QBUS 6840 Lecture 11

State-Space Models

Professor Junbin Gao

The University of Sydney Business School

Simple Exponential Forecasting: Review

- From Lecture 5, we know that the simple exponential smoothing (aka EWMA (exponential weighted moving average)) forecasts minimise the MSE for the following statistical model

$$\begin{aligned} y_t &= \mu_t + \varepsilon_t; \qquad \varepsilon_t \sim \mathcal{N}(0, \sigma_e^2) \\ \mu_{t+1} &= \mu_t + \xi_t \qquad \xi_t \sim \mathcal{N}(0, \sigma_\xi^2) \end{aligned}$$

  where $\{\varepsilon_t\}$ and $\{\xi_t\}$ are two independent Gaussian white noise processes.

- This model is our first example of the State-space models
- μ_t is a pure random walk with initial value μ_1 and y_t is an observed version of μ_t.
- μ_t is referred to as the "level" of the time series, which is not observable. It is called the `state`.

Inference

- The purpose is to infer relevant properties of the μ_t's from a knowledge of the observations $\{y_1, y_2, ..., y_T\}$
- We assume initially that

$$\mu_1 \sim \mathcal{N}(u_1, P_1)$$

  where u_1, P_1, σ_e^2 and σ_ξ^2 are known.
- The model for y_t is non-stationary in this case.
- Three types of inference: Denote $\mathcal{F}_t = \{y_1, y_2, ..., y_t\}$
  - Filtering: recovering the state μ_t given $\mathcal{F}_t$
  - Predicting: forecasting μ_{t+h} or y_{t+h} for $h > 0$ given $\mathcal{F}_t$
  - Smoothing: estimating μ_t given the full set of information $\mathcal{F}_T$.

Notations

- Denote

$$u_{t|j} = E(\mu_t|\mathcal{F}_j) \text{ and } P_{t|j} = \text{var}(\mu_t|\mathcal{F}_j)$$

- Similarly we can define $y_{t|j} = E(y_t|\mathcal{F}_j)$ the conditional mean of y_t given $\mathcal{F}_j$.
- One-step ahead forecast error and its variance given $\mathcal{F}_{t-1}$

$$e_t = y_t - u_{t|t-1} \text{ and } F_t = \text{var}(e_t|\mathcal{F}_{t-1})$$

  Since e_t is independent of $\mathcal{F}_{t-1}$ (see the next page), we have

$$F_t = \text{var}(e_t|\mathcal{F}_{t-1}) = \text{var}(e_t)(unconditional)$$

The Kalman Filter

- First

$$y_{t|t-1} = E(y_t|\mathcal{F}_{t-1}) = E(\mu_t + \varepsilon_t|\mathcal{F}_{t-1}) = E(\mu_t|\mathcal{F}_{t-1}) = u_{t|t-1}$$

- It is easy to show that

$$E(e_t|\mathcal{F}_{t-1}) = E(y_t - u_{t|t-1}|\mathcal{F}_{t-1}) = y_{t|t-1} - u_{t|t-1} = 0$$

hence

$$\begin{aligned} E(e_t) =& E(E(e_t|\mathcal{F}_{t-1})) = 0 \\ \mathrm{corr}(e_t, y_j) =& E(e_t y_j) - E(e_t)E(y_j) \\ =& E(e_t y_j) = E(E(e_t|\mathcal{F}_{t-1}) y_j) = 0 \end{aligned}$$

so e_t and y_j are independent for $j = 1, 2, ..., t-1$.

