Box-Jenkins Seasonal Modeling

Five-step Iterative Procedures

- Stationarity Checking and Differencing
- Model Identification
- Parameter Estimation
- Diagnostic Checking
- Forecasting

Step One: Stationarity Checking and Differencing

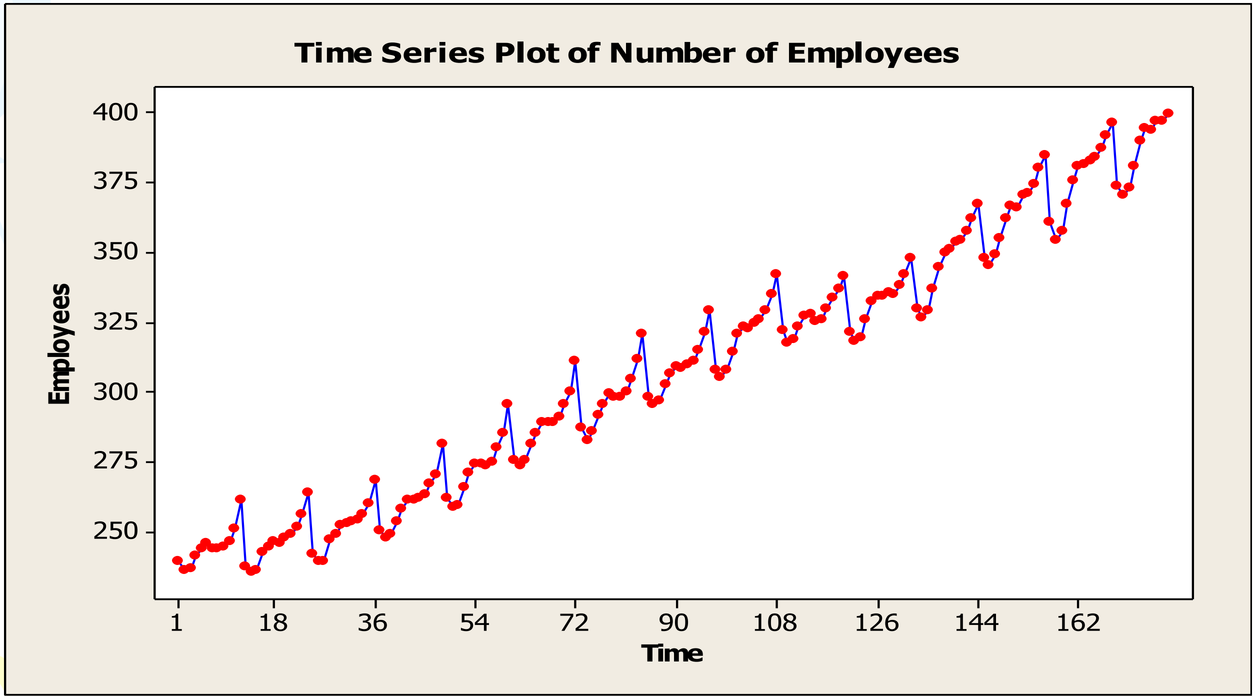
Seasonality

- Seasonality is defined as a pattern that repeats itself over fixed intervals of time.
- In general, seasonality can be found by identifying a large autocorrelation coefficient or a large partial autocorrelation coefficient at the seasonal lag.
- Often autocorrelations at multiples of the seasonal lag will also be significant.

Example: Employees

178 monthly values of the number of people in Wisconsin employed in “trade” are reported from 1961 to 1975.

- Dataset: employees.txt
- SAS program: ARIMA_employees.sas



Example: Employees

Autocorrelations

Lag	Covariance	Correlation	-1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1	Std Error
0	2174.219	1.00000	********************	0
1	2111.301	0.97106	. *******************	0.074953
2	2046.143	0.94109	. *******************	0.127330
3	1990.467	0.91549	. ******************	0.161754
4	1953.651	0.89855	. ******************	0.188630
5	1923.082	0.88449	. ******************	0.211313
6	1894.387	0.87130	. *****************	0.231178
7	1857.165	0.85418	. *****************	0.248944
8	1822.990	0.83846	. *****************	0.264898
9	1795.368	0.82575	. *****************	0.279410
10	1781.604	0.81942	. ****************	0.292800
11	1766.588	0.81252	. ****************	0.305411
12	1754.960	0.80717	. ****************	0.317323
13	1689.253	0.77695	. ****************	0.328655
14	1622.604	0.74629	. ***************	0.338817
15	1565.605	0.72008	. **************	0.347929
16	1526.444	0.70207	. **************	0.356203
17	1493.548	0.68694	. **************.	0.363894
18	1462.579	0.67269	. ************* .	0.371108
19	1424.437	0.65515	. ************* .	0.377896
20	1390.875	0.63971	. ************* .	0.384224
21	1363.633	0.62718	. ************* .	0.390162
22	1347.737	0.61987	. ************ .	0.395785
23	1328.662	0.61110	. ************ .	0.401202
24	1312.463	0.60365	. ************ .	0.406398
25	1248.092	0.57404	. *********** .	0.411404
26	1183.569	0.54437	. *********** .	0.415880
27	1127.791	0.51871	. ********** .	0.419864
28	1090.018	0.50134	. ********** .	0.423448
29	1058.246	0.48672	. ********** .	0.426770
30	1027.732	0.47269	. ********* .	0.429877
31	990.402	0.45552	. ********* .	0.432787
32	955.643	0.43953	. ********* .	0.435473
33	926.976	0.42635	. ********* .	0.437958
34	910.295	0.41868	. ******** .	0.440283
35	891.731	0.41014	. ******** .	0.442514
36	876.729	0.40324	. ******** .	0.444645

"." marks two standard errors

Example: Employees

```
                           Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1        0.97106    |                 .  |******************* |
  2       -0.03275    |                 . *|  .                 |
  3        0.06141    |                 .  |* .                 |
  4        0.13652    |                 .  |***                 |
  5        0.04323    |                 .  |* .                 |
  6        0.02889    |                 .  |* .                 |
  7       -0.04794    |                 . *|  .                 |
  8        0.03610    |                 .  |* .                 |
  9        0.04774    |                 .  |* .                 |
 10        0.09973    |                 .  |**.                 |
 11       -0.00141    |                 .  |  .                 |
 12        0.05761    |                 .  |* .                 |
 13       -0.41758    |            ********|  .                 |
 14        0.00616    |                 .  |  .                 |
 15        0.01128    |                 .  |  .                 |
 16        0.03402    |                 .  |* .                 |
 17        0.02505    |                 .  |* .                 |
 18        0.01620    |                 .  |  .                 |
 19       -0.00012    |                 .  |  .                 |
 20        0.03095    |                 .  |* .                 |
 21        0.02301    |                 .  |  .                 |
 22        0.02393    |                 .  |  .                 |
 23        0.00786    |                 .  |  .                 |
 24        0.01425    |                 .  |  .                 |
 25       -0.24382    |               *****|  .                 |
 26        0.01064    |                 .  |  .                 |
 27       -0.01865    |                 .  |  .                 |
 28        0.01931    |                 .  |  .                 |
 29        0.01201    |                 .  |  .                 |
 30        0.00269    |                 .  |  .                 |
 31        0.00312    |                 .  |  .                 |
 32       -0.00752    |                 .  |  .                 |
 33        0.01229    |                 .  |  .                 |
 34        0.02577    |                 .  |* .                 |
 35        0.03366    |                 .  |* .                 |
 36        0.01177    |                 .  |  .                 |
```

