Handout 5

Methods for Nonstationary Time Series Transformations and ARIMA Models

Class notes for Statistics 451: Applied Time Series
Iowa State University

February 14, 2007
14h 46min

Dealing with Nonconstant Variance

- A common reason: variability increases with level (e.g., σ_a^2 is a percent of the level)

  This type of variance-nonstationarity can <u>sometimes</u> be handled with a transformation of the response.

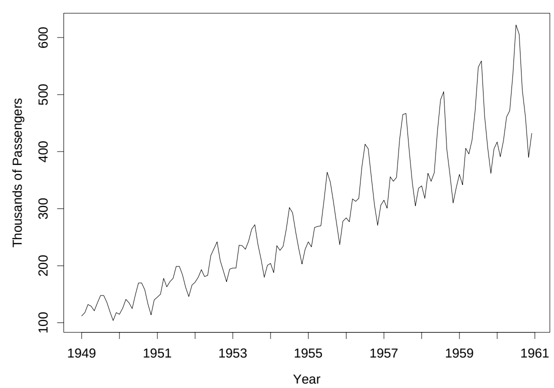
- Can also have variability changing independently of level.

  This type of nonstationarity may be handled by models allowing for nonstationary variance (e.g. Autoregressive Conditional Heteroscedasticity (ARCH) and Generalized Autoregressive Conditional Heteroscedasticity (GARCH) Models (not covered here in detail).

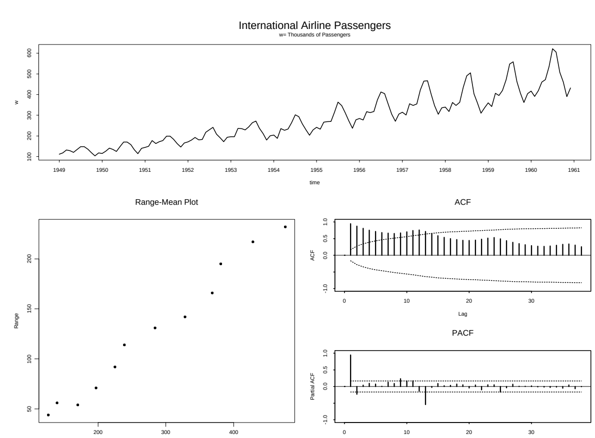
Variability Increasing with Level

- **Example:** If the standard deviation of sales is 10% of level, then for sales $\pm\,\sigma_a$ we have
  * 100 ± 10 or $[90, 110]$ when sales are 100
  * 1000 ± 100 or $[900, 1100]$ when sales are 1000
- On the log scale,
  * $[\log(90), \log(110)] = [2.2, 2.4]$ when sales are 100
  * $[\log(900), \log(1100)] = [4.5, 4.7]$ when sales are 1000
- Common problem when the "dynamic range" of a time series [measured by the ratio $\max(Z_t)/\min(Z_t)$] is large (say more than 3 or 4).

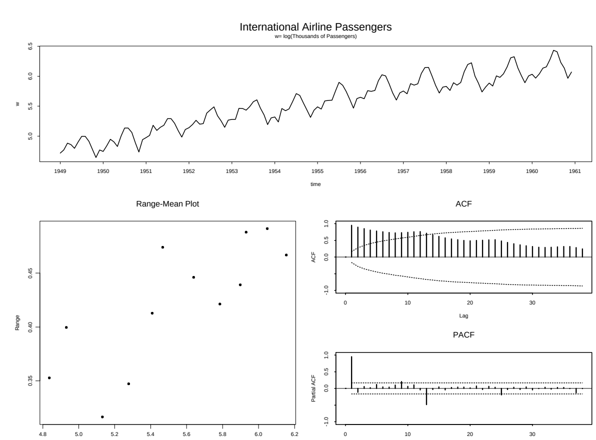
Number of international airline passengers from 1949 to 1960



Plot of the airline data along with a range-mean plot and plots of sample ACF and sample PACF. `[iden(airline.d)]`



Plot of the logarithms of the airline data along with a range-mean plot and plots of sample ACF and sample PACF. `[iden(airline.d,gamma=0)]`



Examples of Power Transformations

γ	Transformation	Possible Application
-1	$Z_t \sim -1/(Z_t^* + m)$	Very long upper tail
$-.333$	$Z_t \sim -1/\sqrt[3]{Z_t^* + m}$	Slightly stronger than log
0	$Z_t \sim \log(Z_t^* + m)$	Percentage change
.333	$Z_t \sim \sqrt[3]{Z_t^* + m}$	Slightly weaker than log
.5	$Z_t \sim \sqrt{Z_t^* + m}$	Count data
1	$Z_t \sim Z_t^*$	No transformation
2	$Z_t \sim [Z_t^* + m]^2$	?

