Module 6
Methods for Nonstationary Time Series
Transformations, Differencing, and ARIMA Model
Identification

Class notes for Statistics 451: Applied Time Series
Iowa State University
Copyright 2015 W. Q. Meeker.

Dealing with Nonconstant Variance

- A common reason: Variability increases with level
- \[ \text{Range} = \text{Maximum} - \text{Minimum} \]

Example: If the standard deviation of sales is 10% of level,
then for sales \( \geq 100 \), we have

Variability increasing with level

Number of international airline passengers from 1949 to 1960

<table>
<thead>
<tr>
<th>Year</th>
<th>Thousands of Passengers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949</td>
<td>100</td>
</tr>
<tr>
<td>1951</td>
<td>200</td>
</tr>
<tr>
<td>1953</td>
<td>300</td>
</tr>
<tr>
<td>1955</td>
<td>400</td>
</tr>
<tr>
<td>1957</td>
<td>500</td>
</tr>
<tr>
<td>1959</td>
<td>600</td>
</tr>
</tbody>
</table>

Plot of the airline data along with a range-mean plot

<table>
<thead>
<tr>
<th>Time</th>
<th>w</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949</td>
<td>100</td>
</tr>
<tr>
<td>1950</td>
<td>200</td>
</tr>
<tr>
<td>1951</td>
<td>300</td>
</tr>
<tr>
<td>1952</td>
<td>400</td>
</tr>
<tr>
<td>1953</td>
<td>500</td>
</tr>
<tr>
<td>1954</td>
<td>600</td>
</tr>
<tr>
<td>1955</td>
<td>700</td>
</tr>
<tr>
<td>1956</td>
<td>800</td>
</tr>
<tr>
<td>1957</td>
<td>900</td>
</tr>
<tr>
<td>1958</td>
<td>1000</td>
</tr>
<tr>
<td>1959</td>
<td>1100</td>
</tr>
<tr>
<td>1960</td>
<td>1200</td>
</tr>
<tr>
<td>1961</td>
<td>1300</td>
</tr>
</tbody>
</table>

International Airline Passengers

<table>
<thead>
<tr>
<th>Time</th>
<th>w= Thousands of Passengers</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949</td>
<td>100</td>
</tr>
<tr>
<td>1950</td>
<td>200</td>
</tr>
<tr>
<td>1951</td>
<td>300</td>
</tr>
<tr>
<td>1952</td>
<td>400</td>
</tr>
<tr>
<td>1953</td>
<td>500</td>
</tr>
<tr>
<td>1954</td>
<td>600</td>
</tr>
<tr>
<td>1955</td>
<td>700</td>
</tr>
<tr>
<td>1956</td>
<td>800</td>
</tr>
<tr>
<td>1957</td>
<td>900</td>
</tr>
<tr>
<td>1958</td>
<td>1000</td>
</tr>
<tr>
<td>1959</td>
<td>1100</td>
</tr>
<tr>
<td>1960</td>
<td>1200</td>
</tr>
<tr>
<td>1961</td>
<td>1300</td>
</tr>
</tbody>
</table>

Transformations
Nonconstant Variance and an Introduction to
Segment 1
Module 6
Plot of the logarithms of the airline data along with a range-mean plot and plots of sample ACF and sample PACF.

\[ \text{idem(airline.tsd,gamma=0)} \]

<table>
<thead>
<tr>
<th>Mean</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2</td>
<td>0.35</td>
</tr>
<tr>
<td>5.4</td>
<td>0.40</td>
</tr>
<tr>
<td>5.6</td>
<td>0.45</td>
</tr>
</tbody>
</table>

**Range-Mean Plot**

<table>
<thead>
<tr>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>1949</td>
</tr>
<tr>
<td>1950</td>
</tr>
<tr>
<td>1951</td>
</tr>
<tr>
<td>1952</td>
</tr>
<tr>
<td>1953</td>
</tr>
<tr>
<td>1954</td>
</tr>
<tr>
<td>1955</td>
</tr>
<tr>
<td>1956</td>
</tr>
<tr>
<td>1957</td>
</tr>
<tr>
<td>1958</td>
</tr>
<tr>
<td>1959</td>
</tr>
<tr>
<td>1960</td>
</tr>
<tr>
<td>1961</td>
</tr>
</tbody>
</table>

