Stat 443. Time Series and Forecasting.

Key ideas: autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA), stationary time series, variogram.

$\{\epsilon_t\}$ is a white noise innovation sequence with mean 0, variance $\sigma_\epsilon^2 > 0$, in all stochastic models in this set of slides.

Autoregressive (AR) of order p or AR(p): regression on previous p.
AR(1): $Y_t - \mu = \phi(Y_{t-1} - \mu) + \epsilon_t$.
AR(2): $Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \phi_2(Y_{t-2} - \mu) + \epsilon_t$.
AR(p): $Y_t - \mu = \sum_{i=1}^{p} \phi_j(Y_{t-i} - \mu) + \epsilon_t$.
Here ϵ_t is an innovation random variable, independent of the past $(Y_{t-1}, Y_{t-2}, \ldots)$, and $\{\epsilon_t\}$ is iid with mean 0, variance σ_ϵ^2.

Moving average (MA) of order q or MA(q).
For MA(q), the moving average is actually weighted sum of q (time-shifting) consecutive terms of a white noise sequence.
Let $\{\epsilon_t\}$ be a white noise sequence with mean 0, variance σ_ϵ^2.
MA(1): $Y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$, $Y_{t-1} = \mu + \epsilon_{t-1} + \theta\epsilon_{t-2}$.
MA(2): $Y_t = \mu + \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2}$.
MA(q): $Y_t = \mu + \epsilon_t + \sum_{j=1}^{q} \theta_j\epsilon_{t-j}$ (notation in R). Some books use notation $Y_t = \mu + \epsilon_t - \sum_{j=1}^{q} \theta_j\epsilon_{t-j}$.
Is ϵ_t independent of $(Y_{t-1}, Y_{t-2}, \ldots)$?

autoregressive moving-average ARMA(p, q), non-negative integers p, q; $\phi_p \neq 0$, $\theta_q \neq 0$.

$$Y_t = \mu + \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p(Y_{t-p} - \mu) + \epsilon_t + \theta_1\epsilon_{t-1} + \cdots + \theta_q\epsilon_{t-q}.$$

$$Y_t - \mu = \sum_{i=1}^{p} \phi_i(Y_{t-i} - \mu) + \epsilon_t + \sum_{j=1}^{q} \theta_j\epsilon_{t-j}.$$

Take a null sum $\sum_{i=1}^{0} x_i$ or $\sum_{j=1}^{0} x_j$ as 0.

If $q = 0$, ARMA(p, q) is same as AR(p), p positive integer.

If $p = 0$, ARMA(p, q) is same as MA(q), q positive integer.

If $p = q = 0$, ARMA(p, q), $Y_t = \mu + \epsilon_t$.

autoregressive integrated moving-average ARIMA(p, d, q), non-negative integers p, q; $d \in \{0, 1, 2\}$.

$d = 0$: no differencing
$d = 1$: one differencing operation
$d = 2$: two differencing operations

ARIMA($p, 0, q$) is ARMA(p, q).
ARIMA($p, 1, q$) is ARMA(p, q) after first difference
ARIMA($p, 2, q$) is ARMA(p, q) after second difference

Exponential smoothing and ARMA

Simple exponential smoothing: differenced series is special case of MA(1) with restricted range on coefficient of ϵ_{t-1}

$$Y_t - Y_{t-1} = \epsilon_t - \theta\epsilon_{t-1}, \quad 0 < \theta < 1.$$

Holt linear exponential smoothing: twice differenced series is MA(2) or ARIMA(0,2,2) with restrictions on coefficients
Damped linear trend with $\phi \in (0,1)$: (once) differenced series $W_t = Y_t - Y_{t-1}$ is ARMA(1,2) or ARMA(1,1,2)

$$(Y_t - Y_{t-1}) - \phi(Y_{t-1} - Y_{t-2}) = \epsilon_t + \theta_1\epsilon_{t-2} + \theta_2\epsilon_{t-2}$$

Winters additive seasonal, periodicity c: difference at lag c of differenced series is MA$(c+1)$ with restrictions on coefficients

So exponential smoothing methods provide explanations for use of differencing in time series modelling. Also, differencing is sometimes useful for assessing leading variables in multiple time series.

