Stat 443. Time Series and Forecasting.

Key topic: Some results on conditions for AR time series to be stationary; the conditions extend to ARMA.

Introduction to some tools for studying properties of ARMA

- Covariances of linear combinations (included in this set of slides for your review).
- Recursion equations from stochastic representation
- Equations for serial correlations with different lags
- Equations for serial conditional correlations (later)

Review for covariances

(a) If X, Y are random variables,
$\text{Cov}\,(X, Y) = \text{Cov}\,(Y, X) = \text{E}\,[(X - \mu_X)(Y - \mu_Y)] = \text{E}\,[(X - \mu_X)Y]$.

(b) X is a random variable,
$L(\cdot) = \text{Cov}\,(X, \cdot) = \text{E}\,[(X - \mu_X)\cdot]$ is a linear operator over random variables; that is: if $c_1, \ldots, c_m$ are reals and $Y_1, \ldots, Y_m$ are random variables,

$$\text{Cov}\,(X, c_1 Y_1 + \cdots + c_m Y_m) = \sum_{j=1}^{m} c_j \text{Cov}\,(X, Y_j)$$

(c) By symmetry of $\text{Cov}\,(\cdot, \cdot)$, then $\text{Cov}\,(\cdot, X)$ is a linear operator.

(d) $\text{Var}\,(X) = \text{Cov}\,(X, X)$: use the bilinear property of Cov to derive $\text{Var}\,(c_1 Y_1 + c_2 Y_2)$ when $X = c_1 Y_1 + c_2 Y_2$:

$$
\begin{aligned}
X &= c_1Y_1 + c_2Y_2 \\
&\text{Cov}\,(X, c_1Y_1 + c_2Y_2) = c_1\text{Cov}\,(X, Y_1) + c_2\text{Cov}\,(X, Y_2) \\
&= c_1\text{Cov}\,(c_1Y_1 + c_2Y_2, Y_1) + c_2\text{Cov}\,(c_1Y_1 + c_2Y_2, Y_2) \\
&= c_1^2\text{Cov}\,(Y_1, Y_1) + c_1c_2\text{Cov}\,(Y_2, Y_1) + c_2c_1\text{Cov}\,(Y_1, Y_2) + c_2^2\text{Cov}\,(Y_2, Y_2) \\
&= c_1^2\text{Var}\,(Y_1) + c_2^2\text{Var}\,(Y_2) + 2c_1c_2\text{Cov}\,(Y_1, Y_2) \\
&= \text{Var}\,(c_1Y_1 + c_2Y_2)
\end{aligned}
$$

Extend to the result for covariance of linear combinations of random variables:

$$
\text{Cov}\,(\sum_{i=1}^{m} c_iX_i, \sum_{j=1}^{n} d_jY_j) = \sum_{i=1}^{m}\sum_{j=1}^{n} c_id_j\text{Cov}\,(X_i, Y_j).
$$

From this, get the variance of a single linear combination.

Next slides:

Derivation of conditions on the AR coefficients $\phi_1, \phi_2, \ldots$ in order that the AR(p) time series is stationary

Idea: Look at AR(1) and AR(2), from which general results for AR(p) can be conjectured/obtained.

AR(1): let $\tilde{Y}_t = Y_t - \mu$. $\tilde{Y}_i = \phi \tilde{Y}_{i-1} + \epsilon_i$. Then

$$
\begin{aligned}
\tilde{Y}_t &= \phi \tilde{Y}_{t-1} + \epsilon_t = \phi(\phi \tilde{Y}_{t-2} + \epsilon_{t-1}) + \epsilon_t \\
&= \phi^2 \tilde{Y}_{t-2} + \phi \epsilon_{t-1} + \epsilon_t = \phi^2(\phi \tilde{Y}_{t-3} + \epsilon_{t-2}) + \phi \epsilon_{t-1} + \epsilon_t \\
&= \phi^3 \tilde{Y}_{t-3} + \phi^2 \epsilon_{t-2} + \phi \epsilon_{t-1} + \epsilon_t \\
&= \phi^k \tilde{Y}_{t-k} + \sum_{j=0}^{k-1} \phi^j \epsilon_{t-j} \\
\tilde{Y}_{i+k} &= \phi^k \tilde{Y}_i + \sum_{j=0}^{k-1} \phi^j \epsilon_{i+k-j} \text{ shifting subscript index}
\end{aligned}
$$

If $|\phi| > 1$ and $\tilde{Y}_i$ becomes much larger than 1 for some i, then the time series grows exponentially fast in absolute value.

If $|\phi| > 1$, the time series is not stationary.
If $\phi = \pm 1$, the variance of $\tilde{Y}_{i+k}$ increases with k for fixed i, so the time series is not stationary.
Hence $|\phi| < 1$ is a necessary condition for (weak) stationarity.

AR(1) acf in the case of stationarity

AR(1): let $\tilde{Y}_t = Y_t - \mu$. $\tilde{Y}_i = \phi \tilde{Y}_{i-1} + \epsilon_i$.

$$\tilde{Y}_{i+k} = \phi^k \tilde{Y}_i + \sum_{j=0}^{k-1} \phi^j \epsilon_{i+k-j}$$

If time series is stationary, let $\sigma_Y^2 = \text{Var}\,(Y_t) = \text{Var}\,(\tilde{Y}_t)$. Apply covariance with Y_i to the above equation.

$$\begin{aligned} \gamma_k &:= \text{Cov}\,(\tilde{Y}_{i+k}, \tilde{Y}_i) = \phi^k \text{Cov}\,(\tilde{Y}_i, \tilde{Y}_i) + \sum_{j=0}^{k-1} \phi^j \text{Cov}\,(\epsilon_{i+k-j}, \tilde{Y}_i) \\ &= \phi^k \sigma_Y^2 + 0 \\ \rho_k &= \text{Cor}\,(\tilde{Y}_{i+k}, \tilde{Y}_i) = \gamma_k / \sigma_Y^2 = \phi^k \end{aligned}$$

Since a correlation is between -1 and 1, this is consistent with the necessary condition for weak stationarity that $|\phi| < 1$.