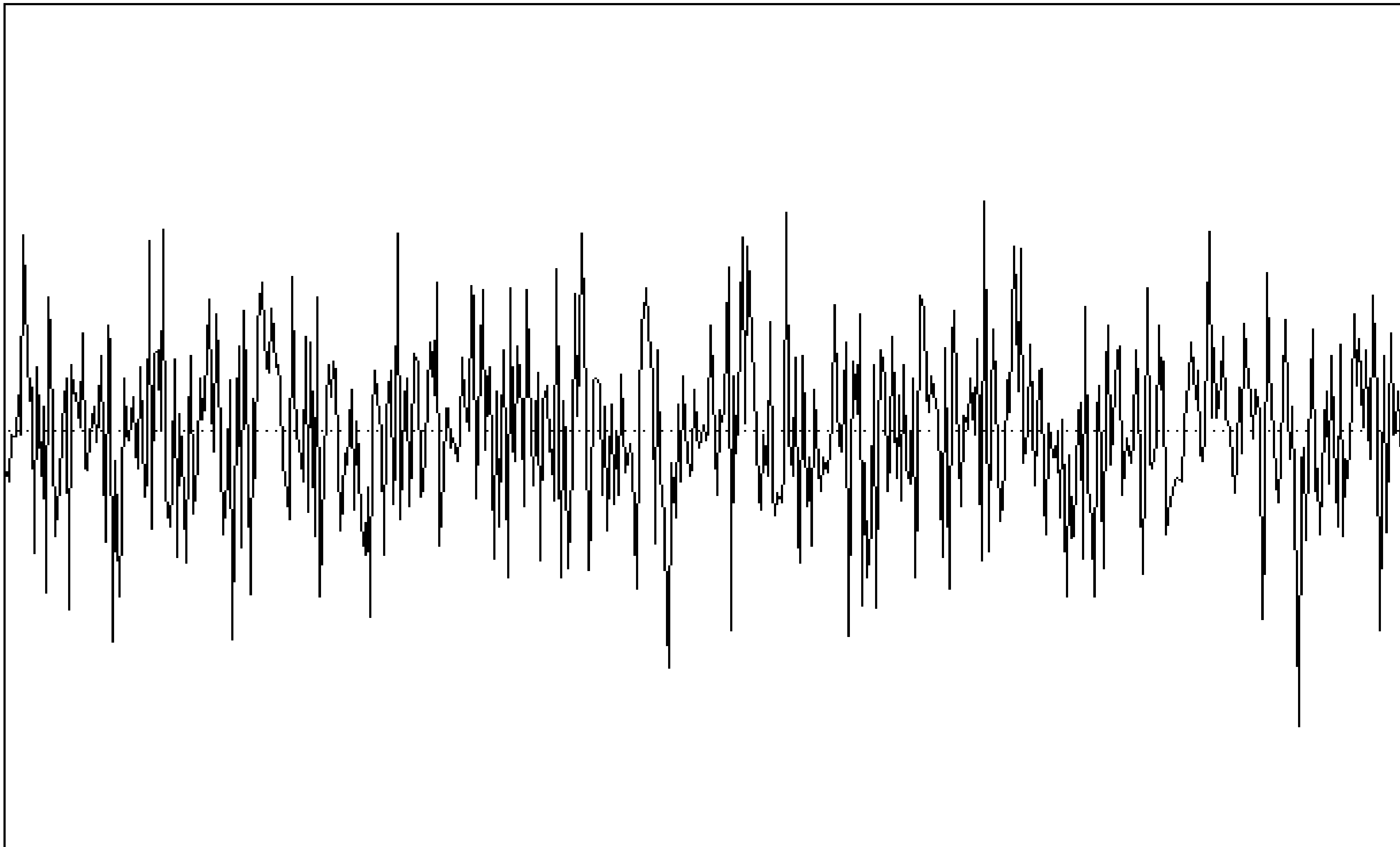
Back

Continue

Assessing Stationarity

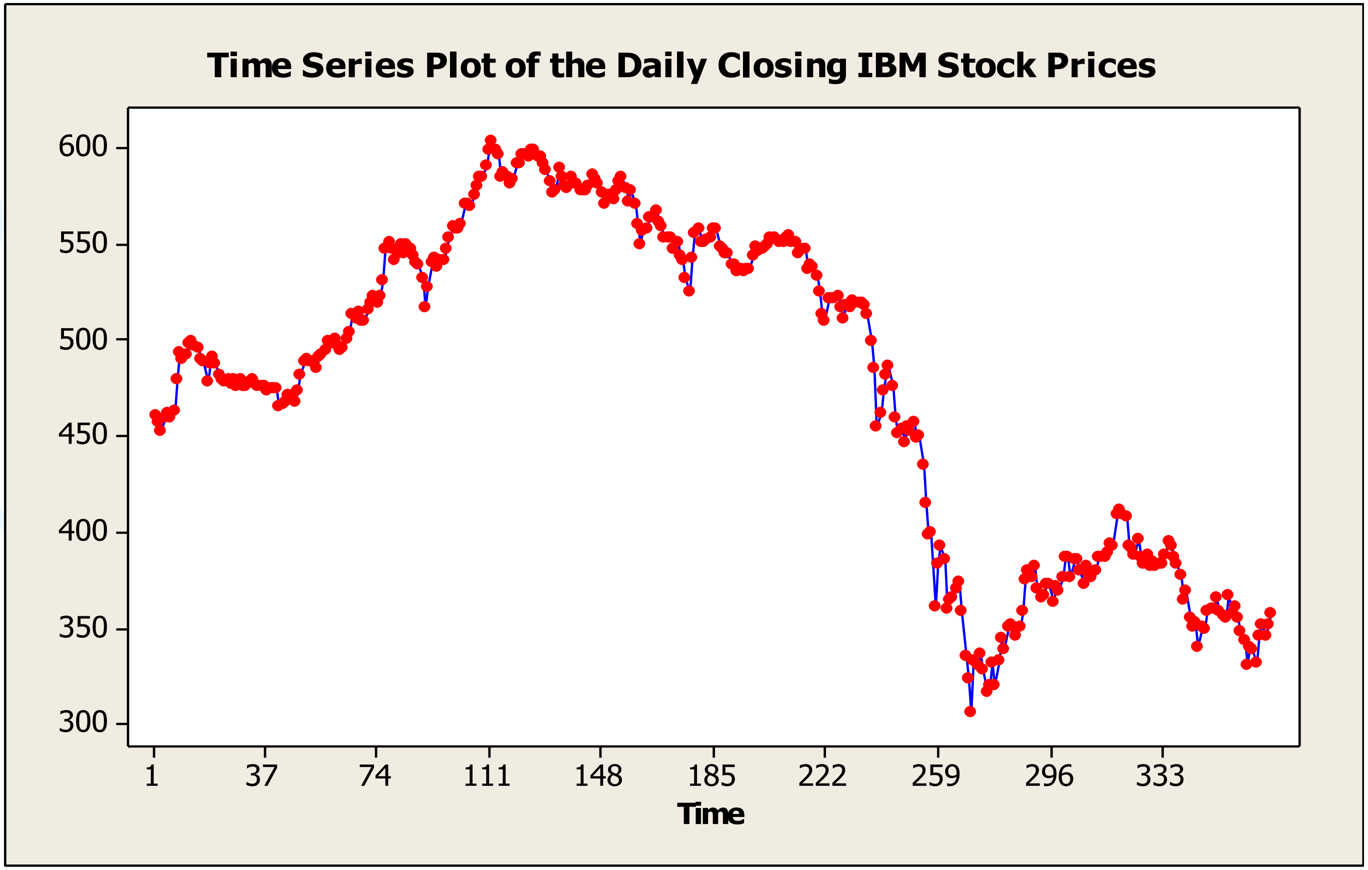
- Using a time series plot
  - If there is no evidence of a change in the mean over time, the series is stationary in the mean.
  - If the plotted series shows no obvious change in the variance over time, it is stationary in the variance.

Example: Stationary Series



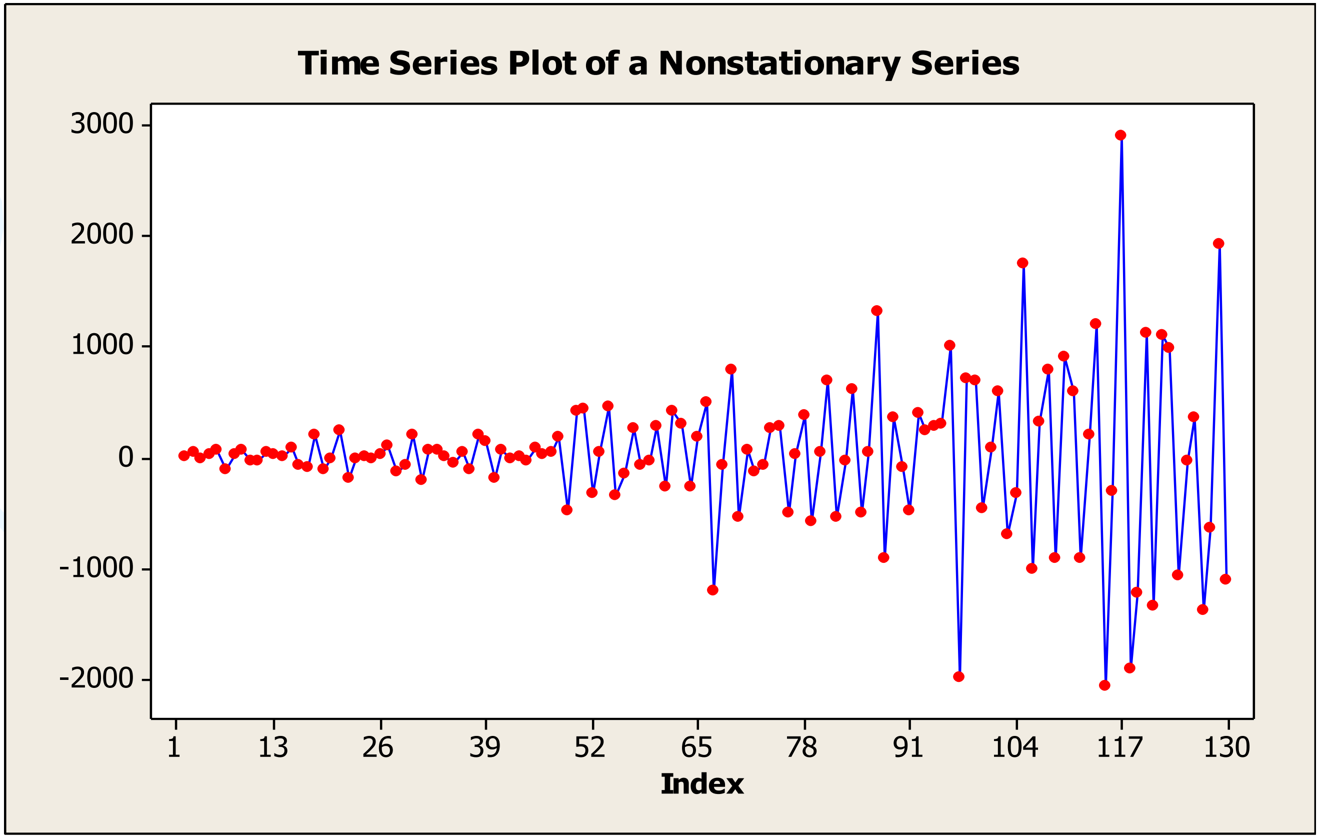
Example: Non-stationary Series

- Non-stationary in the mean



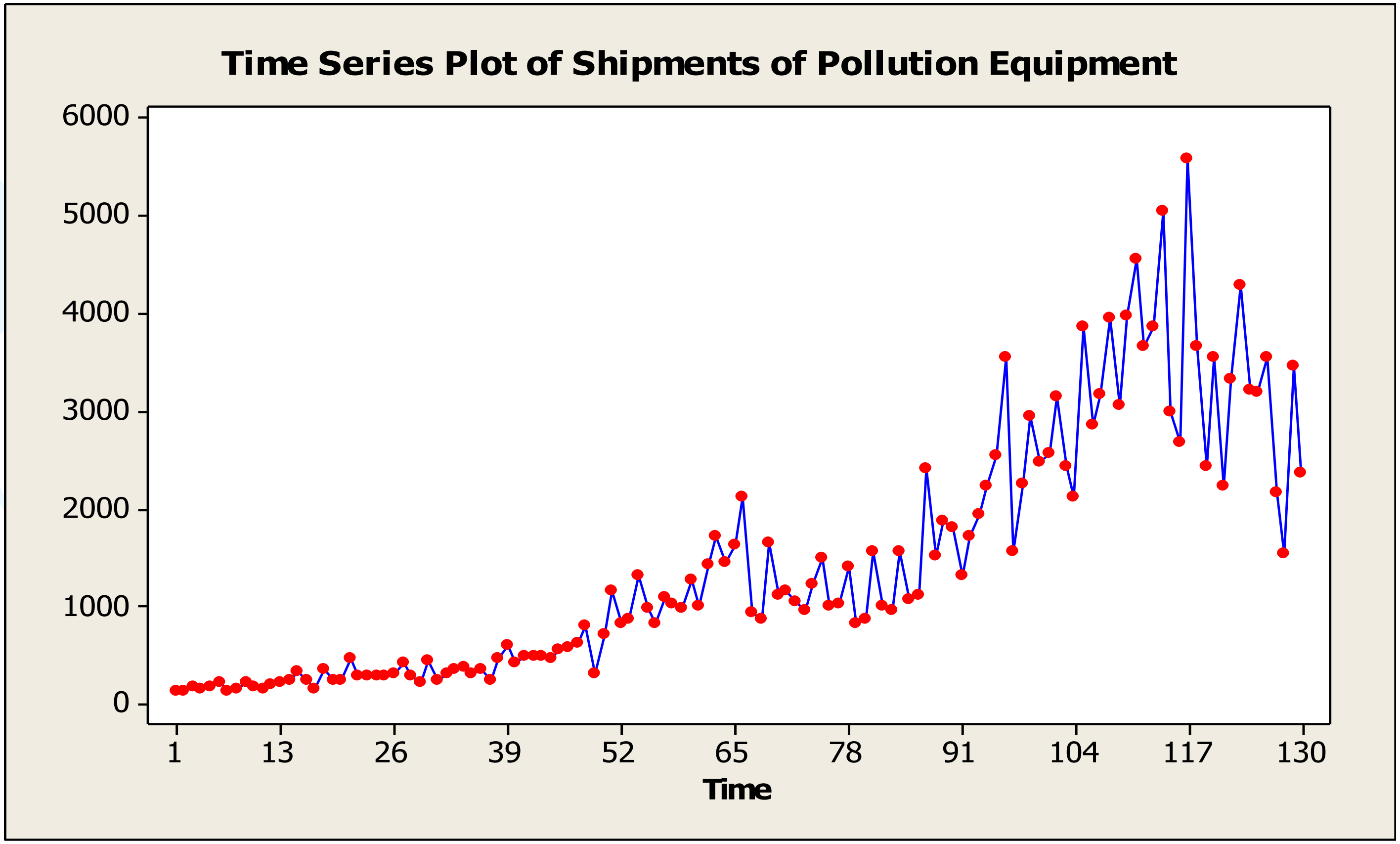
Example: Non-stationary Series

- Non-stationary in the variance



Example: Non-stationary Series

- Non-stationary in both the mean and the variance



Assessing Stationarity

- A time series plot is often enough to convince a forecaster that the data are stationary or non-stationary. The ACF and PACF plots can readily expose non-stationarity in the mean.
  - The autocorrelations of stationary data drop to zero relatively quickly.
  - For a typical pattern of non-stationary series
    - ρ_1 is very large and positive (ACF plot).
    - ρ_k's are relatively large and positive, until k gets big enough (ACF plot).
    - PACF plot displays a large spike close to 1 at lag 1.

Example: Employees

Remove Nonstationarity

- The use of Box-Jenkins modeling techniques requires a stationary process.
- "Pre-differencing transformation" is often used to stabilize the seasonal variation of the time series. A common transformation is of the form:

$$y_t^* = \ell n(y_t)$$

Remove Nonstationarity

- Regular differencing
  - Let y_t^* represent an appropriate predifferencing transformation.
  - If y_t^* exhibits the pattern of nonstationarity, "first regular differencing" can be applied.

$$z_t = y_t^* - y_{t-1}^*$$

Example: Employees

Example: Employees

- Is "pre-differencing transformation" necessary?
- Take first differences

Period(s) of Differencing	1
Mean of Working Series	0.902825
Standard Deviation	7.210001
Number of Observations	177
Observation(s) eliminated by differencing	1

Example: Employees

```
                                         Autocorrelations

Lag   Covariance  Correlation  -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1   Std Error

  0    51.984116    1.00000    |                    |********************|          0
  1     1.341360    0.02580    |                 .  |* .                 |   0.075165
  2   -10.104648    -.19438    |                ****|  .                 |   0.075215
  3   -16.397040    -.31542    |              ******|  .                 |   0.078001
  4    -6.537721    -.12576    |                 ***|  .                 |   0.084902
  5     0.720104    0.01385    |                 .  |  .                 |   0.085948
  6    11.646511    0.22404    |                 .  |****                |   0.085961
  7     0.382655    0.00736    |                .   |   .                |   0.089199
  8    -5.583873    -.10741    |                . **|   .                |   0.089202
  9   -15.804044    -.30402    |              ******|   .                |   0.089930
 10    -9.291756    -.17874    |                ****|   .                |   0.095560
 11     2.139864    0.04116    |                .   |*  .                |   0.097431
 12    46.868231    0.90159    |                .   |******************  |   0.097529
 13     0.801322    0.01541    |               .    |    .               |   0.136736
 14    -9.690318    -.18641    |               .****|    .               |   0.136746
 15   -15.285807    -.29405    |              ******|     .              |   0.138174
 16    -6.236594    -.11997    |              .   **|     .              |   0.141665
 17     0.881801    0.01696    |              .     |     .              |   0.142238
 18    10.680823    0.20546    |              .     |**** .              |   0.142250
 19     0.496121    0.00954    |              .     |     .              |   0.143917
 20    -4.968756    -.09558    |              .   **|     .              |   0.143920
 21   -14.320935    -.27549    |              ******|     .              |   0.144278
 22    -8.286359    -.15940    |              .  ***|     .              |   0.147220
 23     1.685671    0.03243    |              .     |*    .              |   0.148192
 24    42.361435    0.81489    |              .     |****************    |   0.148232
 25     0.121192    0.00233    |             .      |      .             |   0.171686
 26    -9.063339    -.17435    |             .   ***|      .             |   0.171686
 27   -13.777984    -.26504    |             . *****|      .             |   0.172684
 28    -6.266858    -.12055    |             .    **|      .             |   0.174967
 29     1.370233    0.02636    |             .      |*     .             |   0.175436
 30    10.070043    0.19371    |             .      |****  .             |   0.175458
 31     0.413689    0.00796    |             .      |      .             |   0.176662
 32    -4.482652    -.08623    |             .    **|      .             |   0.176664
 33   -13.450155    -.25874    |             . *****|      .             |   0.176902
 34    -7.366474    -.14171    |             .   ***|      .             |   0.179027
 35     1.845753    0.03551    |             .      |*     .             |   0.179660
 36    39.150698    0.75313    |             .      |***************     |   0.179699

                                 "." marks two standard errors
```