Stationary sequence of random variables

Definition: Strictly stationary. A sequence $Y_1, Y_2, \ldots$ is strictly stationary (invariance to tome shifts) if random vectors $(Y_t, \ldots, Y_{t+j})$ and $(Y_{t+h}, \ldots, Y_{t+j+h})$ have the same joint distribution for $t = 1, 2, \ldots$, $j = 0, 1, 2, \ldots$, $h = 1, 2, \ldots$. By marginalization, this implies that for $1 \le t_1 < t_2 < \ldots < t_m$ and $m = 2, \ldots$,

$$(Y_{t_1}, Y_{t_2}, \ldots, Y_{t_m}) \stackrel{d}{=} (Y_{t_1+h}, Y_{t_2+h}, \ldots, Y_{t_m+h}), \;\; h = 1, 2, \ldots, \qquad (1)$$

where $\stackrel{d}{=}$ is the symbol for equal in distribution.

Remarks. 1. Take $j = 0$ (or $m = 1$ in (1)), to get that Y_t and Y_{t+h} have the same distribution for $t = 1, 2, \ldots$, $h = 1, 2, \ldots$. There is a common cumulative distribution function F_Y for the Y_t's. Then we can write $\mu_Y = \mathrm{E}\,(Y_t)$ and $\sigma_Y^2 = \mathrm{Var}\,(Y_t)$ for all t.

2. Take $j = 1$, to get that (Y_t, Y_{t+1}) and (Y_{t+h}, Y_{t+h+1}) have the same distribution for $t = 1, 2, \ldots$, $h = 1, 2, \ldots$. Then we can write $\gamma_1 = \mathrm{Cov}\,(Y_t, Y_{t+1}) = \mathrm{Cov}\,(Y_{t+h}, Y_{t+h+1})$, and $\rho_1 = \rho(1) = \mathrm{Cor}\,(Y_t, Y_{t+1}) = \mathrm{Cor}\,(Y_{t+h}, Y_{t+h+1})$ for the lag 1 serial correlation.

3. In (1), take $m = 2$, $t_1 = t$, $t_2 = t + k$ where k is a positive integer. Then $(Y_t, Y_{t+k}) \overset{d}{=} (Y_{t+h}, Y_{t+k+h})$ and we can write $\gamma_k = \text{Cov}\,(Y_t, Y_{t+k}) = \text{Cov}\,(Y_{t+h}, Y_{t+h+k})$, and $\rho_k = \rho(k) = \text{Cor}\,(Y_t, Y_{t+k}) = \text{Cor}\,(Y_{t+h}, Y_{t+h+k})$ for the lag k serial correlation.

Definition: Weakly stationary. A sequence $Y_1, Y_2, \ldots$ is weakly stationary or second order stationary if $\text{E}\,(Y_t)$, $\text{Var}\,(Y_t)$ do not depend on t, and $\text{Cov}\,(Y_t, Y_{t+k})$ is a function of k, independent of t.

Autocorrelation is only well-defined for a stationary sequence.

Otherwise $\overline{y}$ is not estimating a "population mean" and averaging the products $(y_1 - \overline{y})(y_2 - \overline{y})$, $(y_2 - \overline{y})(y_3 - \overline{y}), \ldots, (y_{n-1} - \overline{y})(y_n - \overline{y})$ might not have meaning.

Data sets: which can be considered as stationary? Use a combination of context of data and time series plot of data to determine if an assumption of stationarity is plausible.

1. Vancouver monthly average temperature. How about deseasonalized series?

2. Economic time series (for example, monthly unemployment rate, monthly CPI)

3. Financial returns: for example, $y_t = \log(p_t/p_{t-1})$ is log return for the S&P market index p_t.

4. Daily water flow of a river at a measuring station. Yearly maximum of daily water flow at same station.

Stochastic models

Assume that $\{\epsilon_t\}$ have a constant positive variance.

AR(2): (weak) stationary or not, or maybe need some condition for stationarity?

MA(1): (weak) stationary or not

MA(2): (weak) stationary or not

Is $\{Y_t\}$ stationary if $\{W_t = Y_t - Y_{t-1}\}$ is MA(1), such as $W_t = \epsilon_t + \theta\epsilon_{t-1}$.

Variogram, pp 68–69 of Bisgaard and Kulahci (2011). *Time Series Analysis and Forecasting by Example*, Wiley. In this book, the variogram is presented as a method to assess stationarity for a numerical time series.