AR(2): let $\tilde{Y}_t = Y_t - \mu$. If stationary, then

$$
\begin{aligned}
\tilde{Y}_t &= \phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \epsilon_t \\
\sigma_Y^2 &= \phi_1^2 \sigma_Y^2 + \phi_2^2 \sigma_Y^2 + 2\phi_1\phi_2\rho_1\sigma_Y^2 + \sigma_\epsilon^2 \\
\sigma_\epsilon^2 &= (1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2\rho_1)\sigma_Y^2 > 0 \\
0 &< (1 - \phi_1^2 - \phi_2^2 - 2\phi_1\phi_2\rho_1) \\
\mathrm{Cov}\,(\tilde{Y}_t, \tilde{Y}_{t-1}) &= \phi_1 \mathrm{Cov}\,(\tilde{Y}_{t-1}, \tilde{Y}_{t-1}) + \phi_2 \mathrm{Cov}\,(\tilde{Y}_{t-2}, \tilde{Y}_{t-1}) + \mathrm{Cov}\,(\epsilon_t, \tilde{Y}_{t-1}) \\
\rho_1 \sigma_Y^2 &= \phi_1 \sigma_Y^2 + \phi_2 \rho_1 \sigma_Y^2 + 0 \\
\rho_1 &= \phi_1/(1 - \phi_2) \in (-1, 1) \\
\mathrm{Cov}\,(\tilde{Y}_t, \tilde{Y}_{t-2}) &= \phi_1 \mathrm{Cov}\,(\tilde{Y}_{t-1}, \tilde{Y}_{t-2}) + \phi_2 \mathrm{Cov}\,(\tilde{Y}_{t-2}, \tilde{Y}_{t-2}) + \mathrm{Cov}\,(\epsilon_t, \tilde{Y}_{t-2}) \\
\rho_2 \sigma_Y^2 &= \phi_1 \rho_1 \sigma_Y^2 + \phi_2 \sigma_Y^2 + 0
\end{aligned}
$$

Examples of equations involving autocovariances/autocorrelations

If $\mathrm{E}\,(\tilde{Y}_t) = 0$ for all t, then it looks like expected values can be 0 on both sides. The above equations come from

$$\mathrm{Var}\,(\tilde{Y}_t) = \mathrm{Var}\,(\phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \epsilon_t)$$

$$\mathrm{Cov}\,(\tilde{Y}_t, \tilde{Y}_{t-1}) = \mathrm{Cov}\,(\phi_1 \tilde{Y}_{t-1} + \phi_2 \tilde{Y}_{t-2} + \epsilon_t, \tilde{Y}_{t-1})$$

One condition for weak stationarity of AR(2) is that $|\phi_2| < 1$, but $|\phi_1|$ can be > 1. A proof of the ϕ_2 condition will be given later.
AR(p) extension: one condition for stationarity is that $|\phi_p| < 1$.

AR(2): if $|\phi_2| < 1$ and $\rho_1 = \phi_1/(1-\phi_2) \in (-1, 1)$, and then

$$
\begin{aligned}
&-1 < \phi_1/(1-\phi_2) < 1 \\
&\phi_2 - 1 < \phi_1 < 1 - \phi_2 \\
&\phi_1 + \phi_2 < 1 \\
&\phi_2 - \phi_1 < 1
\end{aligned}
$$

The three equations in blue define the triangular region of the (ϕ_1, ϕ_2) space for AR(2) to be stationary.

Exercise: draw the region (to be shown later).

Pseudo-code for AR(p) forecast rmse; R function: `arp_fc = function(train,holdout,arvec,mu)`
Forecast is linear in the previous p observations (assuming stationary):

- input `train` with size n_{train}, `holdout` with size $n_{holdout}$, estimates $\widehat{\boldsymbol{\phi}} = (\widehat{\phi}_1, \ldots, \widehat{\phi}_p)^T$ and $\widehat{\mu}$.
- Join last p observations of `train` and the vector `holdout` to get a vector $\mathbf{z}$ of length $p + n_{holdout}$. Subtract $\widehat{\mu}$ from each element: $\mathbf{z} \leftarrow \mathbf{z} - \widehat{\mu}\mathbf{1}$.
- sse$\leftarrow$ 0
- for i in $1, \ldots, n_{holdout}$:
- $\mathbf{zprev} \leftarrow (z_{i+p-1}, \ldots, z_{i+1}, z_i)^T$, fc$\leftarrow$ $\widehat{\mu} + \widehat{\boldsymbol{\phi}}^T \mathbf{zprev}$ (same as $\widehat{\mu} + \sum_{j=1}^{p} \widehat{\phi}_j (y_{n_{train}+i-j} - \widehat{\mu})$);
- fcvec[i]$\leftarrow$ fc; yt $\leftarrow$holdout[i];
- fcerror$\leftarrow$ yt-fc; sse$\leftarrow$ sse$+$ fcerror^2.
- end for; return rmse=sqrt(sse/$n_{holdout}$) and fcvec

Later to check if holdout set moving 1-step forecasts and rmse can be obtain via another call to arima with appropriate inputs

Summary

Start of tools for studying dependence properties of ARMA and ARIMA.

- Recursion equations from stochastic representation (slide 5)
- Equations for serial correlations with different lags (slide 7)
- Equations for serial conditional correlations (later)