Example: Employees

```
                         Partial Autocorrelations

 Lag   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1       0.02580    |                 .  |* .                 |
   2      -0.19518    |                ****|  .                 |
   3      -0.31666    |              ******|  .                 |
   4      -0.18893    |                ****|  .                 |
   5      -0.14977    |                 ***|  .                 |
   6       0.05427    |                 .  |* .                 |
   7      -0.11691    |                 .**|  .                 |
   8      -0.14022    |                 ***|  .                 |
   9      -0.34980    |             *******|  .                 |
  10      -0.42848    |           *********|  .                 |
  11      -0.53214    |         ***********|  .                 |
  12       0.75935    |                 .  |***************     |
  13      -0.05378    |                 . *|  .                 |
  14       0.03904    |                 .  |* .                 |
  15       0.11033    |                 .  |**.                 |
  16       0.08907    |                 .  |**.                 |
  17       0.00483    |                 .  |  .                 |
  18      -0.06527    |                 . *|  .                 |
  19       0.04622    |                 .  |* .                 |
  20      -0.04072    |                 . *|  .                 |
  21       0.06118    |                 .  |* .                 |
  22       0.01489    |                 .  |  .                 |
  23      -0.11010    |                 .**|  .                 |
  24       0.01351    |                 .  |  .                 |
  25      -0.05537    |                 . *|  .                 |
  26      -0.02404    |                 .  |  .                 |
  27      -0.01628    |                 .  |  .                 |
  28      -0.06433    |                 . *|  .                 |
  29       0.02361    |                 .  |  .                 |
  30      -0.00076    |                 .  |  .                 |
  31      -0.01543    |                 .  |  .                 |
  32      -0.04057    |                 . *|  .                 |
  33      -0.06671    |                 . *|  .                 |
  34      -0.01294    |                 .  |  .                 |
  35      -0.00254    |                 .  |  .                 |
  36       0.07857    |                 .  |**.                 |
```

Seasonal Differencing

- With seasonal data which is non-stationary, it may be appropriate to take seasonal differences.
- A seasonal difference is the difference between an observation and the corresponding observation from the previous year.

  $$z_t = y_t^* - y_{t-L}^*$$

  where s is the number of seasons. For monthly data, *L*=12, for quarterly data, *L*=4, and so on.
- The differencing can be repeated to obtain second-order seasonal differencing, although this is rarely needed.

Example: Employees

- Non-stationarity in the mean remains in the seasonal level. Try to remove it with a further first seasonal difference.

$$z_t = (y_t^* - y_{t-1}^*) - (y_{t-L}^* - y_{t-1-L}^*) = (y_t^* - y_{t-1}^*) - (y_{t-12}^* - y_{t-13}^*)$$

Example: Employees

```
Period(s) of Differencing                        1,12
Mean of Working Series                       0.087273
Standard Deviation                           1.438735
Number of Observations                            165
Observation(s) eliminated by differencing          13

                         Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Er

  0      2.069959        1.00000    |                    |********************|
  1      0.380397        0.18377    |                 .  |****                |       0.077
  2     -0.056837        -.02746    |                 . *|  .                 |       0.080
  3     -0.021478        -.01038    |                 .  |  .                 |       0.080
  4     -0.290834        -.14050    |                 ***|  .                 |       0.080
  5    -0.0045074        -.00218    |                 .  |  .                 |       0.081
  6      0.200142        0.09669    |                 .  |**.                 |       0.081
  7      0.041474        0.02004    |                 .  |  .                 |       0.082
  8      0.187094        0.09039    |                 .  |**.                 |       0.082
  9      0.197702        0.09551    |                 .  |**.                 |       0.083
 10    0.00045630        0.00022    |                 .  |  .                 |       0.083
 11     -0.144889        -.07000    |                 . *|  .                 |       0.083
 12     -0.572732        -.27669    |              ******|  .                 |       0.084
 13     -0.200208        -.09672    |                . **|   .                |       0.089
 14      0.056730        0.02741    |                .   |*  .                |       0.090
 15     0.0061858        0.00299    |                .   |   .                |       0.090
 16      0.287759        0.13902    |                .   |***.                |       0.090
 17      0.049923        0.02412    |                .   |   .                |       0.091
 18     -0.209991        -.10145    |                . **|   .                |       0.091
 19     -0.198252        -.09578    |                . **|   .                |       0.092
 20     -0.113819        -.05499    |                .  *|   .                |       0.092
 21     -0.039443        -.01906    |                .   |   .                |       0.093
 22     -0.039793        -.01922    |                .   |   .                |       0.093
 23      0.106062        0.05124    |                .   |*  .                |       0.093
 24     -0.165247        -.07983    |                . **|   .                |       0.093

                       "." marks two standard errors
```

Example: Employees

```
                         Partial Autocorrelations

 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1        0.18377    |                 .  |****                |
   2       -0.06337    |                 . *|  .                 |
   3        0.00689    |                 .  |  .                 |
   4       -0.14706    |                 ***|  .                 |
   5        0.05599    |                 .  |* .                 |
   6        0.07694    |                 .  |**.                 |
   7       -0.00961    |                 .  |  .                 |
   8        0.08102    |                 .  |**.                 |
   9        0.07084    |                 .  |* .                 |
  10        0.00110    |                 .  |  .                 |
  11       -0.07361    |                 . *|  .                 |
  12       -0.25948    |               *****|  .                 |
  13        0.01555    |                 .  |  .                 |
  14        0.00615    |                 .  |  .                 |
  15       -0.03042    |                 . *|  .                 |
  16        0.09184    |                 .  |**.                 |
  17       -0.01900    |                 .  |  .                 |
  18       -0.04354    |                 . *|  .                 |
  19       -0.08056    |                 .**|  .                 |
  20        0.03107    |                 .  |* .                 |
  21        0.03687    |                 .  |* .                 |
  22       -0.06723    |                 . *|  .                 |
  23        0.03476    |                 .  |* .                 |
  24       -0.18999    |                ****|  .                 |
```

Back

Remarks for Removing Nonstationarity

- When both seasonal and first differences are applied, it makes no difference which is done first—the result will be the same.
- If differencing is used, the differences should be interpretable. First differences are the change between one observation and next. Seasonal differences are the change from one year to the next.

Example: Employees

- Do seasonal differencing: $Z_t = y_t - y_{t-12}$

Period(s) of Differencing	12
Mean of Working Series	10.3759
Standard Deviation	5.005722
Number of Observations	166
Observation(s) eliminated by differencing	12

Example: Employees

```
                                        Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0     25.057251       1.00000     |                    |********************|             0
  1     23.551046       0.93989     |                 .  |******************* |      0.077615
  2     21.750363       0.86803     |               .    |*****************   |      0.129102
  3     19.984942       0.79757     |              .     |****************    |      0.160454
  4     18.383410       0.73366     |             .      |***************     |      0.182782
  5     17.031926       0.67972     |            .       |**************      |      0.199736
  6     15.647808       0.62448     |           .        |************        |      0.213216
  7     14.141135       0.56435     |           .        |***********         |      0.223963
  8     12.707374       0.50713     |           .        |**********          |      0.232372
  9     11.123315       0.44392     |          .         |*********.          |      0.238946
 10      9.421701       0.37601     |          .         |******** .          |      0.243864
 11      7.755107       0.30950     |          .         |******   .          |      0.247332
 12      6.024674       0.24044     |          .         |*****    .          |      0.249654
 13      5.018099       0.20027     |          .         |****     .          |      0.251045
 14      4.119250       0.16439     |          .         |***      .          |      0.252005
 15      3.165849       0.12634     |          .         |***      .          |      0.252651
 16      2.245328       0.08961     |          .         |**       .          |      0.253031
 17      1.057665       0.04221     |          .         |*        .          |      0.253222
 18     -0.103884       -.00415     |          .         |         .          |      0.253264
 19     -0.936067       -.03736     |          .        *|         .          |      0.253265
 20     -1.623877       -.06481     |          .        *|         .          |      0.253298
 21     -2.257332       -.09009     |          .       **|         .          |      0.253398
 22     -2.941722       -.11740     |          .       **|         .          |      0.253591
 23     -3.670260       -.14647     |          .      ***|         .          |      0.253918
 24     -4.472118       -.17848     |          .     ****|         .          |      0.254426

                              "." marks two standard errors
```

Example: Employees

```
                        Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1        0.93989    |                 .  |******************* |
  2       -0.13177    |                 ***|  .                 |
  3       -0.01687    |                 .  |  .                 |
  4        0.01467    |                 .  |  .                 |
  5        0.03995    |                 .  |* .                 |
  6       -0.05817    |                 . *|  .                 |
  7       -0.06967    |                 . *|  .                 |
  8        0.00167    |                 .  |  .                 |
  9       -0.09306    |                 .**|  .                 |
 10       -0.08076    |                 .**|  .                 |
 11       -0.03285    |                 . *|  .                 |
 12       -0.07481    |                 . *|  .                 |
 13        0.19780    |                 .  |****                |
 14       -0.05541    |                 . *|  .                 |
 15       -0.04737    |                 . *|  .                 |
 16       -0.00425    |                 .  |  .                 |
 17       -0.10667    |                 .**|  .                 |
 18       -0.01782    |                 .  |  .                 |
 19        0.04612    |                 .  |* .                 |
 20        0.00470    |                 .  |  .                 |
 21       -0.05160    |                 . *|  .                 |
 22       -0.06924    |                 . *|  .                 |
 23       -0.01942    |                 .  |  .                 |
 24       -0.09627    |                 .**|  .                 |
```

Example: Employees

- Is stationarity achieved after first seasonal differencing?

Stationarity

- In general, if ACF does both of the following, the series should be considered stationary.
  - Cuts off fairly quickly or dies down fairly quickly at the nonseasonal level
    - Arbitrarily define the "nonseasonal level" as lags 1 through L–3.
      - For monthly data, lags 1 through 9
      - For quarterly data, lags 1, 2, and possibly 3
  - Cuts off fairly quickly or dies down fairly quickly at the seasonal level
    - Define the "seasonal level" as lags equal to (or nearly equal to) L, $2L$, $3L$, and $4L$.
      - Define the lags L, $2L$, $3L$, and $4L$ to be "exact seasonal lags"
      - Define the lags $L - 2$, $L - 1$, $L + 1$, $L + 2$, $2L - 2$, $2L - 1$, $2L + 1$, $2L + 2$, $3L - 2$, $3L - 1$, $3L + 1$, $3L + 2$, $4L - 2$, $4L - 1$, $4L + 1$, and $4L + 2$ to be "near seasonal lags"

Example: Employees

Summary of Removing Nonstationarity

- "Pre-differencing transformation" is often used to stabilize the seasonal variation of the time series. A common transformation is of the form:

$$y_t^* = \ell n(y_t)$$

Summary of Removing Nonstationarity

- First regular difference

$$z_t = y_t^* - y_{t-1}^*$$

- First seasonal difference, where *L* is the number of seasons in a year

$$z_t = y_t^* - y_{t-L}^*$$

- First seasonal and first regular difference

$$z_t = \left(y_t^* - y_{t-1}^*\right) - \left(y_{t-L}^* - y_{t-L-1}^*\right)$$

- One can also obtain second and higher order differences by simply applying the same rule.
- For most practical purposes, a maximum of two differences (including both regular and seasonal differences) will transform the data into a stationary series.