A variogram doesn't assume stationarity, but assumes that the distribution of $Y_{t+k} - Y_t$ does not depend on t for any positive integer k. Hence it implies that if $\{Y_{t+1} - Y_t\}$ is stationary even if $\{Y_t\}$ is non-stationary.

$\{Y_t\}$ stationary implies $\{Y_{t+1} - Y_t\}$ is stationary implies

$$(Y_t, Y_{t+k}) \stackrel{d}{=} (Y_1, Y_{1+k}) \Rightarrow Y_{t+k} - Y_t \stackrel{d}{=} Y_{1+k} - Y_1 \ \forall t$$

$$(Y_{t_1}, Y_{t_1+k}, \ldots, Y_{t_m}, Y_{t_m+k}) \stackrel{d}{=} (Y_1, Y_{1+k}, \ldots, Y_{t_m-t_1+1}, Y_{t_m-t_1+1+k}) \Rightarrow$$

$$(Y_{t_1+k} - Y_{t_1}, \ldots, Y_{t_m+k} - Y_{t_m}) \stackrel{d}{=} (Y_{1+k} - Y_1, \ldots, Y_{t_m-t_1+1+k} - Y_{t_m-t_1+1})$$
$$\forall t_1 < \ldots < t_m$$

Example: Let $Y_t = Y_{t-1} + \epsilon_t + \theta\epsilon_{t-1}$ for all t.
1. $\{Y_t\}$ is non-stationary:

$$Y_{t-1} = Y_{t-2} + \epsilon_{t-1} + \theta\epsilon_{t-2}$$
$$\mathrm{Cov}\,(Y_{t-1}, \epsilon_{t-1}) = \mathrm{Cov}\,(Y_{t-2}, \epsilon_{t-1}) + \mathrm{Cov}\,(\epsilon_{t-1}, \epsilon_{t-1}) + \theta\mathrm{Cov}\,(\epsilon_{t-2}, \epsilon_{t-1}) = 0 + \sigma_\epsilon^2 + 0$$
$$\mathrm{Var}\,(Y_t) = \mathrm{Var}\,(Y_{t-1}) + \mathrm{Var}\,(\epsilon_t) + \theta^2\mathrm{Var}\,(\epsilon_{t-1}) + 2\theta\mathrm{Cov}\,(Y_{t-1}, \epsilon_{t-1})$$
$$= \mathrm{Var}\,(Y_{t-1}) + (1 + \theta^2 + 2\theta)\sigma_\epsilon^2 = \mathrm{Var}\,(Y_{t-1}) + (1+\theta)^2\sigma_\epsilon^2 \text{ increasing in } t.$$

2. $Y_{t+1} - Y_t = \epsilon_{t+1} + \theta\epsilon_t$ has distribution not depending on t.
$Y_{t+2} - Y_t = Y_{t+2} - Y_{t+1} + Y_{t+1} - Y_t = \epsilon_{t+2} + (\theta + 1)\epsilon_{t+1} + \theta\epsilon_t$ has distribution that doesn't depend on t.
For integer $k \geq 3$, $Y_{t+k} - Y_t = \sum_{i=1}^{k}(\epsilon_{t+k} - \theta\epsilon_{t+k-1}) = \epsilon_{t+k} + \sum_{i=1}^{k-1}(\theta + 1)\epsilon_{t+i} + \theta\epsilon_t$ has distribution that doesn't depend on t.

Variogram G_k, $k \geq 1$

Def: $G_k = \text{Var}\,(Y_{t+k} - Y_t)/\text{Var}\,(Y_{t+1} - Y_t)$ (not depending on t) for $k = 1, 2, \ldots$ so that $G_1 = 1$.

For a stationary time series,

$$\begin{aligned} \text{Var}\,(Y_{t+k} - Y_t) &= \text{Var}\,(Y_{t+k}) + \text{Var}\,(Y_t) - 2\text{Cov}\,(Y_{t+k}, Y_t) \\ &= 2\text{Var}\,(Y_t) - 2\text{Var}\,(Y_t)\text{Cor}\,(Y_{t+k}, Y_t) \\ &= 2\sigma_Y^2 - 2\sigma_Y^2\rho_k = 2\sigma_Y^2(1 - \rho_k) \end{aligned}$$

and $G_k = (1 - \rho_k)/(1 - \rho_1)$.