Box-Cox Family of Transformations

$$Z_t = \begin{cases} \frac{(Z_t^*+m)^\gamma - 1}{\gamma} & \gamma \neq 0 \\ \log(Z_t^* + m) & \gamma = 0 \end{cases}$$

where Z_t^* is the original, untransformed time series, γ is primary transformation parameter, and log is natural log (i.e., base e).

- The quantity m is typically chosen to be 0. Choose $m > 0$ if some $Z_t^* < 0$ or $Z_t^* \approx 0$. Increasing m has the effect of weakening the transformation.
- For $\gamma < 0$, the γ in the denominator of the transformation assures that Z_t is an increasing function of Z_t^* so that a plot of Z_t has the same direction of trend as Z_t^*.
- Because

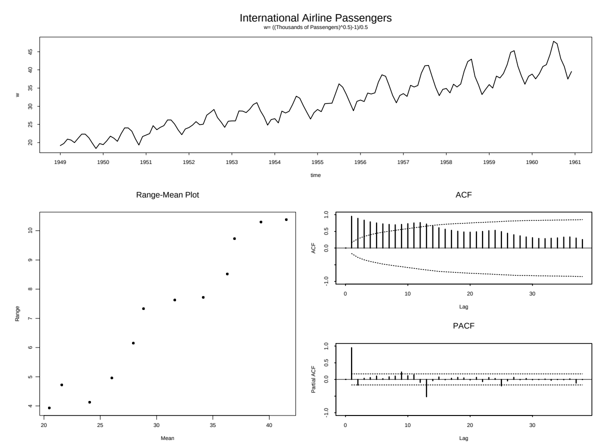
$$\lim_{\gamma \to 0} \frac{(Z_t^* + m)^\gamma - 1}{\gamma} = \log(Z_t^* + m)$$

the transformation is a continuous function in γ.

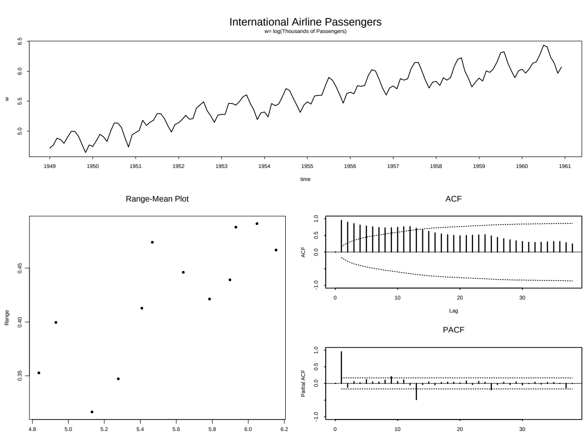
Range-Mean Plot

1. Divide realization into groups (4 to 12 in each group). If there is a natural seasonal period (e.g., 12 observations per year for monthly data), use it to choose the groups.
2. Transform data with given γ (and perhaps m)
3. Compute the mean and range in each group.
4. Plot the ranges versus the means
5. Repeat for different values of γ

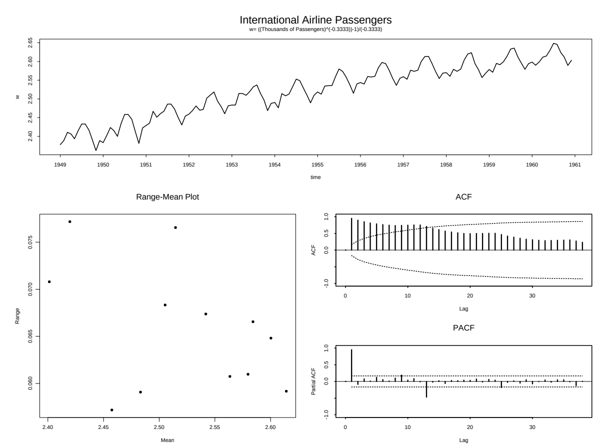
Plot of the squareroots ($\gamma = .5$) of the airline data along with a range-mean plot and plots of sample ACF and sample PACF. `[iden(airline.d,gamma=.5)]`



Plot of the logarithms ($\gamma = 0$) of the airline data along with a range-mean plot and plots of sample ACF and sample PACF. `[iden(airline.d,gamma=0)]`

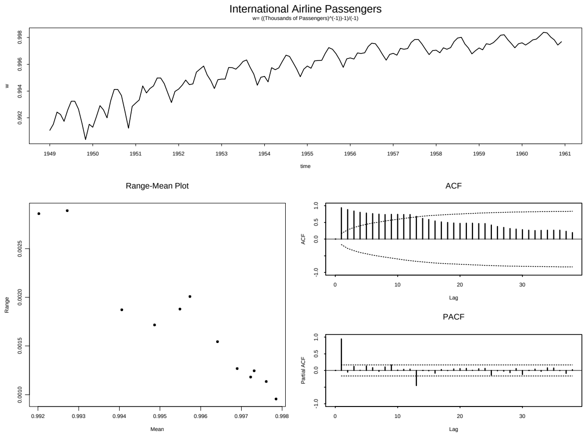


Plot of the reciprocal cube-root ($\gamma = -.3333$) of the airline data. `[iden(airline.d,gamma=-.3333)]`



Iden plot of the reciprocal ($\gamma = -1$) of the airline data.

