

Rob J Hyndman
George Athanasopoulos

FORECASTING

PRINCIPLES AND PRACTICE

A comprehensive introduction to the latest forecasting methods using R. Learn to improve your forecast accuracy using dozens of real data examples.



3RD EDITION

 **OTexts**
Open Texts Publishing

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

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

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

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

Then

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

Matrix formulation

Least squares estimation

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

Matrix formulation

Least squares estimation

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

Differentiate wrt $\boldsymbol{\beta}$ gives

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

Matrix formulation

Least squares estimation

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

Differentiate wrt $\boldsymbol{\beta}$ gives

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

(The “normal equation”.)

Matrix formulation

Least squares estimation

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

Differentiate wrt β gives

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

(The “normal equation”.)

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

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

Likelihood

If the errors are iid and normally distributed, then

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

Likelihood

If the errors are iid and normally distributed, then

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

So the likelihood is

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

Likelihood

If the errors are iid and normally distributed, then

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

So the likelihood is

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

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

Likelihood

If the errors are iid and normally distributed, then

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

So the likelihood is

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

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

So MLE = OLS.

Multiple regression forecasts

Optimal forecasts

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

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

Multiple regression forecasts

Optimal forecasts

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

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

Forecast variance

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

Multiple regression forecasts

Optimal forecasts

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

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

Forecast variance

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

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

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