

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 FOR PRACTICE

## 7. Time series regression models

### 7.5 Selecting predictors

[OTexts.org/fpp3/](https://OTexts.org/fpp3/)

# Comparing regression models

Computer output for regression will always give the  $R^2$  value. This is a useful summary of the model.

- It is equal to the square of the correlation between  $y$  and  $\hat{y}$ .
- It is often called the “coefficient of determination”.
- It can also be calculated as follows:

$$R^2 = \frac{\sum(\hat{y}_t - \bar{y})^2}{\sum(y_t - \bar{y})^2}$$

- It is the proportion of variance accounted for (explained) by the predictors.

# Comparing regression models

However ...

- $R^2$  does not allow for “degrees of freedom”.
- Adding *any* variable tends to increase the value of  $R^2$ , even if that variable is irrelevant.

# Comparing regression models

However ...

- $R^2$  does not allow for “degrees of freedom”.
- Adding *any* variable tends to increase the value of  $R^2$ , even if that variable is irrelevant.

To overcome this problem, we can use *adjusted*  $R^2$ :

$$\bar{R}^2 = 1 - (1 - R^2) \frac{T - 1}{T - k - 1}$$

where  $k$  = no. predictors and  $T$  = no. observations.

# Comparing regression models

However ...

- $R^2$  does not allow for “degrees of freedom”.
- Adding *any* variable tends to increase the value of  $R^2$ , even if that variable is irrelevant.

To overcome this problem, we can use *adjusted*  $R^2$ :

$$\bar{R}^2 = 1 - (1 - R^2) \frac{T - 1}{T - k - 1}$$

where  $k$  = no. predictors and  $T$  = no. observations.

**Maximizing  $\bar{R}^2$  is equivalent to minimizing  $\hat{\sigma}^2$ .**

$$\hat{\sigma}^2 = \frac{1}{T - k - 1} \sum_{t=1}^T \varepsilon_t^2$$

# Akaike's Information Criterion

$$\text{AIC} = -2 \log(L) + 2(k + 2)$$

where  $L$  is the likelihood and  $k$  is the number of predictors in the model.

# Akaike's Information Criterion

$$\text{AIC} = -2 \log(L) + 2(k + 2)$$

where  $L$  is the likelihood and  $k$  is the number of predictors in the model.

- AIC penalizes terms more heavily than  $\bar{R}^2$ .
- Minimizing the AIC is asymptotically equivalent to minimizing MSE via **leave-one-out cross-validation** (for any linear regression).

## Corrected AIC

For small values of  $T$ , the AIC tends to select too many predictors, and so a bias-corrected version of the AIC has been developed.

$$\text{AIC}_C = \text{AIC} + \frac{2(k+2)(k+3)}{T-k-3}$$

As with the AIC, the  $\text{AIC}_C$  should be minimized.



# Bayesian Information Criterion

$$\text{BIC} = -2 \log(L) + (k + 2) \log(T)$$

where  $L$  is the likelihood and  $k$  is the number of predictors in the model.

# Bayesian Information Criterion

$$\text{BIC} = -2 \log(L) + (k + 2) \log(T)$$

where  $L$  is the likelihood and  $k$  is the number of predictors in the model.

- BIC penalizes terms more heavily than AIC
- Also called SBIC and SC.
- Minimizing BIC is asymptotically equivalent to leave- $v$ -out cross-validation when  $v = T[1 - 1/(\log(T) - 1)]$ .

# Leave-one-out cross-validation

For regression, leave-one-out cross-validation is faster and more efficient than time-series cross-validation.

- Select one observation for test set, and use *remaining* observations in training set. Compute error on test observation.
- Repeat using each possible observation as the test set.
- Compute accuracy measure over all errors.