[iden(airline.d,gamma=-1)]



Effects of Doing a Box-Cox Transformation

- Changes the relationship between amount of variability and level (as reflected in a range-mean plot)
- Changes the shape of a trend line (e.g., exponential or percentage trend versus linear trend)
- Changes the shape of the distribution of the residuals (e.g., from lognormal to normal)

Usual Procedure for Deciding on the Use of a Transformation

1. Try $\gamma = 1$ (no transformation)
2. Try $\gamma = .5, .333, 0$ (moderate to strong transformations)
3. Try $\gamma = -.333, -.5, -1$ (even stronger transformations)
4. Choose one tentative value of γ.

- Use the range-mean plot as a rough guide. It is not necessary to obtain 0 correlation between means and ranges.
- Choose one value for γ and use if for initial modeling. Then experiment with other values with your chosen model (usually the chosen model does not depend strongly on the choice of γ).
- Box-Cox transformations useful in other kinds of data analysis (e.g., regression analysis).
- Some computer programs attempt to estimate γ.

Dealing with Nonconstant Level (Mean)

- Fit a trend line (possibly after a transformation)
- Analyze differences (changes) instead of the actual time series [e.g., fit a model to $W_t = Z_t - Z_{t-1}$]

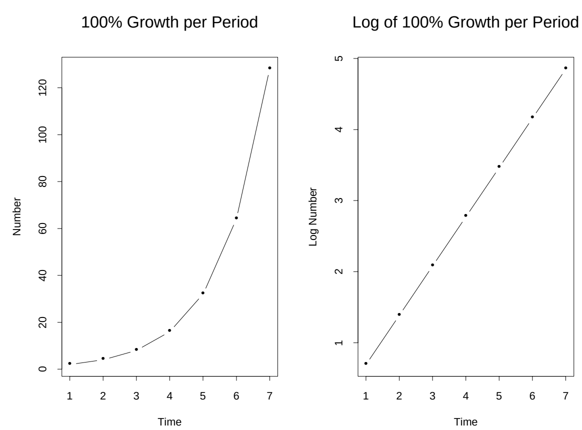
Example of Exponential (Percentage) Growth Number of Bacteria in a Dish Doubling Every Time Period (100% Growth)

t	Amount Z_t	Difference Amount $Z_t - Z_{t-1}$	Log Amount $\log(Z_t)$	Difference Log Amount $\log(Z_t) - \log(Z_{t-1})$
1	2	NA	0.69	NA
2	4	2	1.39	.70
3	8	4	2.08	.69
4	16	8	2.77	.69
5	32	16	3.47	.70
6	64	32	4.16	.69
7	128	64	4.85	.69

Differences of Z_t also grow exponentially

Differences of $\log(Z_t)$ are constant (except for roundoff)

Plots showing the effect of a log transformation on exponential growth



Exponential (Percentage) Growth

$$
\begin{aligned}
Z_t^* &= \beta_0^*[\beta_1^*]^t \\
Z_t = \log(Z_t^*) &= \log(\beta_0^*) + \log([\beta_1^*]^t) \\
Z_t &= \log(\beta_0^*) + \log(\beta_1^*)t \\
Z_t &= \beta_0 + \beta_1 t
\end{aligned}
$$

where $\beta_0 = \log(\beta_0^*)$ and $\beta_1 = \log(\beta_1^*)$.

Growth is rate in percent is

$$
\begin{aligned}
100\left[\frac{Z_{t+1}^* - Z_t^*}{Z_t^*}\right] &= 100\left[\frac{\beta_0^*[\beta_1^*]^{t+1} - \beta_0^*[\beta_1^*]^t}{\beta_0^*[\beta_1^*]^t}\right] \\
&= 100[\beta_1^* - 1] \\
&= 100[\exp(\beta_1) - 1] \approx 100\beta_1
\end{aligned}
$$

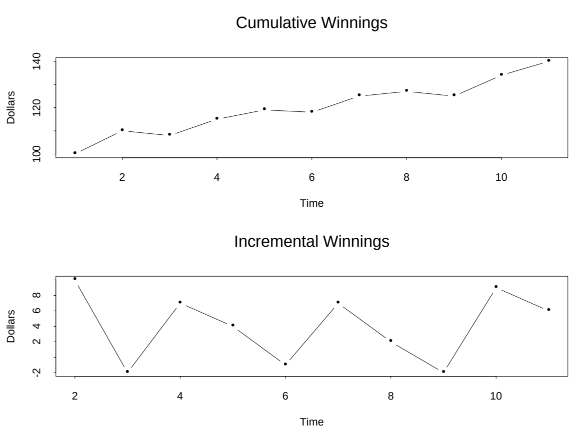
for small β_1

Gambler's Winnings

t	"Integrated" Z_t	Incremental $W_t = Z_t - Z_{t-1}$
1	100	NA
2	110	+10
3	108	−2
4	115	+7
5	119	+4
6	118	−1
7	125	+7
8	127	+2
9	125	−2
10	134	+9
11	140	+6