If close to white noise, then $G_k \approx 1$.

For a stationary ARMA process where $|\rho_k|$ is eventually geometrically decreasing to 0, then $G_k \to 1/(1-\rho_1)$ as $k \to \infty$ (asymptote to a constant).

For a process that is MA(1) after differencing,

$$
\begin{aligned}
\text{Var}\,(Y_{t+k} - Y_t) &= \text{Var}\,(\epsilon_{t+k}) + \sum_{i=1}^{k-1} (\theta + 1)^2 \text{Var}\,(\epsilon_{t+i}) + \theta^2 \text{Var}\,(\epsilon_t) \\
&= \sigma_\epsilon^2 [1 + (k-1)(1+\theta)^2 + \theta^2], \quad k \geq 2; \\
\text{Var}\,(Y_{t+1} - Y_t) &= \sigma_\epsilon^2 (1 + \theta^2), \\
G_k &= \frac{[1 + (k-1)(1+\theta)^2 + \theta^2]}{(1+\theta^2)} = O(k), \quad k \to \infty
\end{aligned}
$$

The idea is that $\{G_k\}$ has a different pattern for a stationary sequence and one that is stationary after differencing.

Sample version $\widehat{G}_k$: for training set of size n

Var $(Y_{t+k} - Y_t)$ is estimated by

$$s_{d_k}^2 = \sum_{i=1}^{n-k} (y_{i+k} - y_i - \bar{d}_k)^2/(n-k-1), \quad \bar{d}_k = (n-k)^{-1} \sum_{i=1}^{n-k} (y_{i+k} - y_i).$$

Then

$$\widehat{G}_k = s_{d_k}^2 / s_{d_1}^2.$$

Compare this sample versions

$$\widehat{H}_k = (1 - \widehat{\rho}_k)/(1 - \widehat{\rho}_1).$$

These two should be similar if the data time series is (close to) stationary; see examples generated from variogram-examples.Rmd file.

Some examples in variogram-examples.pdf (variogram-examples.Rmd).

An R function `variogram` that outputs G and H is included in the Rmd file. The plots use a plotting symbol 1 for $\widehat{G}_k$ and symbol 2 for $\widehat{H}_k$.

The function is applied to: Vancouver monthly total precipitation, returns of corporate stocks, quarterly unemployment rate, monthly CPI, `sunspots` (R data set for monthly counts), `sunspots.year` (R data set for annual averages), some simulated AR(1) time series, one simulated AR(2) time series.

The plots of $\widehat{G}_k$ and $\widehat{H}_k$ should be approximately the same if the data are a realization of a stationary time series.

If $\widehat{G}_k$ and $\widehat{H}_k$ are similar, then the suggestion is that (a) the time series data are a realization of a stationary time series, or (b) the time series data can be modelled by a stationary time series model.
Distinction to be made later.

If $\widehat{G}_k$ and $\widehat{H}_k$ are quite different, then the suggestion is that the time series data are a realization of a non-stationary time series.

From stochastic model MA(1) to forecast rule.
Suppose $Z_t = Y_t - \mu = \epsilon_t + \theta\epsilon_{t-1}$ for all t, where $-1 < \theta < 1$. How to write $Y_{t+1} = g_t(Y_1, \ldots, Y_t; \theta) + \epsilon_t$?

Let B be the backward shift operator defined as $Bx_t = x_{t-1}$ for any variable x. As preview of later analysis of ARMA time series,