The Kalman Filter

- When $\mathcal{F}_t$ is fixed $\Rightarrow \mathcal{F}_{t-1}, y_t$ are fixed $\Rightarrow e_t$ and $\mathcal{F}_{t-1}$ are fixed, and vice versa.
- Consequently

$$u_{t|t} := E(\mu_t|\mathcal{F}_t) = E(\mu_t|\mathcal{F}_{t-1}, e_t) \text{ and}$$
$$P_{t|t} := \mathrm{var}(\mu_t|\mathcal{F}_t) = \mathrm{var}(\mu_t|\mathcal{F}_{t-1}, e_t)$$

- Since all variables are normally distributed, this gives the base for applying the famous theorem on Normal distributions: For all Normal variables,

$$E(x|y,z) = E(x|y) + \Sigma_{xz|y}\Sigma_{zz|y}^{-1}(z - \mu_{z|y}) \text{ and}$$
$$\mathrm{var}(x|y,z) = \mathrm{var}(x|y) - \Sigma_{xz|y}\Sigma_{zz|y}^{-1}\Sigma_{zx|y}$$

The Kalman Filter

- It is sufficient to consider the joint conditional distribution of (μ_t, e_t) given $\mathcal{F}_{t-1}$.
- The conditional distribution of e_t given $\mathcal{F}_{t-1}$ is normal with mean zero and variance given in F_t, and that of μ_t given $\mathcal{F}_{t-1}$ is also normal with mean $u_{t|t-1}$ and variance $P_{t|t-1}$.
- Thus the joint distribution of (μ_t, e_t) given $\mathcal{F}_{t-1}$ is also normal.
- We will use the previous theorem to calculate both $u_{t|t} = E(\mu_t|\mathcal{F}_t)$ and $P_{t|t} = \mathrm{var}(\mu_t|\mathcal{F}_t)$, by taking $x = \mu_t$ and $y = \mathcal{F}_{t-1}$ and $z = e_t$. Hence

$$E(\mu_t|\mathcal{F}_t) = E(\mu_t|\mathcal{F}_{t-1}, e_t) = E(\mu_t|\mathcal{F}_{t-1}) + \mathrm{Cov}(\mu_t, e_t|\mathcal{F}_{t-1})\mathrm{var}(e_t)^{-1}e_t$$

and

$$\mathrm{var}(\mu_t|\mathcal{F}_t) = \mathrm{var}(\mu_t|\mathcal{F}_{t-1}, e_t) = \mathrm{var}(\mu_t|\mathcal{F}_{t-1}) - \mathrm{Cov}(\mu_t, e_t|\mathcal{F}_{t-1})^2\mathrm{var}(e_t)^{-1}$$

The Kalman Filter

- Let us calculate the conditional covariance

$$\begin{aligned}
\mathrm{Cov}(\mu_t, e_t|\mathcal{F}_{t-1}) &= E(\mu_t(y_t - u_{t|t-1})|\mathcal{F}_{t-1}) = E(\mu_t(\mu_t + \varepsilon_t - u_{t|t-1})|\mathcal{F}_{t-1}) \\
&= E(\mu_t^2 + \mu_t\varepsilon_t - \mu_t u_{t|t-1}|\mathcal{F}_{t-1}) \\
&= E(\mu_t^2|\mathcal{F}_{t-1}) + E(\mu_t\varepsilon_t|\mathcal{F}_{t-1}) - E(\mu_t|\mathcal{F}_{t-1})u_{t|t-1} \\
&= E(\mu_t^2|\mathcal{F}_{t-1}) + 0 - E(\mu_t|\mathcal{F}_{t-1})^2 \\
&= \mathrm{var}(\mu_t|\mathcal{F}_{t-1})] = P_{t|t-1}
\end{aligned}$$

- And

$$\begin{aligned}
F_t \equiv \mathrm{var}(e_t|\mathcal{F}_{t-1}) &= \mathrm{var}(\mu_t + \varepsilon_t - u_{t|t-1}|\mathcal{F}_{t-1}) \\
&= \mathrm{var}(\mu_t - u_{t|t-1}|\mathcal{F}_{t-1}) + \mathrm{var}(\varepsilon_t) \\
&= E((\mu_t - u_{t|t-1})^2|\mathcal{F}_{t-1}) + \sigma_e^2 \\
&= \mathrm{var}(\mu_t|\mathcal{F}_{t-1}) + \sigma_e^2 = P_{t|t-1} + \sigma_e^2
\end{aligned}$$