$$
\begin{aligned}
H_0: \mu_W = 0 \quad \bar{W} &= \frac{\sum W_i}{10} = \frac{40}{10} = 4 \\
S_W &= \sqrt{\frac{\sum (W_i - 4)^2}{9}} = 4.52 \\
t = \frac{\bar{W} - 0}{S_{\bar{W}}} = \frac{\bar{W} - 0}{S_W/\sqrt{10}} &= 2.79
\end{aligned}
$$

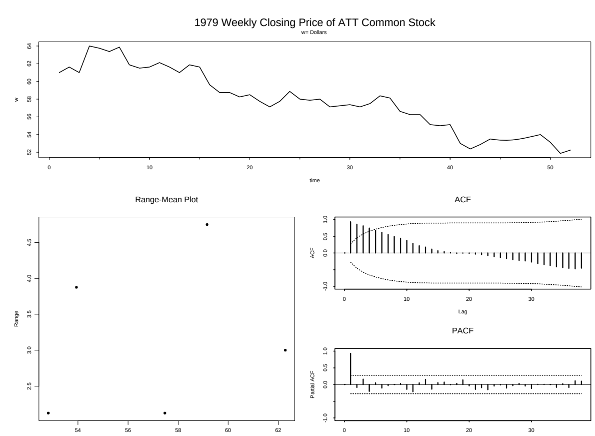
Plots of the gambler's cumulative winnings and incremental winnings



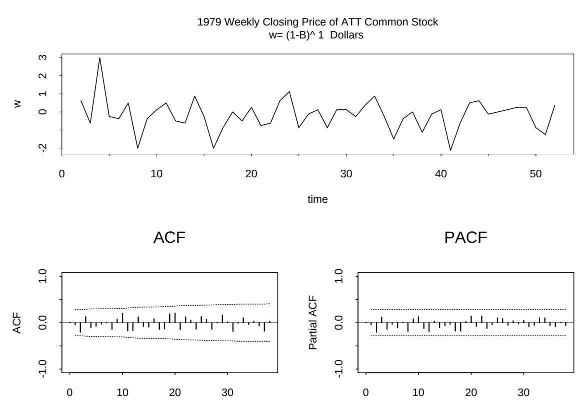
Forecasting the Gambler's Cumulative Winnings

- Because $W_t = (1-B)Z_t = Z_t - Z_{t-1}$ appears to be stationary, we can build an ARMA model for W_t and use $Z_t = Z_{t-1} + W_t$.
- If W_t is $W_t = \theta_0 + a_t$, then $Z_t = Z_{t-1} + \theta_0 + a_t$ is a "a random walk with drift" with E(increase) $= \theta_0$ per period.
- Estimated from data: $\widehat{\theta}_0 = \bar{W} = 4$, $\widehat{\sigma}_a = S_W = 4.52$
- A forecast (or prediction) interval for W_t can be translated into a prediction interval for Z_t (because we know Z_{t-1})
- One-step-ahead forecast for Z_t is $\widehat{Z}_t = Z_{t-1} + \widehat{\theta}_0$
- One-step-ahead $100(1-\alpha)\%$ prediction interval: $\widehat{Z}_t \pm t_{(1-\alpha/2, n_w - 1)} S_{(W-\bar{W})}$
- $S_{(W-\bar{W})} = \sqrt{S_W^2 + S_{\bar{W}}^2} = \sqrt{S_W^2 + S_W^2/n_w}$
- $S_{(W-\bar{W})} \approx S_W = \widehat{\sigma}_a$ for large n_w

Function `iden` Output for 1979 Weekly Closing Prices of AT&T Common stock. [iden(att.d)]



Function `iden` Output for the First Differences of 1979 Weekly Closing Prices of AT&T Common Stock. [iden(att.d,d=1)]



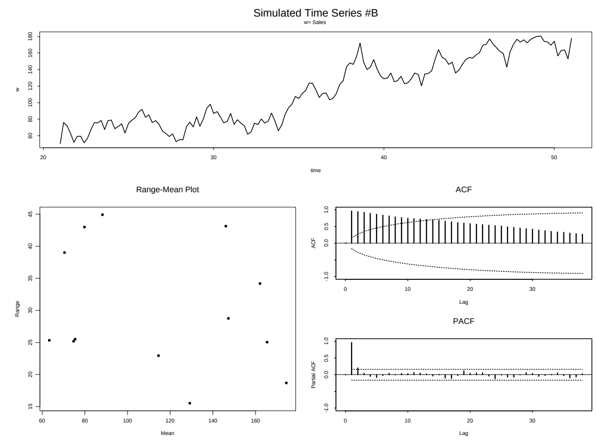
Forecasting a Random Walk

- If the model for $W_t = (1 - \text{B})Z_t = Z_t - Z_{t-1}$ is $W_t = a_t$, then