Backshift Notation

- Backward shift operator, *B*

$$By_t = y_{t-1}$$

- *B*, operating on y_t, has the effect of shifting the data back one period.
- Two applications of *B* to y_t shifts the data back two periods.

$$B(By_t) = B^2 y_t = y_{t-2}$$

- *m* applications of *B* to y_t shifts the data back *m* periods.

$$B^m y_t = y_{t-m}$$

Backshift Notation

- The backward shift operator is convenient for describing the process of differencing.

$$\Delta y_t = y_t - y_{t-1} = y_t - By_t = (1-B)y_t$$

$$\Delta^2 y_t = y_t - 2y_{t-1} + y_{t-2} = (1-2B+B^2)y_t = (1-B)^2 y_t$$

- In general, a *d*th-order can be written as

$$\Delta^d y_t = (1-B)^d y_t$$

- An *s*th difference

$$(1-B^s)y_t$$

- The backshift notation is convenient because the terms can be multiplied together to see the combined effect.

$$(1-B)(1-B^s)y_t = (1-B-B^s+B^{s+1})y_t = y_t - y_{t-1} - y_{t-s} + y_{t-s-1}$$

Step Two: Model Identification

ARIMA models for time series data

- Three basic Box-Jenkins models for a stationary time series $\{y_t\}$
  - Autoregressive models (AR)
  - Moving Average models (MA)
  - Autoregressive Moving Average models (ARMA)
- Autoregressive Integrated Moving Average models (ARIMA)
  - ARMA models extended to nonstationary series by allowing differencing of the data series

Autoregressive Models (AR)

- Autoregressive model of order *p* (AR(*p*))

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t$$

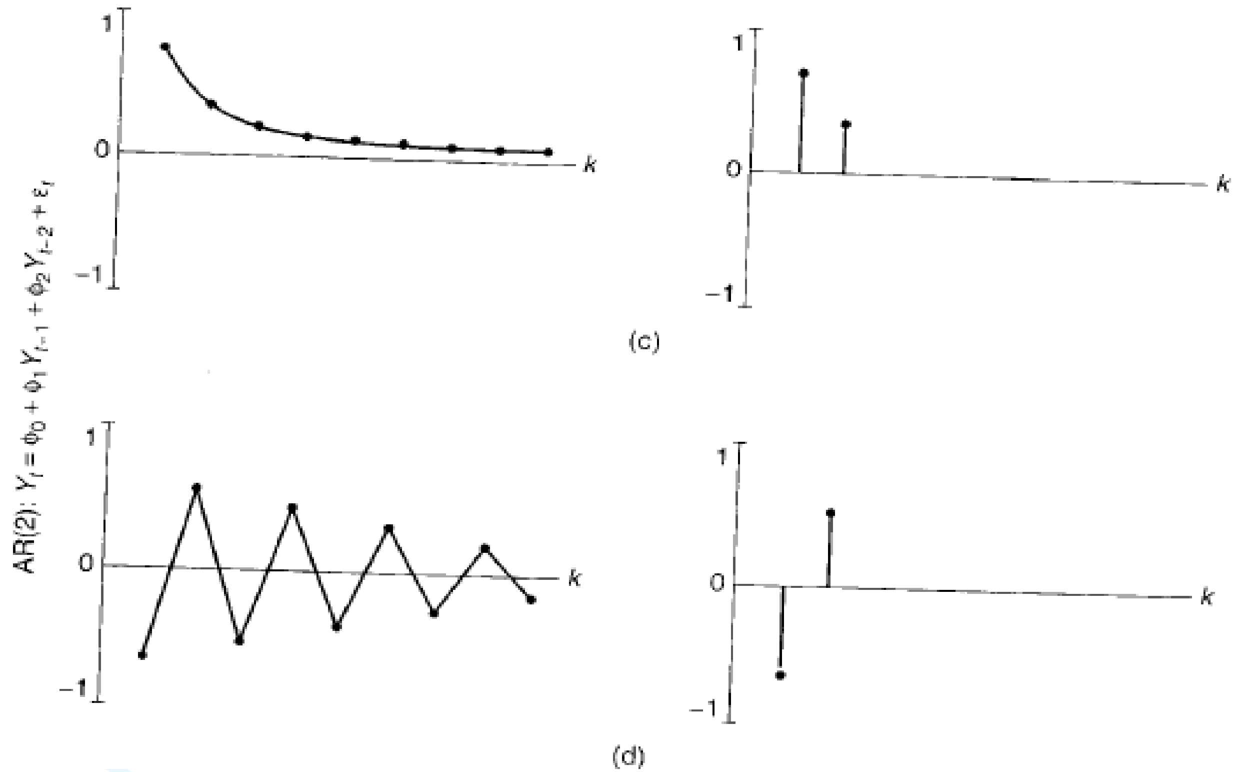
  i.e., y_t depends on its *p* previous values
- Using backshift notation

$$\phi_p(B) y_t = \delta + \varepsilon_t$$

  where

$$\phi_p(B) = 1 - \phi_1 B - \ldots - \phi_p B^p$$

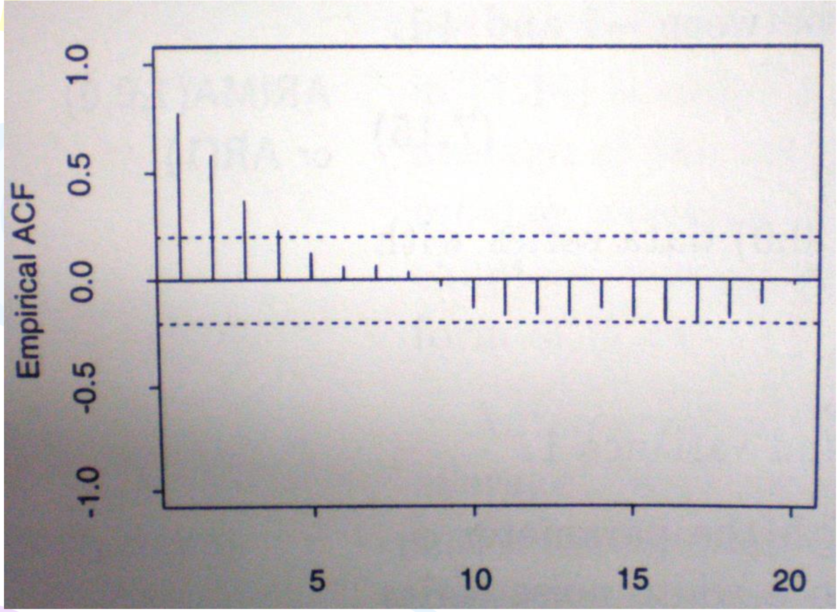
Theoretical ACF and PACF for AR(p)

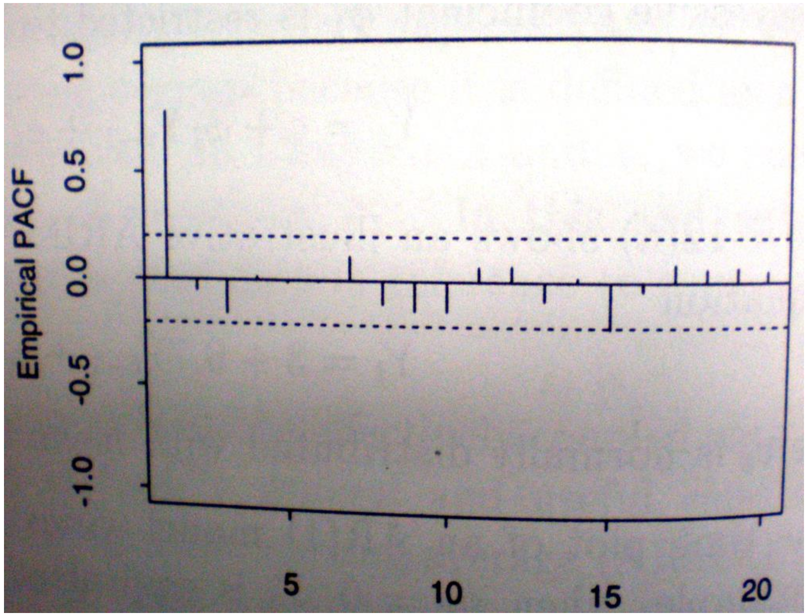


- Characteristics:
  - ACF dies down.
  - PACF cuts off after lag p.

Example for AR(1)

- Empirical ACF and PACF





$$y_t = 3 + 0.7 y_{t-1} + \varepsilon_t$$

Moving Average Models (MA)

- Moving Average model of order q (MA(q))

$$y_t = \mu + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

i.e., y_t depends on q previous random error terms.

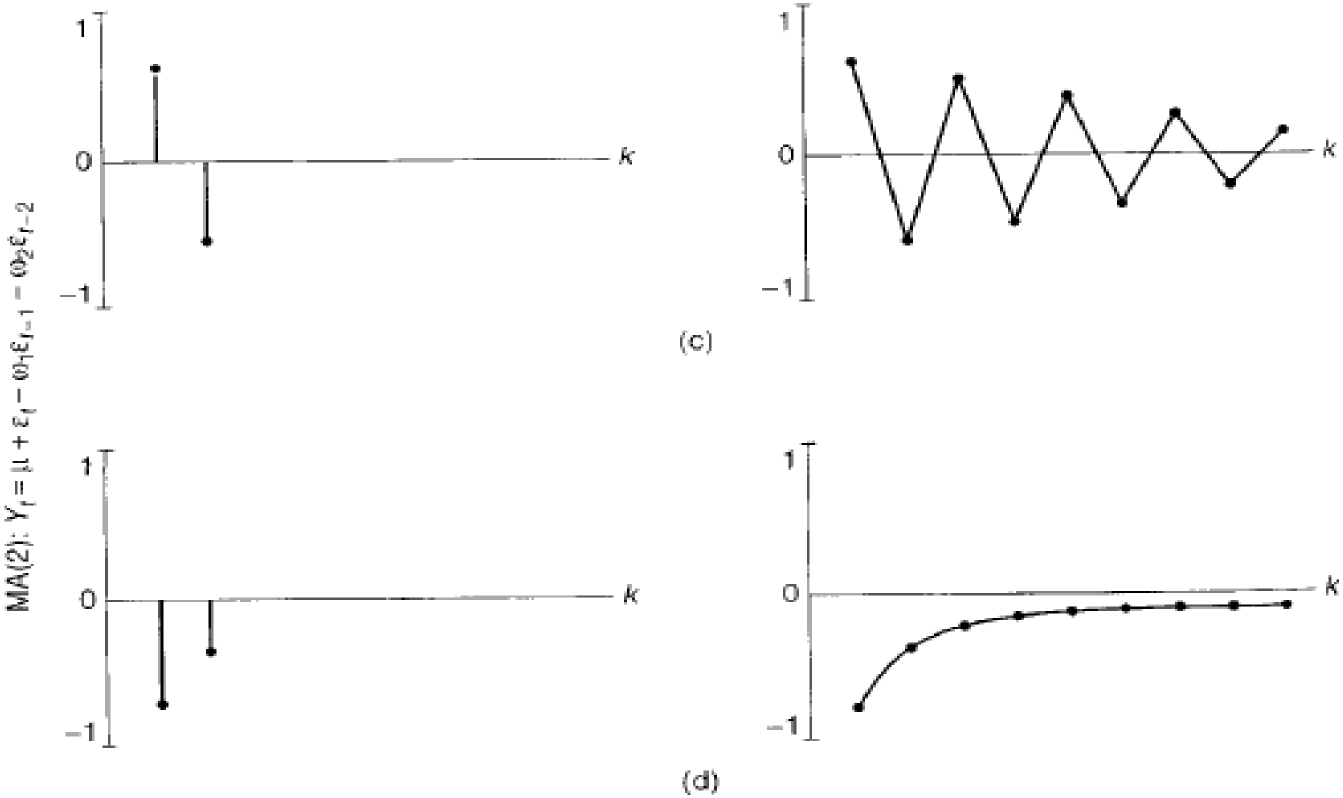
- Using backshift notation

$$y_t = \mu + \theta_q(B)\varepsilon_t$$

where

$$\theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \theta_q B^q$$

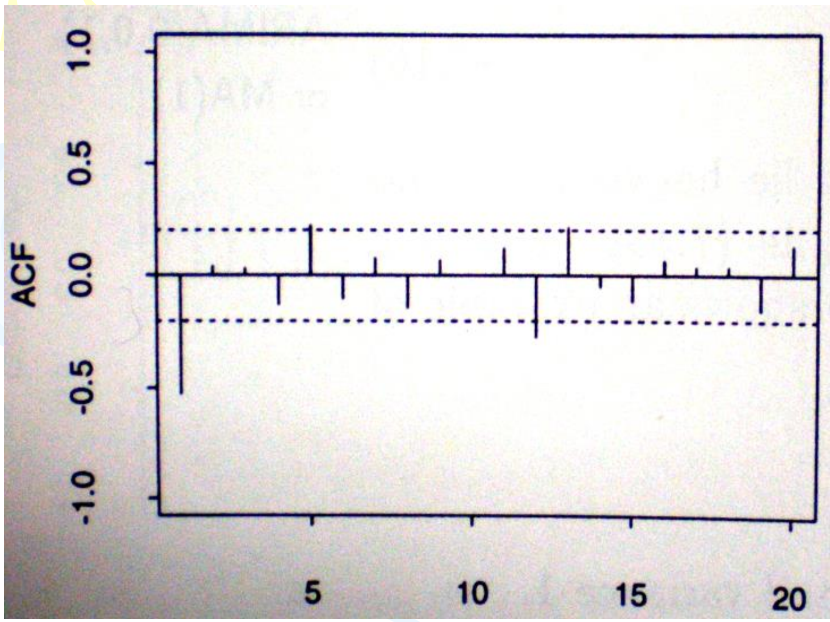
Theoretical ACF and PACF for MA(q)