# Cross-validation

## Traditional evaluation

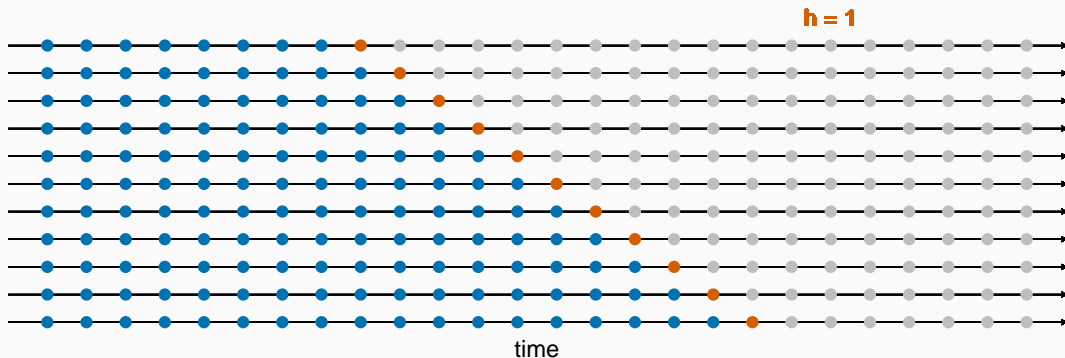


# Cross-validation

## Traditional evaluation



## Time series cross-validation

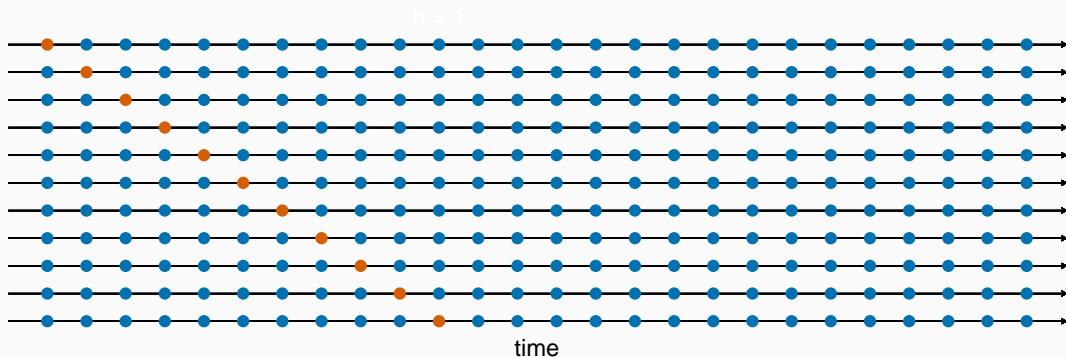


# Cross-validation

## Traditional evaluation



## Leave-one-out cross-validation

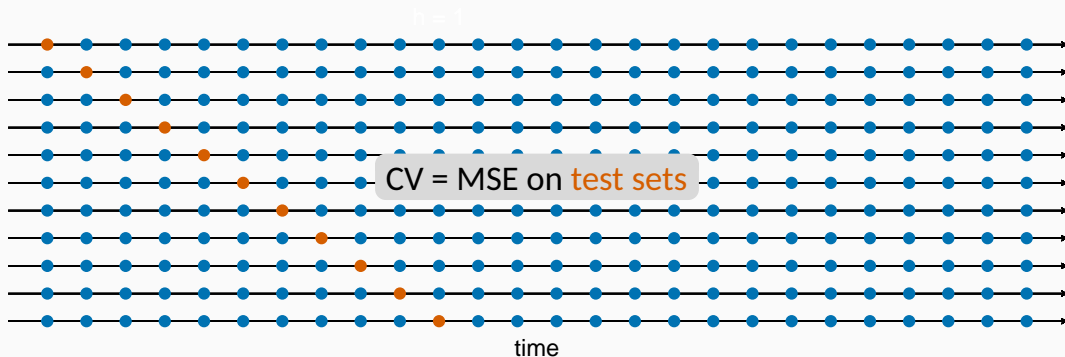


# Cross-validation

## Traditional evaluation



## Leave-one-out cross-validation



# Choosing regression variables

## Best subsets regression

- Fit all possible regression models using one or more of the predictors.
- Choose the best model based on one of the measures of predictive ability (CV, AIC, AICc).



# Choosing regression variables

## Best subsets regression

- Fit all possible regression models using one or more of the predictors.
- Choose the best model based on one of the measures of predictive ability (CV, AIC, AICc).

## Warning!

- If there are a large number of predictors, this is not possible.
- For example, 44 predictors leads to 18 trillion possible models!

# Choosing regression variables

## Backwards stepwise regression

- Start with a model containing all variables.
- Try subtracting one variable at a time. Keep the model if it has lower CV or AICc.
- Iterate until no further improvement.

# Choosing regression variables

## Backwards stepwise regression

- Start with a model containing all variables.
- Try subtracting one variable at a time. Keep the model if it has lower CV or AICc.
- Iterate until no further improvement.

## Notes

- Stepwise regression is not guaranteed to lead to the best possible model.
- Inference on coefficients of final model will be wrong.