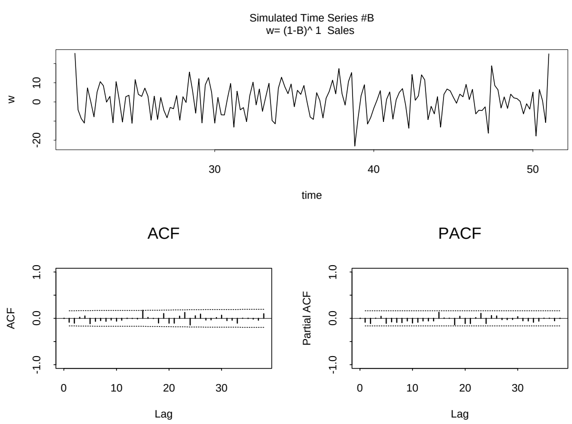
$$\begin{aligned} Z_t &= Z_{t-1} + a_t \\ &= a_{t-1} + a_{t-2} + a_{t-3} + \cdots + a_t \end{aligned}$$

  which is a "a random walk" process. This is like an AR(1) model with $\phi_1 = 1$ (note that model is nonstationary).
- The only parameter of this model is σ_a
- A forecast (or prediction interval) for W_t can be translated into a forecast (or prediction interval) for Z_t (because we know Z_{t-1})
- One-step-ahead forecast for Z_t is $\widehat{Z}_t = Z_{t-1}$
- One-step-ahead prediction interval is: $\widehat{Z}_t \pm t_{(1-\alpha/2,n-1)} \text{S}_W$

Function `iden` Output for Simulated Series B



Function `iden` Output for the First Differences of Simulated Series B



AutoRegressive Integrated Moving Average Model (ARIMA)

- An ARIMA(p, d, q) model is

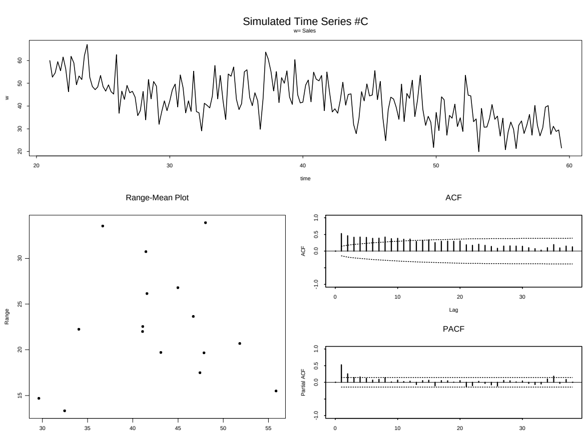
$$\phi_p(\text{B})(1 - \text{B})^d Z_t = \theta_q(\text{B}) a_t$$

  can be viewed as a generalized ARMA model, with d roots on the unit circle.
- Letting $W_t = (1 - \text{B})^d Z_t$ be the "working series" and substituting gives an ARMA model for W_t

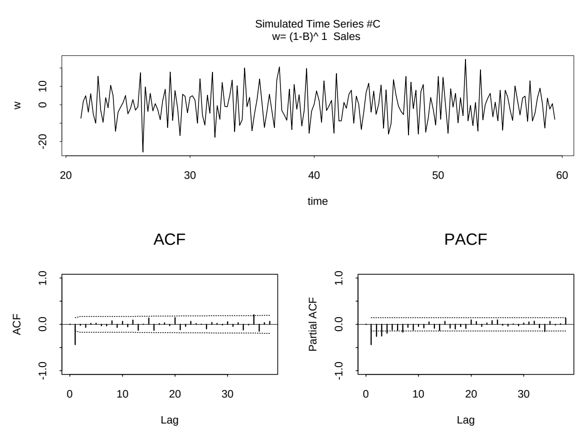
$$\phi_p(\text{B}) W_t = \theta_q(\text{B}) a_t$$

- Can fit an ARMA model to W_t
- Can unscramble the Z_t for forecasting and interpretation

Function `iden` Output for Simulated Series C



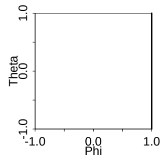
Function `iden` Output for the First Differences of Simulated Series C



Special Case: ARIMA$(0,1,1)$ Model [a.k.a. IMA(1,1)]