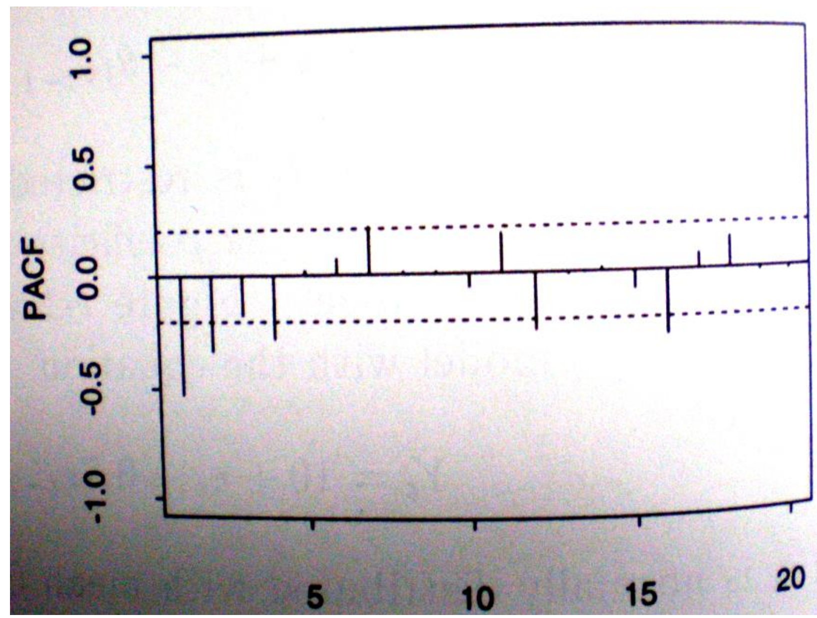


- Characteristics:
  - ACF cuts off after lag q.
  - PACF dies down.

Example for MA(1)

- Empirical ACF and PACF





$$y_t = 10 + \varepsilon_t - 0.7\varepsilon_{t-1}$$

Autoregressive Moving Average Models (ARMA)

- Autoregressive-moving average model of order *p* and *q* (ARMA(*p*,*q*))

$$y_t = \delta + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \cdots - \theta_q \varepsilon_{t-q}$$

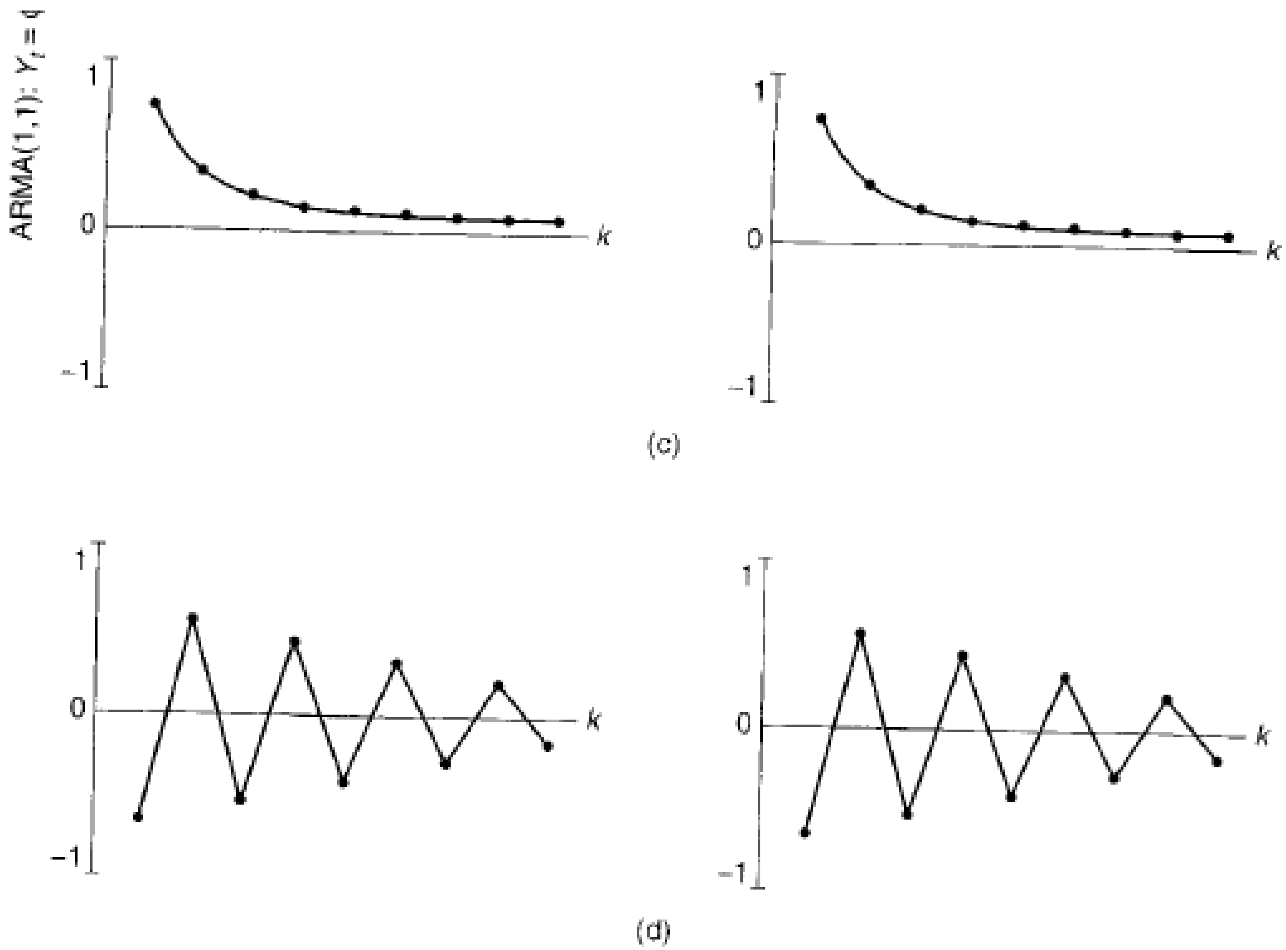
  i.e., y_t depends on its *p* previous values and *q* previous random error terms

- Using backshift notation

$$\phi_p(B) y_t = \delta + \theta_q(B) \varepsilon_t$$

- When *q* = 0, the ARMA(*p*,0) model reduces to AR(*p*); when *p* = 0, the ARMA(0,*q*) model reduces to MA(*q*).

Theoretical ACF and PACF for ARMA(p,q)



- Characteristics:
  - Both ACF and PACF die down.

Summary of the Behaviors of ACF and PACF

Behaviors of ACF and PACF for general non-seasonal models

Process	ACF	PACF
AR(p)	Dies down.	Cuts off after lag p.
MA(q)	Cuts off after lag q.	Dies down.
ARMA(p,q)	Dies down.	Dies down.

Non-seasonal Autoregressive Integrated Moving Average models (ARIMA)

- ARMA models can only be used for stationary data. This class of models can be extended to non-stationary series by allowing differencing the data series. ⇒ ARIMA models
- Backshift notation

$$\phi_p(B)(1-B)^d y_t = \delta + \theta_q(B)\varepsilon_t$$

and $\delta = \mu\phi_p(B)$

e.g. ARIMA(1,1,1)

$$(1-\phi_1 B)(1-B)y_t = \delta + (1-\theta_1 B)\varepsilon_t$$

- The general non-seasonal model: ARIMA(p,d,q)
  - AR: p = order of the autoregressive part
  - I: d = order of integration
  - MA: q = order of the moving average part

Seasonality and ARIMA Models

- One final complexity to add to ARIMA models is seasonality.
- ARIMA notation: ARIMA$(p,d,q)(P,D,Q)_L$
  where L = number of periods per season.

$$\text{ARIMA}\ \underbrace{(p,d,q)}_{\left(\substack{\text{Non-seasoal}\\ \text{part of the}\\ \text{model}}\right)}\ \underbrace{(P,D,Q)_L}_{\left(\substack{\text{Seasoal}\\ \text{part of}\\ \text{the model}}\right)}$$

- Suppose y_t^* is a pre-differencing transformed series. The general stationarity transformation is

$$z_t = \Delta_L^D \Delta^d y_t^* = (1-B^L)^D (1-B)^d y_t^*$$

where D is the order of seasonal differencing and d is the order of non-seasonal differencing.

Seasonality and ARIMA Models

- Using backshift notation

$$\phi_p(B)\,\Phi_p(B^L)z_t = \delta + \theta_q(B)\,\Theta_Q(B^L)\,\varepsilon_t$$

where

$$\phi_p(B) = (1 - \phi_1 B - \phi_2 B^2 - \ldots - \phi_p B^p)$$

is the non-seasonal autoregressive operator of order p,

$$\Phi_P(B^L) = (1 - \Phi_1 B^L - \Phi_2 B^{2L} - \ldots - \Phi_P B^{PL})$$

is the seasonal autoregressive operator of order P,

$$\theta_q(B) = (1 - \theta_1 B - \theta_2 B^2 - \ldots - \theta_q B^q)$$

is the non-seasonal moving average operator of order q,

$$\Theta_Q(B^L) = (1 - \Theta_1 B^L - \Theta_2 B^{2L} - \ldots - \Theta_Q B^{QL})$$

is the seasonal moving average operator of order Q,

and $$\delta = \mu\phi_p(B)\,\Phi_P(B^L)$$

- For example, ARIMA$(1,1,1)(1,1,1)_4$ model:

$$(1 - \phi_1 B)(1 - \Phi_1 B^4)(1 - B)(1 - B^4)y_t = \delta + (1 - \theta_1 B)(1 - \Theta_1 B^4)\varepsilon_t$$

Seasonality and ARIMA Models

- The seasonal part of an AR or MA model will be seen in the seasonal lags of the ACF and PACF.
  - For example, ARIMA(0,0,0)$(0,0,1)_{12}$ will show a spike at lag 12 in the ACF but no other significant spikes. The PACF will show exponential decay in the seasonal lags, i.e., lags 12, 24, 36,…
  - Similarly, ARIMA(0,0,0)$(1,0,0)_4$ will show exponential decay in the seasonal lags of the ACF, and a single significant spike at lag 4 in the PACF.

Summary of the First Two Steps

1. Plot the data.
   - Decide if a pre-differencing transformation is necessary to stabilize the seasonal variation. If necessary, transform the data to achieve stationarity in the variance. The main approach is through a logarithmic transformation of the data.
2. Consider if the (possibly transformed) data appear stationary from the time plot and the ACF.
   - Time plot shows the data scattered horizontally around a constant mean and the ACF drops to near zero quickly.
   - If the time plot is not horizontal, or the ACF does not drop to zero quickly, non-stationarity is implied.

Summary of the First Two Steps

(continued)

3. When the data appear non-stationary, remove nonstationarity by differencing.
   - Check whether the data appear stationary in the non-seasonal level. If not, take the first differences of the data.
   - Check whether the data appear stationary in the seasonal level. If not, take the seasonal differences of the data.
   - For most practical purposes, a maximum of two differences will transform the data into a stationary series.

Summary of the First Two Steps

(continued)

4. When stationarity is achieved, examine the autocorrelation to see if any pattern remains.
   – The patterns of ACF and PACF may suggest a possible model at both seasonal level and non-seasonal level.
   – If there is no clear MA or AR model suggested, a mixture model may be necessary.