- Also

$$\begin{aligned}
P_{t+1|t} &= \mathrm{var}(\mu_{t+1}|\mathcal{F}_t) = \mathrm{var}(\mu_t + \xi_t|\mathcal{F}_t) = \mathrm{var}(\mu_t|\mathcal{F}_t) + \sigma_\xi^2 = P_{t|t} + \sigma_\xi^2 \\
u_{t+1|t} &= E[\mu_{t+1}|\mathcal{F}_t] = E(\mu_t + \xi_t|\mathcal{F}_t) = E(\mu_t|\mathcal{F}_t) = u_{t|t}
\end{aligned}$$

The Kalman Filter: Overall Algorithm

- The overall algorithm

$$
\begin{aligned}
e_t &= y_t - u_{t|t-1} \\
F_t &= P_{t|t-1} + \sigma_e^2 \\
K_t &= P_{t|t-1}/F_t \\
u_{t+1|t} &= u_{t|t-1} + K_t e_t \\
P_{t+1|t} &= P_{t|t-1}(1 - K_t) + \sigma_\xi^2
\end{aligned}
$$

Starting with $u_{1|0}, P_{1|0}, \sigma_e^2$ and σ_ξ^2.

Forecasting

- The theory of forecasting for the local level model: we regard forecasting as filtering the observations $\{y_1, ..., y_n, y_{n+1}, ..., y_{n+J}\}$ using Kalman Filtering and treating the last J observations $\{y_{n+1}, ..., y_{n+J}\}$ as missing.
- Letting

$$\begin{aligned}
\widehat{u}_{n+j} &= E[\mu_{n+j}|\mathcal{F}_n] \\
\widehat{P}_{n+j+1} &= \widehat{P}_{n+j} + \sigma^2_{\xi},\ j = 1, 2, ..., J-1
\end{aligned}$$

with $\widehat{u}_{n+1} = u_{n+1|n}$ and $\widehat{P}_{n+1} = P_{n+1|n}$ obtained from Kalman Filtering

- The forecasts of y are

$$\begin{aligned}
\widehat{y}_{n+j} &= E[y_{n+j}|\mathcal{F}_n] = E[\mu_{n+j}|\mathcal{F}_n] + E[\varepsilon_{n+j}|\mathcal{F}_n] = \widehat{u}_{n+j} \\
\widehat{F}_{n+j} &= \mathrm{var}[\mu_{n+j}|\mathcal{F}_n] + \mathrm{var}[\varepsilon_{n+j}|\mathcal{F}_n] = \widehat{P}_{n+j} + \sigma^2_e,\ j = 1, 2, ..., J
\end{aligned}$$

Forecasting: Example

- First we use the Kalman algorithm from $\{y_1, \ldots, y_n\}$ to calculate all

$$u_{n+1|n}, u_{n|n-1}, u_{n-1|n-2}, \ldots, u_{2|1}, u_{1|0}(\text{assumed})$$
$$P_{n+1|n}, P_{n|n-1}, P_{n-1|n-2}, \ldots, P_{2|1}, u_{1|0}(\text{assumed})$$

- Hence the forecasts for u_{n+j} are

$$\widehat{u}_{n+1} := E[\mu_{n+1}|\mathcal{F}_n] = u_{n+1|n}$$
$$\widehat{u}_{n+2} := E[\mu_{n+2}|\mathcal{F}_n] = E[\mu_{n+1} + \xi_{n+1}|\mathcal{F}_n] = E[\mu_{n+1}|\mathcal{F}_n] = u_{n+1|n}$$
$$\widehat{u}_{n+3} := E[\mu_{n+3}|\mathcal{F}_n] = E[\mu_{n+3} + \xi_{n+2}|\mathcal{F}_n] = E[\mu_{n+2}|\mathcal{F}_n] = u_{n+1|n}$$
$$\vdots$$
$$\widehat{u}_{n+J} := E[\mu_{n+J}|\mathcal{F}_n] = u_{n+1|n}$$

- Hence

$$\widehat{y}_{n+j} = \widehat{u}_{n+j} = u_{n+1|n}$$

- Try how to find all the variances of each forecast?

