MULTIVARIATE TIME SERIES & FORECASTING
Multivariate Stationary Process

Suppose that the vector time series $Y_t = (y_{1t}, y_{2t}, \ldots, y_{mt})$ consists of $m$ univariate time series. Then $Y_t$ with finite first and second order moments is said to be weakly stationary if

(i) $E(Y_t) = E(Y_{t+s}) = \mu$, constant for all $s$
(ii) $\text{Cov}(Y_t) = E[(Y_t - \mu)(Y_t - \mu)'] = \Gamma(0)$
(iii) $\text{Cov}(Y_t, Y_{t+s}) = \Gamma(s)$ depends only on $s$
Vector ARMA models

\[ \Phi(B)Y_t = \delta + \Theta(B)\varepsilon_t, \]

where \( \Phi(B) = I - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p \), \( \Theta(B) = I - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_q B^q \).
\( \varepsilon_t \) represents the sequence of independent random vectors with \( E(\varepsilon_t) = 0 \) and \( \text{Cov}(\varepsilon_t) = \Sigma \).

Stationarity

if the roots of the equation

\[ \det(\Phi(B)) = \det(I - \Phi_1 B - \Phi_2 B^2 - \cdots - \Phi_p B^p) = 0 \]

are all greater than 1 in absolute value

Then: infinite MA representation

\[ Y_t = \mu + \Psi(B)\varepsilon_t \]
\[ = \mu + \sum_{i=0}^{\infty} \Psi_i \varepsilon_{t-i} \]

where \( \Psi(B) = \Phi(B)^{-1}\Theta(B) \), \( \mu = \Phi(B)^{-1}\delta \), and \( \sum_{i=0}^{\infty} \|\Psi_i\|^2 < \infty \).
Invertibility

if the roots of the equation

\[ \text{det}(\Theta(B)) = \text{det}(I - \Theta_1 B - \Theta_2 B^2 - \cdots - \Theta_q B^q) = 0 \]

are all greater than 1 in absolute value

\[
\begin{align*}
\text{ARMA}(1,1) \\
y_{1,t} &= \delta_1 + \phi_{11}y_{1,t-1} + \phi_{12}y_{2,t-1} + \varepsilon_{1,t} - \theta_{11}\varepsilon_{1,t-1} - \theta_{12}\varepsilon_{2,t-1} \\
y_{2,t} &= \delta_2 + \phi_{21}y_{1,t-1} + \phi_{22}y_{2,t-1} + \varepsilon_{2,t} - \theta_{21}\varepsilon_{1,t-1} - \theta_{22}\varepsilon_{2,t-1}
\end{align*}
\]

\[ \Phi(B) = I - \Phi_1 B \]

\[ \Theta(B) = I - \Theta_1 B \]
If case of non stationarity: apply differencing of appropriate degree

\[ \Phi(B)D(B)Y_t = \delta + \Theta(B)\epsilon_t \]

\[ D(B) = \text{diag}\{(1 - B)^{d_1}, (1 - B)^{d_2}, \ldots, (1 - B)^{d_m}\} \]

\[
\begin{bmatrix}
(1 - B)^{d_1} & 0 & \cdots & 0 \\
0 & (1 - B)^{d_2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & (1 - B)^{d_m}
\end{bmatrix}
\]
VAR models (vector autoregressive models) are used for multivariate time series. The structure is that each variable is a linear function of past lags of itself and past lags of the other variables.

\[ X_{t,1} = \alpha_1 + \Phi_{11} X_{t-1,1} + \Phi_{12} X_{t-1,2} + \Phi_{13} X_{t-1,3} + \epsilon_{t,1} \]

\[ X_{t,2} = \alpha_2 + \Phi_{21} X_{t-1,1} + \Phi_{22} X_{t-1,2} + \Phi_{23} X_{t-1,3} + \epsilon_{t,2} \]

\[ X_{t,3} = \alpha_3 + \Phi_{31} X_{t-1,1} + \Phi_{32} X_{t-1,2} + \Phi_{33} X_{t-1,3} + \epsilon_{t,3} \]

Each variable is a linear function of the lag 1 values for all variables in the set.

In a VAR(2) model, the lag 2 values for all variables are added to the right sides of the equations. In the case of three x-variables there would be six variables on the right side of each equation, three lag 1 variables and three lag 2 variables.

In general, for a VAR(p) model, the first p lags of each variable in the system would be used as regression predictors for each variable.