Example 1: Industry Sales

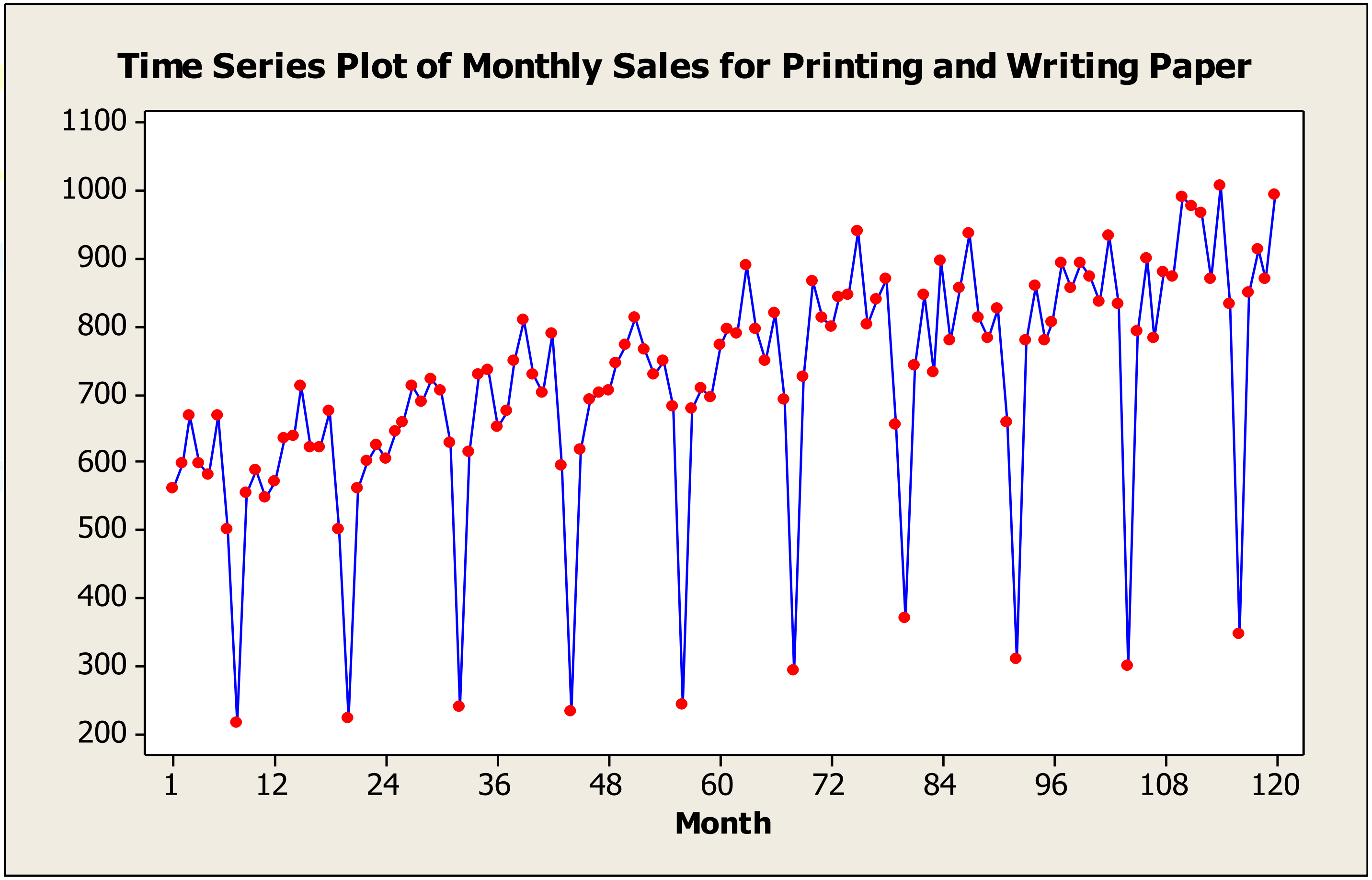
- This example analyzes the monthly industry sales for printing and writing paper between the years 1963 and 1972.
  - Dataset: writing.dat
  - SAS program: ARIMA_paper.sas

SAS program of Example 1

- **data** paper;
- infile 'E:\teaching\MS6217\Box-Jenkins Methodology\writing.dat' DLM='09'X;
- input date r;
- title 'Time Series Plot for Monthly Sales of Writing Paper';
- **proc print** data=paper;
- **run**;
- symbol1 i=join;
- **proc gplot** data=paper;
- plot r*date;
- **run**;
- **proc arima** data=paper out=paperout;
- identify var=r;
- identify var=r(**1**,**12**);
- estimate q=(**1**)(**12**) plot method=ml noconstant;
- forecast id=date;
- **run**;
- **proc print** data=paperout;
- var date r forecast residual u95 l95;
- **run**;

Example 1: Industry Sales

(continued)



Example 1: Industry Sales

(continued)

```
                                              Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0     30041.142        1.00000    |                    |********************|              0
  1     13128.669        0.43702    |                .   |*********           |       0.091287
  2      5032.062        0.16751    |                .   |***.                |       0.107315
  3      7980.640        0.26566    |                .   |*****               |       0.109472
  4      4744.480        0.15793    |               .    |*** .               |       0.114719
  5      2004.329        0.06672    |               .    |*   .               |       0.116516
  6      2946.138        0.09807    |               .    |**  .               |       0.116834
  7      2002.376        0.06665    |               .    |*   .               |       0.117518
  8      4628.363        0.15407    |               .    |*** .               |       0.117833
  9      6745.155        0.22453    |               .    |****.               |       0.119500
 10      3566.337        0.11872    |               .    |**  .               |       0.122965
 11     10204.192        0.33967    |               .    |*******             |       0.123917
 12     23673.899        0.78805    |               .    |****************    |       0.131447
 13      9811.962        0.32662    |             .      |*******             |       0.166219
 14      2363.778        0.07868    |             .      |**    .             |       0.171484
 15      4643.380        0.15457    |             .      |***   .             |       0.171784
 16      2062.207        0.06865    |             .      |*     .             |       0.172939
 17      -181.901        -.00606    |             .      |      .             |       0.173166
 18       644.661        0.02146    |             .      |      .             |       0.173168
 19      -418.796        -.01394    |             .      |      .             |       0.173190
 20      1761.975        0.05865    |             .      |*     .             |       0.173200
 21      3521.549        0.11722    |             .      |**    .             |       0.173365
 22      1072.664        0.03571    |             .      |*     .             |       0.174024
 23      6982.004        0.23241    |             .      |***** .             |       0.174085
 24     18887.134        0.62871    |             .      |*************       |       0.176652
 25      7105.623        0.23653    |            .       |*****  .            |       0.194406
 26       139.948        0.00466    |            .       |       .            |       0.196790
 27      2071.211        0.06895    |            .       |*      .            |       0.196791
 28      -116.533        -.00388    |            .       |       .            |       0.196992
 29     -2561.166        -.08526    |            .     **|       .            |       0.196993
 30     -1148.110        -.03822    |            .      *|       .            |       0.197300
 31     -1494.620        -.04975    |            .      *|       .            |       0.197362
 32       345.171        0.01149    |            .       |       .            |       0.197466
 33      2108.430        0.07018    |            .       |*      .            |       0.197472
 34      -185.315        -.00617    |            .       |       .            |       0.197680
 35      5193.412        0.17288    |            .       |***    .            |       0.197681
 36     15559.649        0.51794    |            .       |**********          |       0.198937
```

Example 1: Industry Sales

(continued)

Partial Autocorrelations

```
Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1
  1        0.43702    |               .   |*********          |
  2       -0.02903    |               .  *|   .               |
  3        0.25115    |               .   |*****              |
  4       -0.06529    |               .  *|   .               |
  5        0.01804    |               .   |   .               |
  6        0.02680    |               .   |*  .               |
  7       -0.00963    |               .   |   .               |
  8        0.16955    |               .   |***.               |
  9        0.09715    |               .   |** .               |
 10       -0.02731    |               .  *|   .               |
 11        0.35138    |               .   |*******            |
 12        0.68864    |               .   |**************     |
 13       -0.38726    |           ********|   .               |
 14       -0.13079    |               .***|   .               |
 15       -0.19058    |               ****|   .               |
 16        0.01961    |               .   |   .               |
 17       -0.02804    |               .  *|   .               |
 18       -0.03943    |               .  *|   .               |
 19       -0.09296    |               . **|   .               |
 20       -0.09448    |               . **|   .               |
 21       -0.00628    |               .   |   .               |
 22        0.05835    |               .   |*  .               |
 23        0.05340    |               .   |*  .               |
 24        0.12646    |               .   |***.               |
 25       -0.06074    |               .  *|   .               |
 26       -0.01692    |               .   |   .               |
 27       -0.02145    |               .   |   .               |
 28        0.01062    |               .   |   .               |
 29       -0.07732    |               . **|   .               |
 30        0.04128    |               .   |*  .               |
 31        0.04507    |               .   |*  .               |
 32        0.01637    |               .   |   .               |
 33        0.02704    |               .   |*  .               |
 34       -0.05001    |               .  *|   .               |
 35        0.03071    |               .   |*  .               |
 36       -0.00168    |               .   |   .               |
```

Example 1: Industry Sales

(continued)

- What patterns can you see from the plots?
- How should you deal with the data?

Example 1: Industry Sales

(continued)

```
Period(s) of Differencing                          12
Mean of Working Series                       35.60183
Standard Deviation                            50.5778
Number of Observations                            108
Observation(s) eliminated by differencing          12

                           Autocorrelations

Lag    Covariance    Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      2558.114      1.00000     |                    |********************|             0
  1       195.992      0.07662     |                .   |** .                |      0.096225
  2       432.944      0.16924     |                .   |***.                |      0.096788
  3       500.149      0.19551     |                .   |****                |      0.099491
  4       327.943      0.12820     |                .   |***.                |      0.102987
  5     72.191703      0.02822     |                .   |*  .                |      0.104454
  6       307.173      0.12008     |                .   |** .                |      0.104525
  7       349.566      0.13665     |                .   |***.                |      0.105794
  8       135.242      0.05287     |                .   |*  .                |      0.107416
  9       196.161      0.07668     |                .   |** .                |      0.107657
 10     78.023328      0.03050     |                .   |*  .                |      0.108161
 11     60.459291      0.02363     |                .   |   .                |      0.108241
 12      -979.276      -.38281     |            ********|   .                |      0.108289
 13    -10.801001      -.00422     |               .    |    .               |      0.120167
 14      -294.950      -.11530     |               .  **|    .               |      0.120169
 15      -395.715      -.15469     |               . ***|    .               |      0.121189
 16      -343.554      -.13430     |               . ***|    .               |      0.123003
 17     -3.463499      -.00135     |               .    |    .               |      0.124354
 18      -393.085      -.15366     |               . ***|    .               |      0.124354
 19      -365.626      -.14293     |               . ***|    .               |      0.126100
 20      -420.682      -.16445     |               . ***|    .               |      0.127591
 21      -379.630      -.14840     |               . ***|    .               |      0.129539
 22      -156.619      -.06122     |               .   *|    .               |      0.131103
 23      -281.543      -.11006     |               .  **|    .               |      0.131368
 24      -266.424      -.10415     |               .  **|    .               |      0.132219

                    "." marks two standard errors
```

Example 1: Industry Sales

(continued)

```
                         Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1        0.07662    |                .   |**  .                |
  2        0.16434    |                .   |***.                |
  3        0.17791    |                .   |****                |
  4        0.08742    |                .   |**  .                |
  5       -0.04070    |                .  *|    .                |
  6        0.05767    |                .   |*   .                |
  7        0.10264    |                .   |**  .                |
  8        0.01533    |                .   |    .                |
  9        0.01462    |                .   |    .                |
 10       -0.03945    |                .  *|    .                |
 11       -0.01842    |                .   |    .                |
 12       -0.44173    |           *********|    .                |
 13       -0.02242    |                .   |    .                |
 14       -0.03025    |                .  *|    .                |
 15       -0.03692    |                .  *|    .                |
 16       -0.07175    |                .  *|    .                |
 17        0.05373    |                .   |*   .                |
 18       -0.02358    |                .   |    .                |
 19       -0.01792    |                .   |    .                |
 20       -0.11047    |                . **|    .                |
 21       -0.00518    |                .   |    .                |
 22        0.10012    |                .   |**  .                |
 23        0.00856    |                .   |    .                |
 24       -0.26387    |               *****|    .                |
```

Example 1: Industry Sales

(continued)

```
Period(s) of Differencing                          1,12
Mean of Working Series                         0.389673
Standard Deviation                             68.54769
Number of Observations                              107
Observation(s) eliminated by differencing            13
```

```
                                  Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      4698.785        1.00000     |                    |********************|             0
  1     -2593.215        -.55189     |         ***********|   .                |      0.096674
  2       223.699        0.04761     |               .    |*   .               |      0.122633
  3       216.012        0.04597     |               .    |*   .               |      0.122806
  4     87.521220        0.01863     |               .    |    .               |      0.122967
  5      -450.312        -.09584     |               .  **|    .               |      0.122993
  6       142.953        0.03042     |               .    |*   .               |      0.123689
  7       284.013        0.06044     |               .    |*   .               |      0.123759
  8      -319.788        -.06806     |               .   *|    .               |      0.124034
  9       184.252        0.03921     |               .    |*   .               |      0.124383
 10      -162.709        -.03463     |               .   *|    .               |      0.124498
 11      1162.223        0.24735     |               .    |*****               |      0.124588
 12     -2086.936        -.44414     |           *********|    .               |      0.129096
 13      1313.279        0.27949     |              .     |******              |      0.142664
 14      -206.928        -.04404     |              .    *|     .              |      0.147693
 15      -155.607        -.03312     |              .    *|     .              |      0.147815
 16      -287.067        -.06109     |              .    *|     .              |      0.147885
 17       625.158        0.13305     |              .     |***  .              |      0.148120
 18      -403.615        -.08590     |              .   **|     .              |      0.149233
 19       160.764        0.03421     |              .     |*    .              |      0.149694
 20      -116.622        -.02482     |              .     |     .              |      0.149767
 21      -130.188        -.02771     |              .    *|     .              |      0.149806
 22       324.875        0.06914     |              .     |*    .              |      0.149854
 23      -205.937        -.04383     |              .    *|     .              |      0.150152
 24       184.075        0.03918     |              .     |*    .              |      0.150271

                          "." marks two standard errors
```

Example 1: Industry Sales *(continued)*

```
                          Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1      -0.55189     |         ***********|   .                |
  2      -0.36953     |             *******|   .                |
  3      -0.20315     |                ****|   .                |
  4      -0.05279     |                .  *|   .                |
  5      -0.12580     |                .***|   .                |
  6      -0.15464     |                .***|   .                |
  7      -0.04740     |                .  *|   .                |
  8      -0.05675     |                .  *|   .                |
  9      -0.01424     |                .   |   .                |
 10      -0.06486     |                .  *|   .                |
 11       0.35398     |                .   |*******             |
 12      -0.13429     |                .***|   .                |
 13      -0.08475     |                . **|   .                |
 14      -0.05507     |                .  *|   .                |
 15      -0.01514     |                .   |   .                |
 16      -0.12420     |                . **|   .                |
 17      -0.05478     |                .  *|   .                |
 18      -0.09415     |                . **|   .                |
 19       0.02577     |                .   |*  .                |
 20      -0.09031     |                . **|   .                |
 21      -0.14458     |                .***|   .                |
 22      -0.06127     |                .  *|   .                |
 23       0.18887     |                .   |****                |
 24      -0.02287     |                .   |   .                |
```

Example 1: Industry Sales

(continued)

- The model is identified to be ARIMA(p,1,q)(P,1,Q)$_{12}$.
- What are the values of p, q, P, and Q?