Forecasting: Example

- Try how to find all the variances of each forecast?

$$
\begin{aligned}
\widehat{F}_{n+1} &= \text{var}[\mu_{n+1}|\mathcal{F}_n] + \text{var}[\varepsilon_{n+1}|\mathcal{F}_n] = P_{n+1|n} + \sigma_e^2, \\
\widehat{F}_{n+2} &= \text{var}[\mu_{n+2}|\mathcal{F}_n] + \text{var}[\varepsilon_{n+2}|\mathcal{F}_n] \\
&= \text{var}[\mu_{n+1}|\mathcal{F}_n] + \text{var}[\xi_{n+1}|\mathcal{F}_n] + \sigma_e^2 \\
&= P_{n+1|n} + \sigma_\xi^2 + \sigma_e^2 \\
\widehat{F}_{n+3} &= P_{n+1|n} + 2\sigma_\xi^2 + \sigma_e^2 \\
&\vdots
\end{aligned}
$$

Estimating σ_e^2 and σ_ξ^2

- Kalman filtering provides an efficient way to evaluate the likelihood function of the data for estimation
- The overall likelihood function

$$
\begin{aligned}
p(y_1, \dots, y_T | \sigma_e^2, \sigma_\xi^2) =& p(y_1 | \sigma_e^2, \sigma_\xi^2) \prod_{t=2}^{T} p(y_t | \mathcal{F}_{t-1}, \sigma_e^2, \sigma_\xi^2)) \\
=& p(y_1 | \sigma_e^2, \sigma_\xi^2) \prod_{t=2}^{T} p(e_t | \mathcal{F}_{t-1}, \sigma_e^2, \sigma_\xi^2))
\end{aligned}
$$

  where $y_1 \sim \mathcal{N}(u_{1|0}, P_{1|0})$ and $e_t = y_t - u_{t|t-1}$ is normally distributed with mean zero and variance F_1
- Consequently assuming $u_{1|0}$ and $P_{1|0}$ are known, then Maximum Likelihood can be conducted to estimate σ_e^2 and σ_ξ^2

Linear Gaussian State Space Model

- Linear Gaussian state space model is defined in three parts:
  - State Equation:

$$\mu_{t+1} = \mathbf{d}_t + T_t \mu_t + R_t \xi_t;\ \xi_t \sim \mathcal{N}(0, Q_t)$$

  - Observation Equation:

$$\mathbf{y}_t = \mathbf{c}_t + Z_t \mu_t + \varepsilon_t;\ \varepsilon_t \sim \mathcal{N}(0, H_t)$$

  - Initial state distribution

$$\mu_1 \sim \mathcal{N}(\mathbf{u}_{1|0}, P_{1|0})$$

- The state μ_t are m-dimensional vector and each observation $\mathbf{y}_t$ is a k-dimensional vector.
- The matrices Z_t, T_t, R_t, H_t and Q_t (please identify their sizes) are independent of $\{\varepsilon_1, ..., \varepsilon_T\}$ and $\{\xi_1, ..., \xi_T\}$.

Properties

- This state space model is linear with Gaussian disturbances: All the variables will be in Gaussian distributions and nice theory can be applied;
- Those matrices in the model usually consist of unknown parameters, or the entire matrices are unknown and to be estimated from the data
- An estimation algorithm has two aspects:
  - recovering the unobservable states in terms of prediction, filtering or smoothing;
  - estimating unknown parameters by using maximum likelihood or other criteria;
- State space models cover a wide range of models and techniques: dynamic regression, ARIMA, UC (Unobserved components), latent variable models and many ad-hoc filters; most are out of our scope