VAR models are a specific case of more general VARMA models. VARMA models for multivariate time series include the VAR structure above along with moving average terms for each variable. More generally yet, these are special cases of ARMAX models which allow for the addition of other predictors that are outside the multivariate set of principal interest.
Vector AR (VAR) models

Vector AR(p) model

\[ \Phi(B)Y_t = \delta + \epsilon_t \]

\[ Y_t = \delta + \sum_{i=1}^{p} \Phi_i Y_{t-i} + \epsilon_t \]

For a stationary vector AR process: infinite MA representation

\[ Y_t = \mu + \Psi(B)\epsilon_t \]

\[ \Psi(B) = I + \Psi_1 B + \Psi_2 B^2 + ... \]

\[ \mu = \Phi(B)^{-1} \delta \]
\[ E(Y_t) = \]
\[ \text{Cov}(t, Y_{t-s}) = 0, \text{ for any } s > 0 \]
\[ \text{Cov}(t, Y_t) = \text{Cov}(t, t) = \]
\[ (s) = \text{Cov}(Y_{t-s}, Y_t) = \sum_{i=1}^{p} \text{Cov}(Y_{t-s}, Y_{t-i}) \]
\[ (0) = \sum_{i=1}^{p} (i) + \]

The Yule-Walker equations can be obtained from the first \( p \) equations
The autocorrelation matrix of \( \text{Var}(p) : \) decaying behavior following a mixture of exponential decay & damped sinusoid
VAR(1) Model

Autocovariance matrix

$$\Gamma(s) = \Gamma(s - 1)\Phi' = (\Gamma(s - 2)\Phi')\Phi' = \cdots = \Gamma(0)(\Phi')'$$

$$\rho(s) = V^{-1/2}\Gamma(s)V^{-1/2}$$
$$= V^{-1/2}\Gamma(0)(\Phi')'V^{-1/2}$$
$$= \rho(0)\nu^{1/2}(\Phi')'V^{-1/2}$$

$V$: diagonal matrix

The eigenvalues of $\Phi$ determine the behavior of the autocorrelation matrix.
1. Data: the pressure readings at two ends of an industrial furnace
   Expected: individual time series to be autocorrelated & cross-correlated
   Fit a multivariate time series model to the data

2. Identify model
   - Sample ACF plots
   - Cross correlation of the time series
Exponential decay pattern: autoregressive model & VAR(1) or VAR(2)
Or ARIMA model to individual time series & take into consideration the cross correlation of the residuals

Time series plots of the pressure readings at both ends of the furnace.

The sample ACF plot for (a) the pressure readings at the front end of the furnace, $y_1$; (b) the pressure readings at the back end of the furnace, $y_2$; (c) the cross correlation between $y_1$ and $y_2$; and (d) the cross correlation between the residuals from the AR(1) model for front pressure and the residuals from the AR(1) model for back pressure.
VAR(1) provided a good fit.

Time series plots of the residuals from the VAR(1) model.

Actual and fitted values for the pressure readings at the front end of the furnace.
Actual and fitted values for the pressure readings at the back end of the furnace.
ARCH/GARCH Models (autoregressive conditionally heteroscedastic)
-a model for the variance of a time series.
-used to describe a changing, possibly volatile variance.
-most often it is used in situations in which there may be short periods of increased variation. (Gradually increasing variance connected to a gradually increasing mean level might be better handled by transforming the variable.)

Identifying an ARCH/GARCH Model in Practice

The best identification tool may be a time series plot of the series. It’s usually easy to spot periods of increased variation sprinkled through the series. It can be fruitful to look at the ACF and PACF of both $y_t$ and $y_t^2$. For instance, if $y_t$ appears to be white noise and $y_t^2$ appears to be AR(1), then an ARCH(1) model for the variance is suggested. If the PACF of the $y_t^2$ suggests AR(m), then ARCH(m) may work.
Not constant variance

Consider AR(p)= model

\[ y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \epsilon_t \]

Errors: uncorrelated, zero mean noise with changing variance

Model \( e_t^2 \) as an AR(l) process

\[ e_t^2 = \xi_0 + \xi_1 e_{t-1}^2 + \xi_2 e_{t-2}^2 + \cdots + \xi_l e_{t-l}^2 + \alpha_t \]

\( \alpha_t \), white noise with zero mean & constant variance

\( e_t \): Autoregressive conditional heteroscedastic process of order l – ARCH(l)
Generalise ARCH model

Consider the error:

\[ e_t = \sqrt{v_t} w_t \]

where \( w_t \) is independent and identically distributed with mean 0 and variance 1, and

\[ v_t = \zeta_0 + \zeta_1 e_{t-1}^2 + \zeta_2 e_{t-2}^2 + \cdots + \zeta_l e_{t-l}^2 \]

conditional variance of \( e_t \)

\[
\text{Var}(e_t | e_{t-1}, \ldots) = E(e_t^2 | e_{t-1}^2, \ldots) \\
= v_t \\
= \zeta_0 + \zeta_1 e_{t-1}^2 + \zeta_2 e_{t-2}^2 + \cdots + \zeta_l e_{t-l}^2
\]

We can also argue that the current conditional variance should also depend on the previous conditional variances as

\[
v_t = \zeta_0 + \zeta_1 v_{t-1} + \zeta_2 v_{t-2} + \cdots + \zeta_k v_{t-k} + \zeta_1 e_{t-1}^2 + \zeta_2 e_{t-2}^2 + \cdots + \zeta_l e_{t-l}^2
\]

e_t: Generalised Autoregressive conditional heteroscedastic process of order k and l–GARCH(k,l)
S&P index

- Initial data: non stationary
- Log transformation of the data
- First differences of the log data

Mean stable
Changes in the variance

No autocorrelation left in the data
ACF & PACF of the squared differences: ARCH (3) model

ACF and PACF plots of the square of the first difference of the log transformation of the weekly close for S&P500 Index from 1995 to 1998.