Example 1: Industry Sales

(continued)

- PACF – exponential decay of the first few lags ⇒ nonseasonal MA(1) ⇒ p = ?, q = ?
- ACF–ρ_1 is significant ⇒ reinforcing the nonseasonal MA(1)
- ACF – ρ_{12} is significant ⇒ seasonal MA(1) ⇒ P = ?, Q = ?

- We end up with the tentative identification:

  ARIMA$(0,1,1)(0,1,1)_{12}$

or $(1-B)(1-B^{12})y_t = \delta + (1-\theta_1 B)(1-\Theta_1 B^{12})\varepsilon_t$

Example 2: Airline

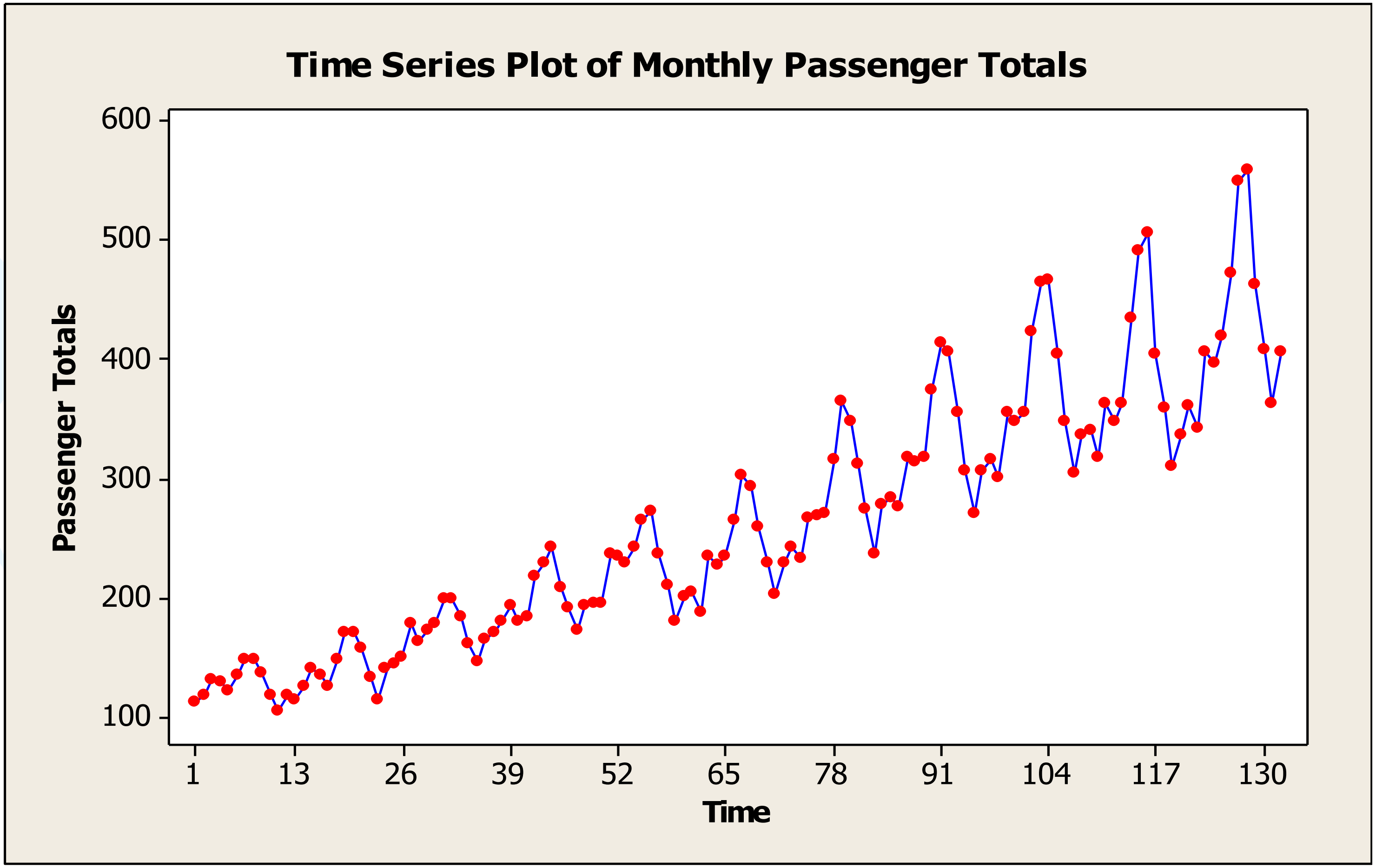
- This example considers the monthly passenger totals in international air travel for an 11-year period.
  - Dataset: airline.txt
  - SAS program: ARIMA_airline.sas

SAS program of Example 2

```
data airline;
infile 'E:\teaching\MS6217\Box-Jenkins Methodology\airline.txt' DLM='09'X;
input date r;
lr=log(r);
title 'Time Series Plot for Logarithms of Monthly Airline Passengers';
proc print data=airline;
run;
symbol1 i=join;
proc gplot data=airline;
plot lr*date;
run;
proc arima data=airline;
identify var=lr;
identify var=lr(1);
identify var=lr(1,12);
estimate q= (1)(12) plot method=ml noconstant;
run;
```

Example 2: Airline

(continued)



Example 2: Airline

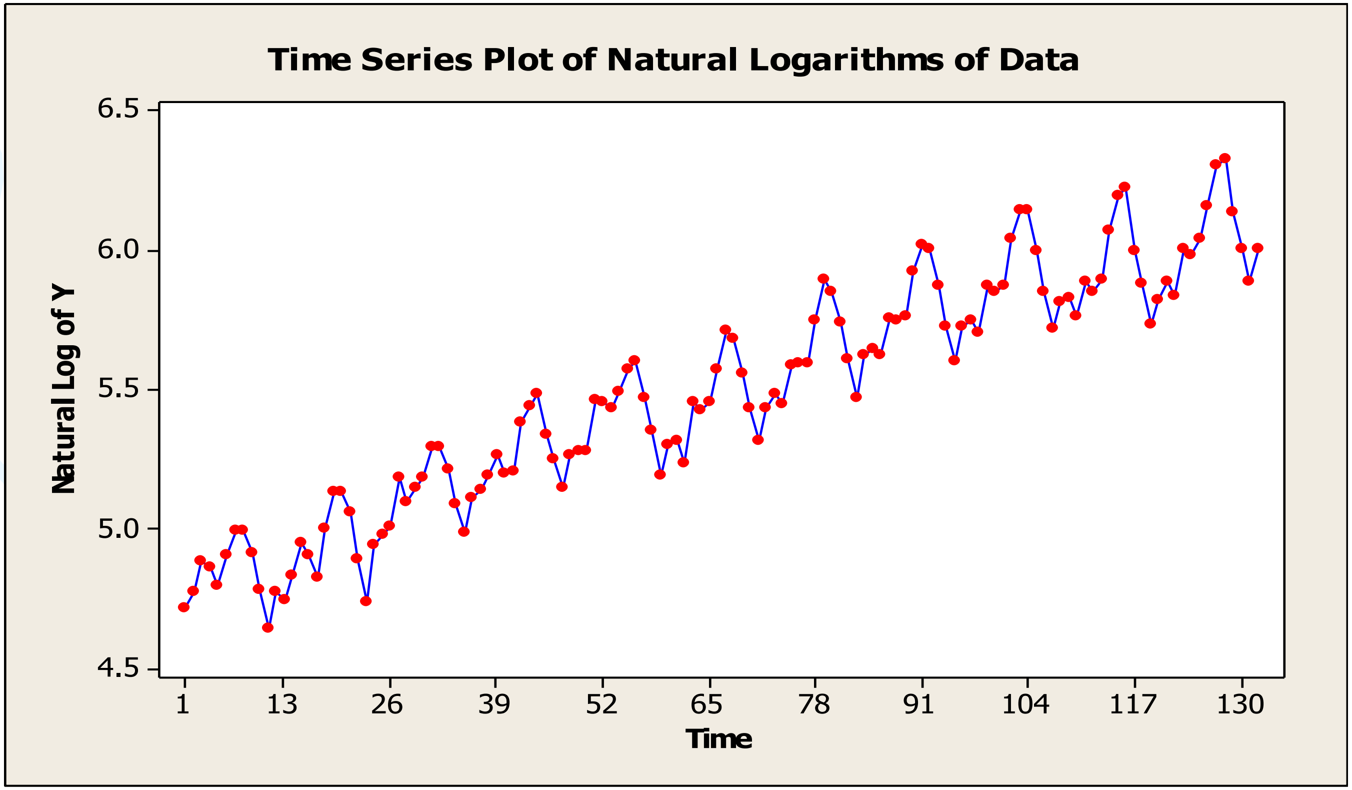
(continued)

- What pattern can you see from this plot?
- How should we deal with the data?

Example 2: Airline

(continued)

- The main approach for achieving stationarity in variance is through a logarithmic transformation of the data.



Example 2: Airline

(continued)

- The magnitude of the fluctuations in the logarithmic transformed data does not vary with time.
- Thus the logarithmic transformation has stabilized the seasonal variation of the time series.

Example 2: Airline

(continued)

```
                                Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      0.171929        1.00000    |                    |********************|             0
  1      0.163046        0.94833    |                 .  |******************* |      0.087039
  2      0.152719        0.88827    |              .     |******************  |      0.145609
  3      0.143920        0.83709    |             .      |*****************   |      0.182090
  4      0.136109        0.79165    |            .       |****************    |      0.209222
  5      0.130713        0.76027    |           .        |***************     |      0.230802
  6      0.126663        0.73671    |          .         |***************     |      0.249053
  7      0.123189        0.71651    |         .          |**************      |      0.265049
  8      0.121189        0.70487    |         .          |**************      |      0.279337
  9      0.122678        0.71354    |        .           |**************      |      0.292502
 10      0.124408        0.72360    |        .           |**************      |      0.305404
 11      0.127268        0.74023    |       .            |***************     |      0.318127
 12      0.128345        0.74650    |       .            |***************     |      0.330918
 13      0.120137        0.69876    |      .             |**************      |      0.343439
 14      0.110500        0.64271    |      .             |*************.      |      0.354046
 15      0.102461        0.59595    |     .              |************  .     |      0.362777
 16      0.094839        0.55162    |     .              |***********   .     |      0.370119
 17      0.089001        0.51766    |     .              |**********    .     |      0.376296
 18      0.084712        0.49271    |     .              |**********    .     |      0.381652
 19      0.081181        0.47218    |     .              |*********     .     |      0.386441
 20      0.079455        0.46214    |    .               |*********      .    |      0.390788
 21      0.080889        0.47048    |    .               |*********      .    |      0.394906
 22      0.082262        0.47847    |    .               |**********     .    |      0.399130
 23      0.084101        0.48916    |    .               |**********     .    |      0.403452
 24      0.084720        0.49276    |    .               |**********     .    |      0.407920

                           "." marks two standard errors
```