Kalman Filtering Algorithm

- The unobserved state μ_t can be estimated from the observations with the Kalman filter:

$$
\begin{aligned}
\mathbf{e}_t &= \mathbf{y}_t - \mathbf{c}_t - Z_t \mathbf{u}_{t|t-1} \\
F_t &= Z_t P_{t|t-1} Z_t^T + H_t \\
K_t &= T_t P_{t|t-1} Z_t^T F_t^{-1} \\
\mathbf{u}_{t+1|t} &= \mathbf{d}_t + T_t \mathbf{u}_{t|t-1} + K_t \mathbf{e}_t \\
L_t &= T_t - K_t Z_t \\
P_{t+1|t} &= T_t P_{t|t-1} L_t^T + R_t Q_t R_t^T
\end{aligned}
$$

starting with given values for $\mathbf{u}_{1|0}$ and $P_{1|0}$, and Z_t^T, L_t^T and R_t^T are transpose of matrices Z_t, L_t and R_t respectively

Example: Regression with Time Varying Coefficients

- Regressors in
$$Z_t = X_t$$
and
$$T_t = I \qquad R_t = I$$
- Regression model with coefficient μ_t following a random walk:
$$\begin{aligned} \mu_{t+1} &= \mu_t + \xi_t \\ y_t &= X_t \mu_t + \varepsilon_t \text{ (linear regression model)} \end{aligned}$$
where we have take $c_t = d_t = 0$.

Example: The Exponential Model

- We have defined

$$
\begin{aligned}
y_t = \mu_t + \varepsilon_t; \quad & \varepsilon_t \sim \mathcal{N}(0, \sigma_e^2) \\
\mu_{t+1} = \mu_t + \xi_t \quad & \xi_t \sim \mathcal{N}(0, \sigma_\xi^2)
\end{aligned}
$$

- In the state equation, we have

$$
T_t = 1; \quad R_t = 1; \quad Q_t = \sigma_\xi^2
$$

- In observation equation

$$
Z_t = 1; \quad H_t = \sigma_e^2
$$

- Tasks `Lecture11_Example01.py`
  - Parameter estimation: σ_e^2 and σ_ξ^2
  - Recovering $\{\mu_t\}$

Example: Holt's Linear Trend Model `Lecture11_Example02.py`

- Let us redefine Holt's Linear Trend Model

$$\begin{aligned} y_t &= \alpha_t + \varepsilon_t; \quad \varepsilon_t \sim \mathcal{N}(0, \sigma_e^2) \\ \alpha_{t+1} &= \alpha_t + \nu_t + \eta_t \quad \eta_t \sim \mathcal{N}(0, \sigma_\eta^2) \\ \nu_{t+1} &= \nu_t + \zeta_t \quad \zeta_t \sim \mathcal{N}(0, \sigma_\xi^2) \end{aligned}$$

- In State equation:

$$\mu_{t+1} = T_t \mu_t + R_t \xi_t, \quad \xi_t \sim \mathcal{N}(0, Q_t)$$

with

$$\mu_t = \begin{pmatrix} \alpha_t \\ \nu_t \end{pmatrix}, \xi_t = \begin{pmatrix} \eta_t \\ \zeta_t \end{pmatrix}, T_t = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, R_t = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, Q_t = \begin{pmatrix} \sigma_\eta^2 & 0 \\ 0 & \sigma_\zeta^2 \end{pmatrix}$$

- In observation equation

$$Z_t = [1 \;\; 0]; \quad H_t = \sigma_e^2$$

Example: MA(1)

- The MA(1) model

$$y_t = \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- Define

$$\mu_t = \begin{pmatrix} y_t \\ \theta_1 \varepsilon_t \end{pmatrix}$$

- State equation:

$$\mu_{t+1} = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} \mu_t + \begin{pmatrix} 1 \\ \theta_1 \end{pmatrix} \varepsilon_{t+1}$$

with

$$Q_t = \sigma_\eta^2$$

- Observation equation

$$y_t = Z_t \mu_t, \quad Z_t = [1 \;\; 0]; \quad H_t = 0$$

Example: ARMA(2,1)

- The ARMA(2,1) model

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- Define

$$\mu_t = \begin{pmatrix} y_t \\ \phi_2 y_{t-1} + \theta_1 \varepsilon_t \end{pmatrix}$$

