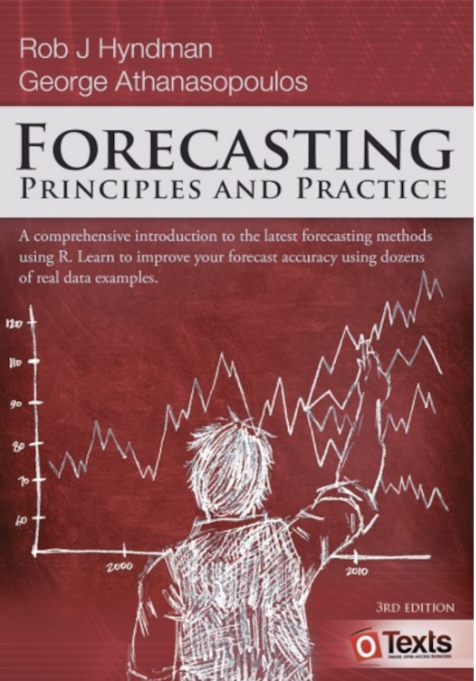

7. Time series regression models

7.8 Correlation, causation and forecasting

OTexts.org/fpp3/

Correlation is not causation

- When x is useful for predicting y, it is not necessarily causing y.
- e.g., predict number of swimmers y using number of ice-creams sold x.
- Correlations are useful for forecasting, even when there is no causality.
- Better models usually involve causal relationships (e.g., temperature x and people z to predict swimmers y).

Multicollinearity

In regression analysis, multicollinearity occurs when:

- Two predictors are highly correlated (i.e., the correlation between them is close to ± 1).
- A linear combination of some of the predictors is highly correlated with another predictor.
- A linear combination of one subset of predictors is highly correlated with a linear combination of another subset of predictors.

Multicollinearity

If multicollinearity exists…

- the numerical estimates of coefficients may be wrong (worse in Excel than in a statistics package)
- don't rely on the *p*-values to determine significance.
- there is no problem with model *predictions* provided the predictors used for forecasting are within the range used for fitting.
- omitting variables can help.
- combining variables can help.