**International Airline Passengers**

\( w = \log(\text{Thousands of Passengers}) \)

**ACF**

<table>
<thead>
<tr>
<th>Lag</th>
<th>ACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
</tbody>
</table>

**PACF**

<table>
<thead>
<tr>
<th>Lag</th>
<th>Partial ACF</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>10</td>
<td>0.5</td>
</tr>
<tr>
<td>20</td>
<td>1</td>
</tr>
</tbody>
</table>

**Examples of Power Transformations**

<table>
<thead>
<tr>
<th>( \gamma )</th>
<th>Transformation Possible Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>( -1 )</td>
<td>( Z_t \sim \frac{-1}{Z_t^* + m} ) Very long upper tail</td>
</tr>
<tr>
<td>( -\frac{3}{3} )</td>
<td>( Z_t \sim \frac{-1}{3} \sqrt{Z_t^* + m} ) Slightly stronger than log</td>
</tr>
<tr>
<td>0</td>
<td>( Z_t \sim \log(Z_t^* + m) ) Percentage change</td>
</tr>
<tr>
<td>( \frac{1}{2} )</td>
<td>( Z_t \sim \sqrt{Z_t^* + m} ) Slightly weaker than log</td>
</tr>
<tr>
<td>1</td>
<td>( Z_t \sim Z_t^* ) No transformation</td>
</tr>
<tr>
<td>2</td>
<td>( Z_t \sim (Z_t^* + m)^2 ) Possible application</td>
</tr>
</tbody>
</table>

**Box-Cox Transformations and the Range-Mean Plot**

1. Divide realization into groups (4 to 12 in each group).
2. Transform data with given \( \gamma \) (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 \( \gamma \).

**Box-Cox Family of Transformations**

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

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

- For \( \gamma > 0 \), the \( \gamma \) in the denominator of the transformation has the same direction of trend as \( Z_t^* \) so that a plot of \( Z_t \) has the same direction of trend as \( Z_t^* \).
- Because the transformation is a continuous function in \( \gamma \), \( \log(Z_t^* + m) \) is a monotonically increasing function of \( Z_t^* \) as \( \gamma \) increases.
- For \( \gamma < 0 \), the denominator of the transformation is a decreasing function of \( Z_t^* \) and the transformation is thus a monotonically decreasing function of \( Z_t^* \).
- For \( \gamma = 0 \), the transformation is the identity function:
  \[ Z_t = \log(Z_t^* + m) \]

**Plot of the square roots (\( \gamma = 0.5 \)) of the airline data** along with a range-mean plot and plots of sample ACF and sample PACF.

\[ \text{idem(airline.tsd,gamma=.5)} \]
Nonconstant Level and Exponential Growth

Segment 3

Module 6

Effects of Doing a Box-Cox Transformation

from Lognormal to Normal
• Changes the shape of the distribution of the residuals (e.g.,
  centred trend versus linear trend)
• Changes the shape of a trend line (e.g., exponential or per-
  cented (as reflected in a range-mean plot)
• Changes the relationship between amount of variability and

usual Procedure for Deciding on the Use of a

4. Choose one tentative value of γ
  3. Try γ = −0.3333 (certain stronger transformations)
  2. Try γ = 0.3333 (moderate to strong transformations)
  1. Try γ = 1 (no transformation)

Box-Cox transformations useful in other kinds of data anal-
• Some computer programs attempt to estimate γ
  - 6.5, 3333 (moderate to strong transformations)
  - 3. Try γ = −0.3333 (certain stronger transformations)
  - 2. Try γ = 0.3333 (moderate to strong transformations)
  - 1. Try γ = 1 (no transformation)
Dealing with Nonconstant Level (Trend or Cycle)

- Fit a trend line (possibly after a transformation)
- Analyze differences (changes) instead of the actual time series \[e.g., \text{fit a model to } W_t = Z_t - Z_{t-1}\]
- If transformation and differencing is used always do the transformation first