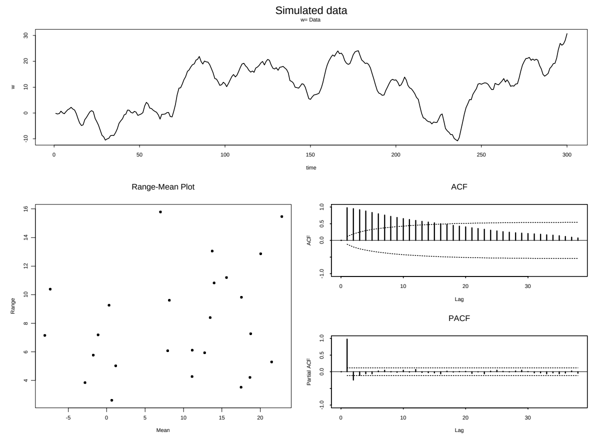
$$\begin{aligned}(1-\mathrm{B})^1 Z_t &= (1-\theta_1\mathrm{B})a_t \\ Z_t &= Z_{t-1} - \theta_1 a_{t-1} + a_t\end{aligned}$$

which is a <u>nonstationary</u> "ARMA(1,1)"

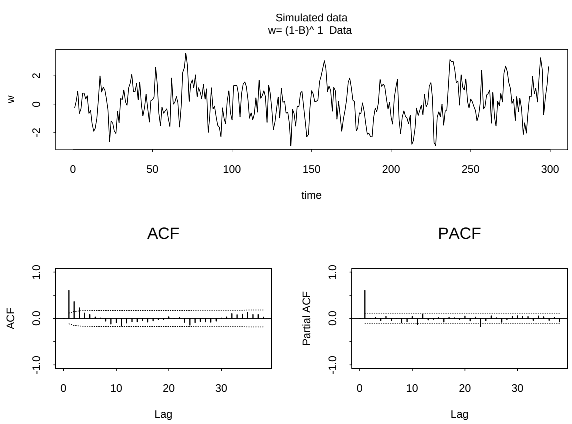


This model is also known as exponentially weighted moving average (EWMA) and its forecast equations can be shown to be equivalent to exponential smoothing.

Function `iden` Output for Simulated Series D

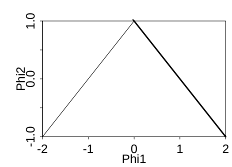


Function `iden` Output for the First Differences of Simulated Series D



Special Case: ARIMA$(1,1,0)$ Model

- $(1-\phi_1\mathrm{B})(1-\mathrm{B})^1 Z_t = a_t$ or
  $(1-\phi_1\mathrm{B}-\mathrm{B}+\phi_1\mathrm{B}^2)Z_t = a_t$, leading to
  $$Z_t = (1+\phi_1)Z_{t-1} - \phi_1 Z_{t-2} + a_t$$
  which is a <u>nonstationary</u> "ARIMA(2,0,0)" or "AR(2)"



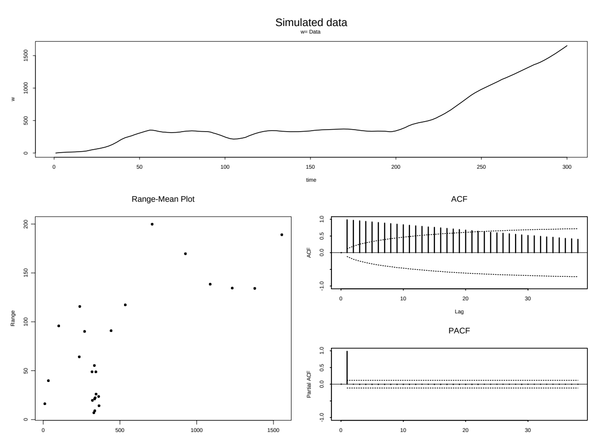
$$Z_t = \phi_1' Z_{t-1} + \phi_2' Z_{t-2} + a_t$$

where $\phi_1' = (1+\phi_1)$ and $\phi_2' = -\phi_1$ or $\phi_2' = 1 - \phi_1'$

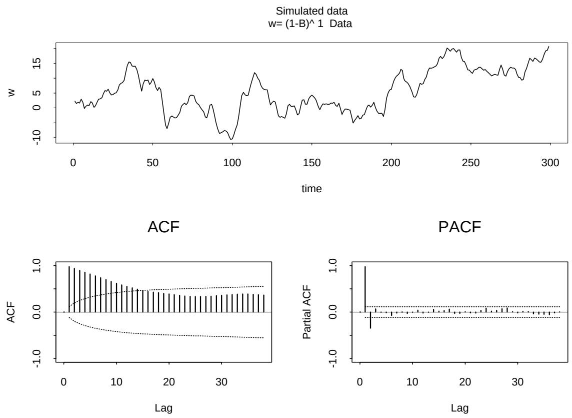
How to Choose d (the amount of differencing)

1. $d=0$ to $d=1$ most common; $d=2$ not common; $d \geq 3$ generally not useful;
2. Plot data versus time, looking for trend and cycle?
3. Consider the physical process (is the process changing?)
4. Examine the ACF for $d=0$, stepping up to $d=1$ or $d=2$, if needed. Look for the smallest d such that ACF "dies down quickly."
5. Fit AR(1) model and test $\mathrm{H}_0 : \phi_1 = 1$ using
   $$t = \frac{\widehat{\phi}_1 - 1}{\mathrm{S}_{\widehat{\phi}_1}}$$
   Need special tables [see Dickey and Fuller 1979 JASA]
6. Minimize S_W
7. Compare forecasts and prediction intervals.