$$
\begin{aligned}
&Z_t = \epsilon_t + \theta B\epsilon_t = (1+\theta B)\epsilon_t \quad (*)\\
&Z_{t+1} = (1+\theta B)\epsilon_{t+1}\\
&(1+\theta B)^{-1} Z_{t+1} = \epsilon_{t+1}\\
&(1 - \theta B + \theta^2 B^2 - \theta^3 B^3 + \cdots) Z_{t+1} = \epsilon_{t+1} \quad (-1 < \theta < 1)\\
&Z_{t+1} - \theta B Z_{t+1} + \theta^2 B^2 Z_{t+1} - \theta^3 B^3 Z_{t+1} + \cdots = \epsilon_{t+1}\\
&Z_{t+1} = \theta Z_t - \theta^2 Z_{t-1} + \theta^3 Z_{t-2} + \cdots + \epsilon_{t+1}\\
&Y_{t+1} = \mu + \theta(Y_t - \mu) - \theta^2(Y_{t-1} - \mu) + \theta^3(Y_{t-2} - \mu) + \cdots + \epsilon_{t+1} \quad (**)\\
&Y_t = \mu + \theta(Y_{t-1} - \mu) - \theta^2(Y_{t-2} - \mu) + \theta^3(Y_{t-3} - \mu) + \cdots + \epsilon_t\\
&\theta Y_{t-1} = \theta\mu + \theta^2(Y_{t-2} - \mu) - \theta^3(Y_{t-3} - \mu) + \theta^4(Y_{t-4} - \mu) + \cdots + \theta\epsilon_{t-1}
\end{aligned}
$$

The coeff. of Y_1 is negligible as t gets larger, or $Y_{t+1} = g_t(\ldots, Y_1, \ldots, Y_t; \theta)$ with infinite past.

Next, use algebra to directly show that (**) implies (*)
(**) is in the form $Y_{t+1} = g_t(\ldots, Y_1, \ldots, Y_t; \theta) + \epsilon_{t+1}$.

From stochastic model MA(2) to forecast rule.
Suppose $Z_t = Y_t - \mu = \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2}$ for all t, where roots of $g(b) = (1 + \theta_1 b + \theta_2 b^2) = 0$ are $\eta_1, \eta_2 \in (-1, 1)$.

$$
\begin{aligned}
& Z_t = (1 - \eta_1 B)(1 - \eta_2 B)\epsilon_t \\
& (1 - \eta_1 B)^{-1}(1 - \eta_2 B)^{-1} Z_t = \epsilon_t \quad \text{(}|\eta_1| < 1,\ |\eta_2| < 1 \text{ for next)} \\
& (1 + \eta_1 B + \eta_1^2 B^2 + \eta_1^3 B^3 + \cdots +)(1 + \eta_2 B + \eta_2^2 B^2 + \eta_2^3 B^3 + \cdots +) Z_t = \epsilon_t \\
& Z_t + \psi_1 Z_{t-1} + \psi_2 Z_{t-2} + \psi_3 Z_{t-3} + \cdots = \epsilon_t \\
& \psi_1 = \eta_1 + \eta_2 \\
& \psi_2 = \eta_1^2 + \eta_1 \eta_2 + \eta_2^2 \\
& \psi_3 = \eta_1^2 + \eta_1^2 \eta_2 + \eta_1 \eta_2^2 + \eta_2^3 \\
& etc
\end{aligned}
$$

Rearrange so that Y_t is alone on left side; then shift subscripts.
Get the form $Y_{t+1} = g_t(\ldots, Y_1, \ldots, Y_t; \theta_1, \theta_2) + \epsilon_{t+1}$, and
$\widehat{y}_{t+1|t} = g_t(\ldots, y_1, \ldots, y_t, \widehat{\theta}_1, \widehat{\theta}_2)$.
Theory for estimation of parameters in ARMA forthcoming.

Appendix

```
# Check notation for MA coefficents in R
nn = 500
set.seed(443)
y1ts = arima.sim(n=nn,list(order=c(0,0,1),ma=0.5))
round(c(acf(y1ts,plot=F,lag.max=5)$acf ),3)
# [1]  1.000  0.351 -0.018 -0.034 -0.055 -0.030
# theoretical rho1 is theta/(1+theta^2)

y2ts = arima.sim(n=nn,list(order=c(0,0,1),ma=-0.5))
round(c(acf(y2ts,plot=F,lag.max=5)$acf ),3)
# [1]  1.000 -0.397 -0.047  0.069 -0.004  0.015

est1 = arima(y1ts,order=c(0,0,1)); print(est1)
#Coefficients:
#         ma1  intercept
#      0.4264    -0.0727
#s.e.  0.0413     0.0642
#sigma^2 estimated as 1.015:  log likelihood = -713.24,  aic = 1432.48

est2 = arima(y2ts,order=c(0,0,1)); print(est2)
#Coefficients:
#          ma1  intercept
#      -0.4981    -0.0283
#s.e.   0.0361     0.0217
#sigma^2 estimated as 0.9303:  log likelihood = -691.54,  aic = 1389.07
```