**Example of Exponential (Percentage) Growth**

<table>
<thead>
<tr>
<th>Time</th>
<th>Number of Bacteria in a Dish</th>
<th>Difference</th>
<th>Amount</th>
<th>Log Amount</th>
<th>Log Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>NA</td>
<td>0.69</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>2</td>
<td>1.39</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>4</td>
<td>2.08</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>16</td>
<td>8</td>
<td>2.77</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>32</td>
<td>16</td>
<td>3.47</td>
<td>.70</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>64</td>
<td>32</td>
<td>4.16</td>
<td>.69</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>128</td>
<td>64</td>
<td>4.85</td>
<td>.69</td>
<td></td>
</tr>
</tbody>
</table>

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**

- For small \(\beta_1\)
  \[100\% \text{ growth rate in percent is } \beta_1 \approx 100 \left[ \frac{1 - e^{\beta_1}}{e^{\beta_1}} \right] \]
  \[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]
  \[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

Expontial (Percentage) Growth

**Example of Exponential (Percentage) Growth**

\[Z_t = \beta_0 + \beta_1 t\]

\[Z_t = \log(\beta_0 + \beta_1 t) = \log(\beta_0) + \log(\beta_1 + t)\]

where \(\beta_0 = \log(\beta_0)\) and \(\beta_1 = \log(\beta_1)\)

\[100 \% \text{ growth } = 100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + \beta_1 + 100)} \right] \]

\[100 \left[ \frac{\log(1 + \beta_1)}{\log(1 + e^{\beta_1})} \right] \]
ARIMA Models Part 1: ARIMA(0, 1, 1)

Forecasting a Random Walk

If the model for \( W_t = (1 - B)Z_t \) is
\[
W_t = \theta_0 + \alpha_t,
\]
then
\[
Z_t = Z_t - 1 + \theta_0 + \alpha_t
\]
which is a "random walk" process. This is like an AR(1) model with \( \phi = 1 \) (note that model is nonstationary).

The only parameter of this model is \( \sigma_\alpha \).

The one-step-ahead point forecast for \( Z_t \) is
\[
\hat{Z}_t = Z_t - 1 + \theta_0
\]
and
\[
\text{One-step-ahead prediction interval for } Z_t \text{ is } [Z_t - \hat{\sigma}_Z, Z_t + \hat{\sigma}_Z]
\]
where \( \hat{\sigma}_Z \) is the estimated standard deviation of \( Z_t \).
To be equivalent to exponential smoothing, this model is also known as an exponentially weighted moving average (EWMA) and its forecast equations can be shown as:

\[ W_t = (1 - \theta B) Z_t = Z_t - \theta Z_{t-1} + \theta a_t \]

This model is known as an exponentially weighted moving average (EWMA)model [a.k.a. IMA(1,1)].

Which is a transformation of ARIMA(1,1,1):

\[ 1 + \theta B \phi B = 1 \]

\[ W_t = Z_t - (1 - \theta)(Z_{t-1} - \theta Z_{t-2} + \theta a_{t-1}) + \theta a_t \]

This model is also known as exponentially weighted moving average (EWMA) and its forecast equations can be shown as:

\[ W_t = (1 - \theta B) Z_t = Z_t - \theta Z_{t-1} + \theta a_t \]

\[ Z_t = Z_{t-1} - \theta Z_{t-2} + \theta a_{t-1} + a_t \]

This model is also known as exponentially weighted moving average (EWMA) and its forecast equations can be shown as:

\[ W_t = (1 - \theta B) Z_t = Z_t - \theta Z_{t-1} + \theta a_t \]

\[ Z_t = Z_{t-1} - \theta Z_{t-2} + \theta a_{t-1} + a_t \]

This model is also known as exponentially weighted moving average (EWMA) and its forecast equations can be shown as:

\[ W_t = (1 - \theta B) Z_t = Z_t - \theta Z_{t-1} + \theta a_t \]

\[ Z_t = Z_{t-1} - \theta Z_{t-2} + \theta a_{t-1} + a_t \]

