nu}_t^2}{\sum_{t=T-h-d+1}^{T} \tilde{\nu}_t^2} \simeq F_h^h$$

F-test based on the sum over squared innovations for two exclusive subsets of the sample of equal length h.

Serial correlation:

- Autocorrelogram based on autocorrelations

$$c_j = \frac{1}{T-d} \sum_{t=d+j+1}^{T} \frac{(\tilde{\nu}_t - m_1)(\tilde{\nu}_{t-j} - m_1)}{m_2}$$

for $j = 1, \ldots, 12$ (or 24) 95% confidence interval: $\left[\frac{1.96}{\sqrt{(T-d)}}, \frac{-1.96}{\sqrt{(T-d)}}\right]$

- Liung Box test

$$Q = (T-d)(T-d+2) \sum_{j=1}^{h} \frac{c_j^2}{T-d-j} \simeq \chi_h^2$$

- Durbin-Watson test

See Durbin and Koopman (2012) Section 2.12 and 7.5 for more details.

Model selection and comparison

Likelihood-based diagnostics:

- Akaike information criterion

$$AIC = \frac{1}{(T-d)}[-2log(L) + 2(q+p)]$$

- Bayesian information criterion

$$BIC = \frac{1}{(T-d)}[-2log(L) + log(T-d)(q+p)]$$

  q: number of hyperparameters (estimated with ML)

  p: number of state variables

  d: number of non-stationary state variables

  L: Likelihood (see Block 10)

- Nested models: Likelihood ratio test

$$LR = 2 * [\ log(L[M_{alt}]) - log(L[M_{null}])\ ] \simeq \chi^2_r$$

  – M_{alt}: extended model under alt. hypothesis

  – M_{null}: reduced model under the null hypothesis

  – r: d.f. $\Rightarrow$ number of parameters equal to zero

  - Example of nested models: if one or more hyperparameters are assumed to be zero
  - Remark: likelihoods are difficult to compare for models with different numbers of non-stationary variables
  - Trick: Exact initialization with smoothed estimates for state variables

- Evaluate the contribution of state variables: plots of the smoothed estimates with 95% confidence interval

See Durbin and Koopman (2012) Section 7.4 for more details.

Exercise

Which models are nested?

1. $y_t = L_t + S_t + I_t$ versus $y_t = L_t + I_t$
2. Model with local linear trend model versus model with a smooth trend model (see Block 3 for definitions)
3. Model with a time dependent dummy seasonal component versus a model with time invariant seasonal component
4. Model with time dependent dummy seasonal component with time dependent trigonometric seasonal component

Software

Software for STM:

- Eviews
- SAS
- R: package KFAS
- Oxmetrics:
  - STAMP
  - Ssfpack

References

Box, G. and Jenkins, G. (1989). *Time series analysis: forecasting an control*. Holden-Day, San Francisco.

Durbin, J. and Koopman, S. (2012). *Time Series Analysis by State Space Methods (second edition)*. Oxford University Press, Oxford.

Harvey, A. (1989). *Forecasting, structural time series models and the Kalman filter*. Cambridge University Press, Cambridge.