- State equation:

$$\mu_{t+1} = \begin{pmatrix} \phi_1 & 1 \\ \phi_2 & 0 \end{pmatrix} \mu_t + \begin{pmatrix} 1 \\ \theta_1 \end{pmatrix} \varepsilon_{t+1}$$

  with

$$Q_t = \sigma_\eta^2$$

- Observation equation

$$y_t = Z_t \mu_t, \quad Z_t = [1 \;\; 0]; \quad H_t = 0$$

Example: ARIMA(2,1,1)

- The ARIMA(2,1,1) model [Note: Here we define $\Delta y_t = y_{t+1} - y_t$]

$$\Delta y_t = \phi_1 \Delta y_{t-1} + \phi_2 \Delta y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- Define the state vector

$$\mu_t = \begin{pmatrix} y_t \\ \Delta y_t \\ \phi_2 \Delta y_{t-1} + \theta_1 \varepsilon_t \end{pmatrix}$$

- State equation:

$$\mu_{t+1} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & \phi_1 & 1 \\ 0 & \phi_2 & 0 \end{pmatrix} \mu_t + \begin{pmatrix} 0 \\ 1 \\ \theta_1 \end{pmatrix} \varepsilon_{t+1}$$

  with

$$Q_t = \sigma_\eta^2$$

- Observation equation

$$y_t = Z_t \mu_t, \quad Z_t = [1 \;\; 0 \;\; 0]; \quad H_t = 0$$

Example: ARIMA(2,2,1)

- The ARIMA(2,1,1) model

$$\Delta^2 y_t = \phi_1 \Delta^2 y_{t-1} + \phi_2 \Delta^2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1}$$

- Define the state vector

$$\mu_t = \begin{pmatrix} y_t \\ \Delta y_t \\ \Delta^2 y_t \\ \phi_2 \Delta^2 y_{t-1} + \theta_1 \varepsilon_t \end{pmatrix}$$

- State equation:

$$\mu_{t+1} = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & \phi_1 & 1 \\ 0 & 0 & \phi_2 & 0 \end{pmatrix} \mu_t + \begin{pmatrix} 0 \\ 0 \\ 1 \\ \theta_1 \end{pmatrix} \varepsilon_{t+1}$$

- Write each equation from the above system
- Write the observation equation

Example: ARMA(p,d,q)

- All ARIMA(p,d,q) models have a state space representation (many forms available)

$$y_t = \Delta^d x_t$$

$$\phi(B) = \theta(B)\varepsilon_t$$

$$y_t = \sum_{j=1}^{r} \phi_j y_{t-j} + \varepsilon_t + \sum_{j=1}^{r-1} \theta_j \varepsilon_{t-j}$$

$$Z_t = [1, 0, ..., 0]$$

$$\mu_t = \begin{pmatrix} y_t \\ \phi_2 y_{t-1} + \cdots + \phi_r y_{t-r+1} + \theta_1 \varepsilon_t + \cdots + \theta_{r-1} \varepsilon_{t-r+2} \\ \phi_3 y_{t-1} + \cdots + \phi_r y_{t-r+2} + \theta_2 \varepsilon_t + \cdots + \theta_{r-1} \varepsilon_{t-r+3} \\ \vdots \\ \phi_r y_{t-1} + \theta_{r-1} \varepsilon_t \end{pmatrix}$$

with $r = \max(p, q+1)$ with $\phi_j = 0$ if $j > p$ and $\theta_j = 0$ if $j > q$.

Example: ARIMA(p,d,q)

- We have

$$T_t = \begin{pmatrix} \phi_1 & 1 & 0 & \cdots & 0 \\ \vdots & & & & \\ \phi_{r-1} & 0 & 0 & \cdots & 1 \\ \phi_r & 0 & 0 & \cdots & 0 \end{pmatrix}$$

  and

$$R_r = R = \begin{pmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{r-1} \end{pmatrix}$$

- $\zeta_t = \varepsilon_{t+1}$
- Observation equation

$$y_t = Z_t \mu_t$$

  with $H_t = 0 \Rightarrow \varepsilon_t = 0$