This model is also known as exponentially weighted moving average (EWMA) and its forecast equations can be shown as:

\[ W_t = (1 - \theta B) Z_t = Z_t - \theta Z_{t-1} + \theta a_t \]

\[ Z_t = Z_{t-1} - \theta Z_{t-2} + \theta a_{t-1} + a_t \]
Module 6
Segment 6
ARIMA Models Part 2: ARIMA(1, 1, 0)

Function iden
Output for Simulated Series D
Mean
Range

-5 0 5 10 15 20
4 6 8 10 12 14 16

Range-Mean Plot
time

w
0 50 100 150 200 250 300

Simulated data
w= Data

ACF
Lag
ACF
0 10 20 30
-1.0 0.0 0.5 1.0

PACF
Lag
Partial ACF
0 10 20 30
-1.0 0.0 0.5 1.0

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

w
0 50 100 150 200 250 300
-2 0 2

Simulated data
w= (1-B)^1 Data

ACF
Lag
ACF
0 10 20 30
-1.0 0.0 1.0

PACF
Lag
Partial ACF
0 10 20 30
-1.0 0.0 1.0

Special Case: ARIMA(1, 1, 0) Model
\[ (1 - \phi L)(1 - L)Z_t = a_t \]
where
\[ b = \phi(1 + \phi) \]
\[ a_t = \phi^2 a_{t-2} + \phi^2 a_{t-1} + a_t \]

leading to
\[ a_t = \phi^2 a_{t-2} + \phi^2 a_{t-1} + a_t \cdot \phi \]

which is nonstationary, \( \text{ARIMA}(2,0,0) \) or \( \text{AR}(2) \)

1. \( d = 0 \) to \( d = 1 \) most common; \( d = 2 \) not common; \( d \geq 3 \) generally not used

2. Plot data-versus-time, looking for trend and cycle
3. Consider the physical process (is the process changing?)
4. Examine the ACF for its smallest \( d \) such that ACF decreases
5. Fit AR(1) model and test \( H_0: \phi = 1 \) using
\[ t = \frac{\hat{\phi} - 1}{\hat{\sigma}} \]

6. Need special tables (see Dickey and Fuller 1979 JASA)

7. Compare forecasts and prediction intervals.
Beware of differencing when differencing is not warranted.

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

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

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

An example of a nonstationary model.

Module 6

Segment 7
Beware of Over Differencing

Trivial model:

\[ Z_t = \theta_0 + a_t, \quad a_t \sim \text{NID}(0, \sigma_a^2) \]

First difference of \( Z_t \):

\[ W_t = (1 - B)Z_t = Z_t - Z_{t-1} = [\theta_0 + a_t] - [\theta_0 + a_{t-1}] = -a_t - a_{t-1} + a_t \]

is a noninvertible MA(1).

Second difference of \( Z_t \):

\[ W_t = (1 - B)^2Z_t = Z_t - 2Z_{t-1} + Z_{t-2} = -2a_t - a_{t-1} + a_{t-2} + a_{t-3} \]

is a noninvertible MA(2).

---

When to Include \( \theta_0 \) 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 \( \theta_0 \) each time period.
• Use of a constant term after differencing is rare. Check:
  - \( H_0: \mu_W = 0 \) by looking at \( t = (W - \mu)/\sigma_W \)
  - The physical nature of the data-generating process
• Situation is similar, but more complicated, with higher order differencing.

---

Exponential Smoothing

Using a Constant Term After Differencing and

Module 6
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

\[ \hat{Z}_t = Z_{t-1} - \hat{\theta}_1 \hat{a}_{t-1} = Z_{t-1} - \hat{\theta}_1 (Z_{t-1} - \hat{Z}_{t-1}) = (1 - \hat{\theta}_1) Z_{t-1} + \hat{\theta}_1 \hat{Z}_{t-1} \]

\[ = \alpha Z_{t-1} + (1 - \alpha) \hat{Z}_{t-1} \]

where \( \alpha = 1 - \theta_1 \). Usually \( 0.0 < \alpha < 1.0 \). This shows why the IMA(1,1) forecast equation is sometimes called "exponentially weighted moving average" (EWMA).

Issues in Applying Exponential Smoothing

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

...