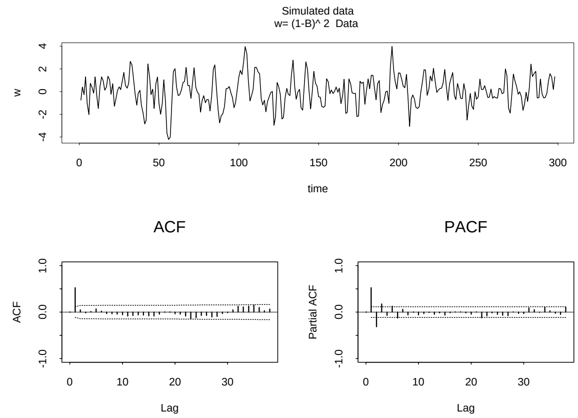
Function `iden` Output for Simulated Series E



Function `iden` Output for the First Differences of Simulated Series E



Function `iden` Output for the Second Differences of Simulated Series E



An Example of a Nonstationary Model

Let $t = 1, 2, \ldots$ be coded time.

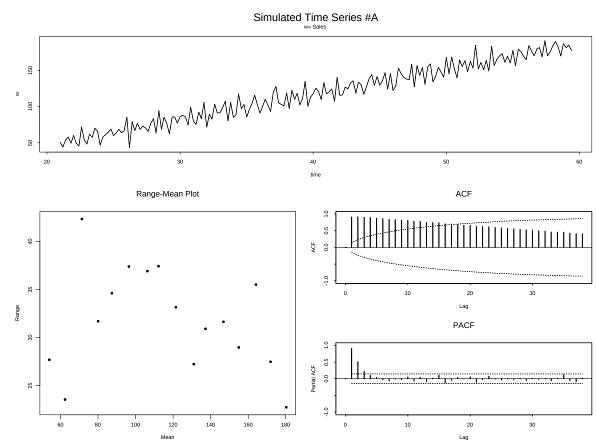
$$\begin{aligned}
Z_t &= \beta_0 + \beta_1 t + a_t, \quad a_t \sim \text{NID}(0, \sigma_a^2) \\
\text{E}(Z_t) &= \beta_0 + \beta_1 t \\
W_t &= (1 - \text{B})Z_t = Z_t - Z_{t-1} \\
&= [\beta_0 + \beta_1 t + a_t] - [\beta_0 + \beta_1(t-1) + a_{t-1}] \\
&= \beta_1 - a_{t-1} + a_t \\
\text{E}(W_t) &= \beta_1 \\
\sigma_W^2 &= \text{Var}(W_t) = 2\sigma_a^2
\end{aligned}$$

Therefore, Z_t is nonstationary, but W_t is stationary. What kind of model does W_t have??

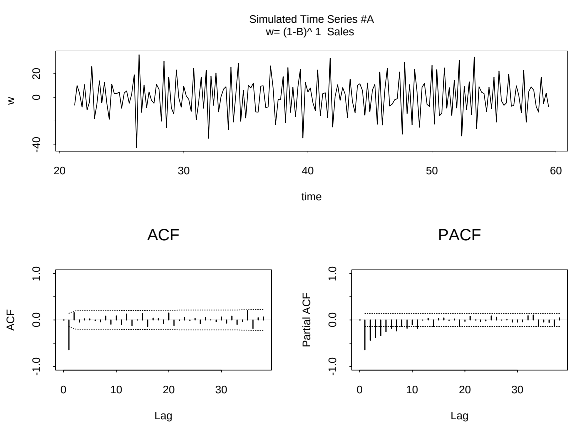
Is there a better solution for fitting a model to Z_t??

Beware of differencing when differencing is not warranted.

Function `iden` Output for Simulated Series A



Function `iden` Output for the First Differences of Simulated Series A



Beware of Over Differencing

Trivial model:

$$Z_t = \theta_0 + a_t$$

First difference of Z_t

$$\begin{aligned}
W_t &= (1 - \text{B})Z_t = Z_t - Z_{t-1} \\
&= [\theta_0 + a_t] - [\theta_0 + a_{t-1}] \\
&= -a_{t-1} + a_t
\end{aligned}$$

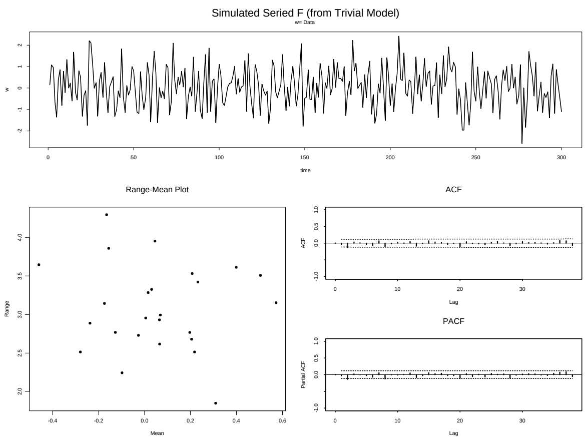
is a noninvertible MA(1)!

