Date: Thursday, November 29, 2018.

Multivariate Time Series

Until now, we considered only univariate time series methods, i.e., we've been forecasting a single variable using only its own history.

Now, imagine you have data on other variables collected for the same duration and at the same frequency as the response series. If these other variables are correlated with the response we may exploit that relationship by using multivariate time series models. This may result in more accurate forecasts.

Depending on how we wish to treat this external information determines which modeling approach to take:

- If we treat these variables as exogenous, i.e., they influence the response, and not the other way around, we can fit SARIMAX models to account for this relationship.
  ↑ exogenous

  Example: Daily BART ridership may depend on daily weather.

- If we treat these variables as endogenous, i.e., they influence the response and the response influences them, we can fit vector autoregression (VAR) models to simultaneously account for all of these dependencies.

  Example: Daily closing prices of AMZN and daily closing prices of APPL.

- If we wish to treat some variables as exogenous and some as endogenous, we can use VARX models.

SARIMAX Models

A SARIMAX model can be thought of as a SARIMA model plus explanatory variables. To begin, consider an ARMAX(p,q):

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} + \beta_1 X_{1,t} + \beta_2 X_{2,t} + \dots + \beta_r X_{r,t}$$

$$\boxed{= \sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t + \sum_{k=1}^{r} \beta_k X_{k,t}}$$

we can determine what X_r looks like using cross-correlation.

$\mathrm{Corr}(Y_t, X_{t+h}) \quad h \in \mathbb{Z}$

* Here, $X_{k,t}$ is an explanatory time series, $k = 1, 2, \dots, r$.
* If $\{Y_t\}$ is not stationary, we simply difference as is appropriate, and this differencing also gets applied to the exogenous variables. Depending on the type of differencing, this gives rise to ARIMAX and SARIMAX models.
* In order to forecast SARIMAX models, you need future values of the exogenous variables, or predicted values of them.

The main limitation of the SARIMAX model is the fact that it ignores the possibility that $\{Y_t\}$ may influence $\{X_{k,t}\}$, $k = 1, 2, \dots, r$. Accounting for direction of influence may provide more accurate forecasts.

Vector Autoregression

In this framework, all variables are treated symmetrically and so, if there are r variables under consideration, we denote them by $\{Y_{1,t}\}, \{Y_{2,t}\}, \dots, \{Y_{r,t}\}$. This methodology provides an equation for each variable, and the number of lags considered is referred to as the order p.

VAR(p):

$$y_{1,t} = c_1 + \sum_{i=1}^{p} \phi_{11,i}\, y_{1,t-i} + \sum_{i=1}^{p} \phi_{12,i}\, y_{2,t-i} + \cdots + \sum_{i=1}^{p} \phi_{1r,i}\, y_{r,t-i} + \varepsilon_{1,t}$$

$$\vdots$$

$$y_{r,t} = c_r + \sum_{i=1}^{p} \phi_{r1,i}\, y_{1,t-i} + \sum_{i=1}^{p} \phi_{r2,i}\, y_{2,t-i} + \cdots + \sum_{i=1}^{p} \phi_{rr,i}\, y_{r,t-i} + \varepsilon_{r,t}$$

where $\{\varepsilon_{k,t}\} \sim WN(0, \sigma_k^2)$ for $k = 1, 2, \ldots, r$.

Let $\vec{y}_t = [y_{1,t}, y_{2,t}, \ldots, y_{r,t}]^T$ ← r×1 vector, and

$$A_i = \begin{bmatrix} \phi_{11,i} & \phi_{12,i} & \cdots & \phi_{1r,i} \\ \vdots & & \ddots & \vdots \\ \phi_{r1,i} & & \cdots & \phi_{rr,i} \end{bmatrix}, \text{ and } i = 1, 2, \ldots, p$$

r×r matrix (not symmetric)

$\vec{\varepsilon}_i = [\varepsilon_{1,t}, \varepsilon_{2,t}, \ldots, \varepsilon_{r,t}]^T$ ← r×1 vector, and

$\vec{c}_i = [c_1, c_2, \ldots, c_r]^T$ ← r×1 vector.

Using this vector-matrix notation, the VAR(p) model can be more succintly written as:

$$\vec{y}_t = \vec{c} + A_1 \vec{y}_{t-1} + A_2 \vec{y}_{t-2} + \cdots + A_p \vec{y}_{t-p} + \vec{\varepsilon}_t .$$

Example: VAR(1) with two variables

$$\left.\begin{aligned} y_{1,t} &= c_1 + \phi_{11,1}\, y_{1,t-1} + \phi_{12,1}\, y_{2,t-1} + \varepsilon_{1,t} \\ y_{2,t} &= c_2 + \phi_{21,1}\, y_{1,t-1} + \phi_{22,1}\, y_{2,t-1} + \varepsilon_{2,t} \end{aligned}\right\} *$$

$$\vec{y}_t = \begin{bmatrix} y_{1,t} \\ y_{2,t} \end{bmatrix} \quad \vec{c} = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} \quad \vec{\varepsilon}_t = \begin{bmatrix} \varepsilon_{1,t} \\ \varepsilon_{2,t} \end{bmatrix}$$

$$A_1 = \begin{bmatrix} \phi_{11,1} & \phi_{12,1} \\ \phi_{21,1} & \phi_{22,1} \end{bmatrix}$$

Then, $\vec{y}_t = \vec{c} + A_1 \vec{y}_{t-1} + \vec{\varepsilon}_t$ yields exactly the same set of equations as defined in (*).

[Break]

Estimation of the parameters is carried out via Least Squares Minimization separately for each equation. In total, there are $2r + pr^2$ parameters which can become very complicated. Thus, choosing p and/or r to be small is preferrable.

Order selection is based on predictive accuracy and/or goodness-of-fit.

* A VARMA model exists but due to its complexity it is not widely used in practice.

(we don't have to do any differencing here, we are exploiting all those variables).

[R]

Cross-correlation plot : ccf.

ARIMA: xreg (external)

tsdiag → time series diagnostic.

Ljung-Box: $\boxed{\begin{array}{l} H_0: \rho(1) = \rho(2) = \rho(3) = \dots = \rho(H) = 0 \\ H_A: \rho(h) \neq 0 \text{ for some } 1 \leq h \leq H \end{array}}$

Stronger test than ACF plot.

$\boxed{\begin{array}{l} H_0: \rho(h) = 0 \\ H_A: \rho(h) \neq 0 \text{ , } h \geq 1. \end{array}}$

→ var(p)

no need to define p if we use lag.max = n