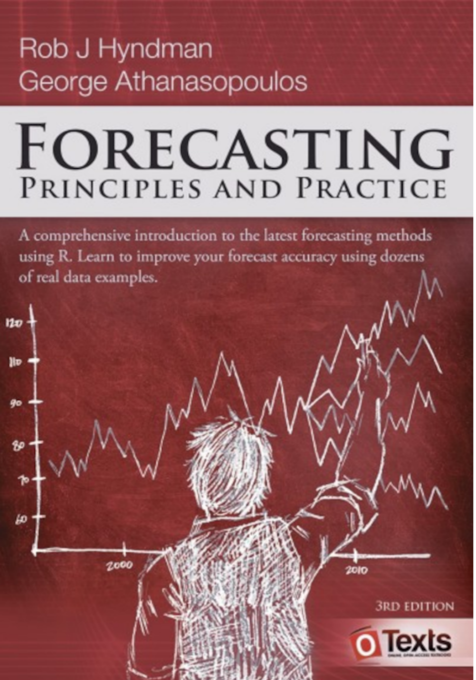

7. Time series regression models

7.9 Matrix formulation

OTexts.org/fpp3/

Matrix formulation

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t.$$

Matrix formulation

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t.$$

Let $\boldsymbol{y} = (y_1, \ldots, y_T)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_T)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ and

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \ldots & x_{k,1} \\ 1 & x_{1,2} & x_{2,2} & \ldots & x_{k,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1,T} & x_{2,T} & \ldots & x_{k,T} \end{bmatrix}.$$

Matrix formulation

$$y_t = \beta_0 + \beta_1 x_{1,t} + \beta_2 x_{2,t} + \cdots + \beta_k x_{k,t} + \varepsilon_t.$$

Let $\boldsymbol{y} = (y_1, \dots, y_T)'$, $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_T)'$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_k)'$ and

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_{1,1} & x_{2,1} & \dots & x_{k,1} \\ 1 & x_{1,2} & x_{2,2} & \dots & x_{k,2} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{1,T} & x_{2,T} & \dots & x_{k,T} \end{bmatrix}.$$

Then

$$\boldsymbol{y} = \boldsymbol{X}\beta + \varepsilon.$$

Matrix formulation

Least squares estimation

Minimize: $(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$

Matrix formulation

Least squares estimation

Minimize: $(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$

Differentiate wrt β gives

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

Matrix formulation

Least squares estimation

Minimize: $(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$

Differentiate wrt β gives

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

(The "normal equation".)

Matrix formulation

Least squares estimation

Minimize: $(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$

Differentiate wrt β gives

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

(The “normal equation”.)

$$\hat{\sigma}^2 = \frac{1}{T-k-1}(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})'(\boldsymbol{y} - \boldsymbol{X}\hat{\beta})$$

Note: If you fall for the dummy variable trap, $(\boldsymbol{X}'\boldsymbol{X})$ is a singular matrix.

Likelihood

If the errors are iid and normally distributed, then

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{I}).$$

Likelihood

If the errors are iid and normally distributed, then

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I}).$$

So the likelihood is

$$L = \frac{1}{\sigma^T(2\pi)^{T/2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)\right)$$

Likelihood

If the errors are iid and normally distributed, then

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X}\beta, \sigma^2 \boldsymbol{I}).$$

So the likelihood is

$$L = \frac{1}{\sigma^T (2\pi)^{T/2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)\right)$$

which is maximized when $(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$ is minimized.

Likelihood

If the errors are iid and normally distributed, then

$$\boldsymbol{y} \sim \mathrm{N}(\boldsymbol{X}\beta, \sigma^2\boldsymbol{I}).$$

So the likelihood is

$$L = \frac{1}{\sigma^T(2\pi)^{T/2}} \exp\left(-\frac{1}{2\sigma^2}(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)\right)$$

which is maximized when $(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta)$ is minimized.

So MLE = OLS.

Cross-validation

Fitted values

$$\begin{aligned}\hat{\boldsymbol{y}} &= \boldsymbol{X}\hat{\beta} \\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y} \\ &= \boldsymbol{H}\boldsymbol{y} \qquad\qquad \text{where } \boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.\end{aligned}$$

Cross-validation

Fitted values

$$\begin{aligned}\hat{\boldsymbol{y}} &= \boldsymbol{X}\hat{\boldsymbol{\beta}}\\ &= \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}\\ &= \boldsymbol{H}\boldsymbol{y} \qquad\qquad \text{where } \boldsymbol{H} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'.\end{aligned}$$

LOO cross-validation MSE

$$\text{CV} = \frac{1}{T}\sum_{t=1}^{T}[e_t/(1-h_t)]^2$$

- e_t = residual at time t (from fitting model to all data)
- $h_1, \dots, h_T$ are the diagonals of $\boldsymbol{H}$.

Multiple regression forecasts

Optimal forecasts

$$\hat{y}^* = \text{E}(y^*|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}^*) = \boldsymbol{x}^*\hat{\boldsymbol{\beta}} = \boldsymbol{x}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

where $\boldsymbol{x}^*$ is a row vector containing the values of the predictors for the forecasts (in the same format as $\boldsymbol{X}$).

Multiple regression forecasts

Optimal forecasts

$$\hat{y}^* = \text{E}(y^*|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}^*) = \boldsymbol{x}^*\hat{\beta} = \boldsymbol{x}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

where $\boldsymbol{x}^*$ is a row vector containing the values of the predictors for the forecasts (in the same format as $\boldsymbol{X}$).

Forecast variance

$$\text{Var}(y^*|\boldsymbol{X}, \boldsymbol{x}^*) = \sigma^2\left[1 + \boldsymbol{x}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{x}^*)'\right]$$

Multiple regression forecasts

Optimal forecasts

$$\hat{y}^* = \text{E}(y^*|\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{x}^*) = \boldsymbol{x}^*\hat{\boldsymbol{\beta}} = \boldsymbol{x}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$$

where $\boldsymbol{x}^*$ is a row vector containing the values of the predictors for the forecasts (in the same format as $\boldsymbol{X}$).

Forecast variance

$$\text{Var}(y^*|\boldsymbol{X}, \boldsymbol{x}^*) = \sigma^2\left[1 + \boldsymbol{x}^*(\boldsymbol{X}'\boldsymbol{X})^{-1}(\boldsymbol{x}^*)'\right]$$

- This ignores any errors in $\boldsymbol{x}^*$.
- 95% prediction intervals assuming normal errors:

$$\hat{y}^* \pm 1.96\sqrt{\text{Var}(y^*|\boldsymbol{X}, \boldsymbol{x}^*)}.$$