Second difference of Z_t

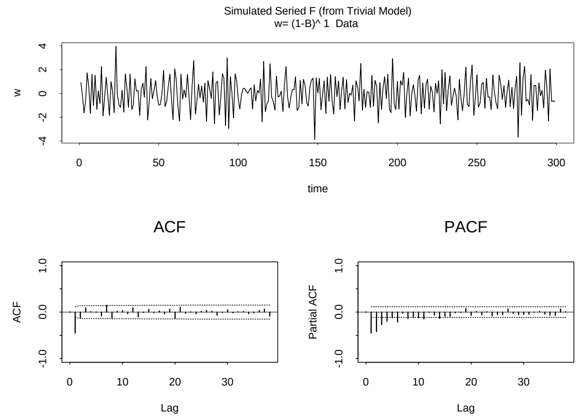
$$\begin{aligned}
W_t &= (1 - \text{B})^2 Z_t = Z_t - 2Z_{t-1} + Z_{t-2} \\
&= -2a_{t-1} + a_{t-2} + a_t
\end{aligned}$$

is a noninvertible MA(2)!!

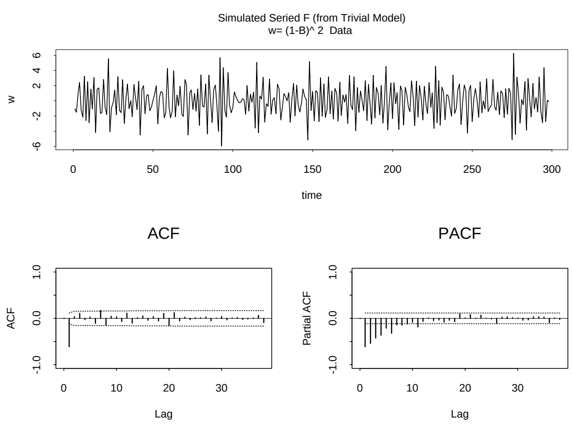
Function iden Output for Simulated Series F



Function iden Output for the First Differences of Simulated Series F



Function iden Output for the Second Differences of Simulated Series F



When to Include θ_0 in an ARIMA Model

- With no differencing ($d = 0$) include a constant term in the model to allow estimation of the process mean.
- If there is differencing ($d = 1$) then a constant term should be included only if there is need to or evidence of deterministic trend. With stationary W_t, E$(W_t) = 0$, and

$$Z_t = Z_{t-1} + \theta_0 + W_t$$

  the deterministic trend is θ_0 each time period.
- Use of a constant term after differencing is rare. Check:
  - $\mathrm{H}_0 : \mu_W = 0$ by looking at $t = (\bar{W} - 0)/S_{\bar{W}}$
  - The physical nature of the data-generating process
- Situation is similar, but more complicated, with higher order differencing.

Relationship Between IMA(1,1) and Exponential Smoothing

IMA(1,1) model, unscrambled

$$Z_t \;=\; Z_{t-1} - \theta_1 a_{t-1} + a_t$$

IMA(1,1) forecast

$$\begin{aligned}
\widehat{Z}_t &= Z_{t-1} - \widehat{\theta}_1 \widehat{a}_{t-1} \\
&= Z_{t-1} - \widehat{\theta}_1 (Z_{t-1} - \widehat{Z}_{t-1}) \\
&= (1 - \widehat{\theta}_1) Z_{t-1} + \widehat{\theta}_1 \widehat{Z}_{t-1} \\
&= \alpha Z_{t-1} + (1 - \alpha) \widehat{Z}_{t-1} \\
&= \alpha Z_{t-1} + \alpha(1 - \alpha) Z_{t-2} + \alpha(1 - \alpha)^2 Z_{t-3} + \cdots
\end{aligned}$$

where $\alpha = 1 - \theta_1$. Usually $0.01 < \alpha < 0.1$. This shows why the IMA(1,1) forecast equation is sometimes called "exponential smoothing" or "exponentially weighted moving average" (EWMA).

Issues in Applying Exponential Smoothing

- Choice of the smoothing constant α.
- Start-up value for the forecasts?
- Single, double, or triple exponential smoothing? (Single exponential smoothing is equivalent to IMA(1,1), double exponential smoothing is equivalent to IMA(2,2), ...)
- Seasonality (Winter's method)
- Prediction intervals or bounds?