Example 2: Airline

(continued)

```
                         Name of Variable = lr

        Period(s) of Differencing                          1
        Mean of Working Series                      0.009812
        Standard Deviation                          0.106108
        Number of Observations                           131
        Observation(s) eliminated by differencing          1


                                Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      0.011259        1.00000    |                    |********************|             0
  1     0.0021204        0.18833    |                 .  |****                |      0.087370
  2    -0.0014316        -.12715    |                .***|   .                |      0.090416
  3    -0.0017371        -.15428    |                .***|   .                |      0.091771
  4    -0.0036710        -.32605    |             *******|   .                |      0.093730
  5    -0.0007441        -.06609    |                .  *|   .                |      0.102022
  6    0.00045622        0.04052    |                .   |*  .                |      0.102348
  7    -0.0011062        -.09826    |                . **|   .                |      0.102470
  8    -0.0038662        -.34339    |             *******|   .                |      0.103187
  9    -0.0012221        -.10855    |                . **|   .                |      0.111570
 10    -0.0013476        -.11969    |                . **|   .                |      0.112373
 11     0.0022423        0.19916    |               .    |****.               |      0.113342
 12     0.0093768        0.83283    |               .    |****************    |      0.115983
 13     0.0022286        0.19794    |              .     |**** .              |      0.155053
 14    -0.0016113        -.14311    |              .  ***|     .              |      0.156970
 15    -0.0012339        -.10959    |              .   **|     .              |      0.157963
 16    -0.0032466        -.28836    |              ******|     .              |      0.158542
 17    -0.0005219        -.04636    |              .    *|     .              |      0.162497
 18    0.00040322        0.03581    |             .      |*     .             |      0.162598
 19    -0.0011741        -.10428    |             .    **|      .             |      0.162658
 20    -0.0035209        -.31272    |             .******|      .             |      0.163167
 21    -0.0011947        -.10612    |             .    **|      .             |      0.167680
 22    -0.0009539        -.08473    |             .    **|      .             |      0.168192
 23     0.0020862        0.18530    |             .      |****  .             |      0.168518
 24     0.0080362        0.71376    |             .      |**************      |      0.170066

                        "." marks two standard errors
```

Example 2: Airline

(continued)

```
                          Name of Variable = lr

            Period(s) of Differencing                        1,12
            Mean of Working Series                       0.001322
            Standard Deviation                           0.044849
            Number of Observations                            119
            Observation(s) eliminated by differencing          13

                             Autocorrelations

Lag    Covariance    Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1    Std Error

  0     0.0020114       1.00000     |                    |********************|          0
  1    -0.0006378       -.31710     |              ******|    .               |   0.091670
  2    0.00021914       0.10895     |                 .  |**  .               |   0.100466
  3    -0.0004343       -.21592     |                ****|    .               |   0.101454
  4    0.00008998       0.04474     |                 .  |*   .               |   0.105244
  5    0.00006247       0.03106     |                 .  |*   .               |   0.105404
  6    0.00008597       0.04274     |                 .  |*   .               |   0.105481
  7    -0.0001303       -.06478     |                 . *|    .               |   0.105626
  8    0.00002510       0.01248     |                 .  |    .               |   0.105960
  9    0.00032664       0.16239     |                 .  |***.                |   0.105972
 10    -0.0001042       -.05179     |                 . *|    .               |   0.108043
 11    0.00014860       0.07388     |                 .  |*   .               |   0.108251
 12    -0.0008254       -.41035     |            ********|    .               |   0.108674
 13    0.00032666       0.16240     |                .   |***  .              |   0.120996
 14    -0.0001072       -.05329     |                .  *|     .              |   0.122814
 15    0.00029374       0.14604     |                .   |***  .              |   0.123008
 16    -0.0002927       -.14552     |                . ***|     .              |   0.124457
 17    0.00018639       0.09267     |                .   |**   .              |   0.125879
 18    -0.0000602       -.02992     |                .  *|     .              |   0.126451
 19    0.00010879       0.05409     |                .   |*    .              |   0.126510
 20    -0.0002837       -.14102     |                . ***|     .              |   0.126704
 21    0.00007227       0.03593     |                .   |*    .              |   0.128016
 22    -0.0001621       -.08061     |                .  **|     .              |   0.128101
 23    0.00043582       0.21667     |                .   |****.               |   0.128527
 24    -0.0000607       -.03017     |                .  *|     .              |   0.131560

                     "." marks two standard errors
```

Example 2: Airline

(continued)

```
                      Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1       -0.31710    |              ******|   .                |
  2        0.00934    |                .   |   .                |
  3       -0.19873    |                ****|   .                |
  4       -0.09252    |                . **|   .                |
  5        0.03361    |                .   |*  .                |
  6        0.03035    |                .   |*  .                |
  7       -0.06064    |                .  *|   .                |
  8       -0.00849    |                .   |   .                |
  9        0.21218    |                .   |****                |
 10        0.04408    |                .   |*  .                |
 11        0.07212    |                .   |*  .                |
 12       -0.35474    |             *******|   .                |
 13       -0.08283    |                . **|   .                |
 14       -0.02168    |                .   |   .                |
 15       -0.01661    |                .   |   .                |
 16       -0.13096    |                .***|   .                |
 17        0.05560    |                .   |*  .                |
 18        0.07705    |                .   |** .                |
 19       -0.00278    |                .   |   .                |
 20       -0.14277    |                .***|   .                |
 21        0.11704    |                .   |** .                |
 22       -0.04138    |                .  *|   .                |
 23        0.18134    |                .   |****                |
 24       -0.10472    |                . **|   .                |
```

Example 2: Airline

(continued)

- At the nonseasonal level ACF has a significant spike at lag 1 and cuts off after lag 1 (with the exception of a spike at lag 3), and the PACF dies down ⇒ nonseasonal MA(1) component
- At the seasonal level both ACF and PACF appear to cut off after lag 12. However, autocorrelation at lag 24 in ACF is smaller than the partial autocorrelation at lag 24 in PACF. ⇒ ACF cuts off after lag 12 and PACF dies down. ⇒ seasonal MA(1) component
- For the logged data, a tentative model would be:

  $\text{ARIMA}(0,1,1)(0,1,1)_{12}$

or $(1 - B)(1 - B^{12})(lny_t) = \delta + (1 - \theta_1 B)(1 - \Theta_1 B^{12})\varepsilon_t$

Recapitulation

1. Make the series stationary.
   - Logarithmic will often take care of nonstationary variance (pre-differencing transformation).
   - Differencing (nonseasonal and/or seasonal) will usually take care of nonstationarity in the mean.
2. Consider nonseasonal aspects.
   - An examination of the ACF and PACF of the stationary series obtained in Step 1 can reveal whether a MA or AR model is feasible.
3. Consider seasonal aspects.
   - An examination of the ACF and PACF at the seasonal lags can help identify MA and AR models for the seasonal aspects of the data, but the indications are not as easy to find as in the case of the nonseasonal aspects.

Step Three: Parameter Estimation

Parameter Estimation

- The method of least squares can be used. However, for models involving an MA component, there is no simple formula that can be applied to obtain the estimates.
- Another method which is frequently used is maximum likelihood.

Parameter Estimation

(continued)

- Given n observations $y_1, y_2, \ldots, y_n$, the likelihood function L is defined to be the probability of obtaining the data actually observed.
- The maximum likelihood estimators (m.l.e.) are those value of the parameters for which the data actually observed are most likely, that is, the values that maximize the likelihood function L.

Parameter Estimation

- Estimate parameters once a tentative model has been selected.
  - Determine whether the parameters are significantly different from zero.
    - t – values

      $$t = \frac{\hat{\theta}}{s_{\hat{\theta}}}$$

      where $\hat{\theta}$ is the estimated coefficient and $s_{\hat{\theta}}$ is the standard error of the estimated coefficient. Reject H_0: $\theta = 0$ if $|t| > t_{\alpha/2,n-r}$
    - p-value: Reject H_0: $\theta = 0$ if p-value < α.
  - If these parameters are not significant, we may have been able to improve the model by dropping the corresponding terms from the model.

Parameter Estimation

- In most Box-Jenkins models, the sample size after differencing is at least 30. Therefore, as a practical rule, it is reasonable to compare the absolute value of a parameter estimate with 2 for the testing.
- To test whether the constant term should be included in the model, i.e., H_0: $\delta = 0$ vs H_a: $\delta \neq 0$
  - If $\left|\dfrac{\bar{z}}{s_z / \sqrt{n'}}\right| > 2$, reject H_0, where $\bar{z}$ is the mean of the working series, s_z is the standard deviation, and n' is the sample size of the working series.

Example 1: Industry Sales

Period(s) of Differencing	1,12
Mean of Working Series	0.389673
Standard Deviation	68.54769
Number of Observations	107
Observation(s) eliminated by differencing	13

Should the constant be included in the model?

Example 1: Industry Sales *(continued)*

- An ARIMA$(0,1,1)(0,1,1)_{12}$ model was identified

The ARIMA Procedure

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.84028	0.05527	15.20	<.0001	1
MA2,1	0.63569	0.10202	6.23	<.0001	12

- The estimated model is

$$(1 - B)(1 - B^{12})y_t = (1 - 0.840B)(1 - 0.636B^{12})\varepsilon_t$$

i.e.,

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \varepsilon_t - 0.840\varepsilon_{t-1} - 0.636\varepsilon_{t-12} + 0.534\varepsilon_{t-13}$$

Exercise

- Example 2: Airline
  - Dataset: airline.dat
  - SAS program: ARIMA_airline.sas

- Write the estimated model. Are all the parameters significant?

Step Four: Diagnostic Checking

Diagnostic Checking

- Residual analysis
  - The residuals left over after fitting the model should be white noise.
- Model selection criteria.
  - AIC, SBC

White Noise Series

- White noise series $\{\varepsilon_t\}$
  - ε's are iid (independent and identically distributed) random variables with the following properties.

$$(1)\ E(\varepsilon_t) = 0 \quad \textit{for all } t.$$

$$(2)\ Var(\varepsilon_t) = \sigma^2 \quad \textit{for all } t.$$

$$(3)\ Cov(\varepsilon_t, \varepsilon_{t+s}) = 0 \quad \textit{for all } t,\ \ s \neq t.$$

Portmanteau Test

- Rather than study the autocorrelation values one at a time, an alternative approach is to consider a whole set of autocorrelation values all at one time, and test to see whether the set is significantly different from a zero set.
- Ljung-Box test

$$Q = n(n+2)\sum_{k=1}^{m}\frac{r_k^2(e)}{n-k} \sim \chi^2_{m-r}$$

where $r_k(e)$ = the residual autocorrelation at lag k

n = number of residuals

k = time lag

m = number of time lags to be tested

r = number of parameters estimated in the model

Example: Random Numbers (WNOISE.DAT)

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	2.44	6	0.8754	0.103	0.099	-0.043	-0.031	-0.183	0.025

$Q(6) = 2.44 < \chi^2_{0.05,6} = 12.59$ $\Rightarrow$ White noise series

Exercise: Pigs slaughtered

- The dataset pigs.dat lists the monthly total number of pigs slaughtered in the state of Victoria, Australia, from January 1990 through August 1995. ACF and PACF plots are given. Is the series a white noise series?
  - SAS program – ARIMA_pigs.sas
  - Dataset – pigs.dat

Exercise: Pigs slaughtered

(continued)

```
                                   Autocorrelations

Lag   Covariance   Correlation   -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0    104001682       1.00000   |                    |********************|              0
  1     37233269       0.35801   |               .    |*******             |       0.121268
  2     29504608       0.28369   |               .    |******              |       0.135925
  3     36875165       0.35456   |              .     |*******             |       0.144370
  4      6933036       0.06666   |              .     |*    .              |       0.156653
  5     19538193       0.18786   |              .     |**** .              |       0.157070
  6     19229857       0.18490   |              .     |**** .              |       0.160340
  7      9514415       0.09148   |             .      |**    .             |       0.163445
  8      2352843       0.02262   |             .      |      .             |       0.164197
  9      3966057       0.03813   |             .      |*     .             |       0.164243
 10     -8420991       -.08097   |             .    **|      .             |       0.164373
 11      -754726       -.00726   |             .      |      .             |       0.164958
 12     18206590       0.17506   |             .      |****  .             |       0.164963
 13     -2300270       -.02212   |             .      |      .             |       0.167673
 14     -4904801       -.04716   |             .     *|      .             |       0.167716
 15     -4710970       -.04530   |             .     *|      .             |       0.167911
 16    -17556435       -.16881   |             .   ***|      .             |       0.168090
 17      3678019       0.03536   |             .      |*     .             |       0.170565

                          "." marks two standard errors
```

Exercise: Pigs slaughtered

(continued)

```
                      Partial Autocorrelations

 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1        0.35801    |               .    |*******              |
   2        0.17839    |               .    |****.                |
   3        0.24553    |               .    |*****                |
   4       -0.17665    |               .****|    .                |
   5        0.13668    |               .    |*** .                |
   6        0.04692    |               .    |*   .                |
   7        0.03399    |               .    |*   .                |
   8       -0.16238    |               . ***|    .                |
   9        0.03352    |               .    |*   .                |
  10       -0.13352    |               . ***|    .                |
  11        0.10077    |               .    |**  .                |
  12        0.18842    |               .    |****.                |
  13       -0.08236    |               .  **|    .                |
  14       -0.15065    |               . ***|    .                |
  15       -0.07418    |               .   *|    .                |
  16       -0.03954    |               .   *|    .                |
  17        0.18738    |               .    |****.                |
```

Autocorrelation Check for White Noise

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	29.74	6	<.0001	0.358	0.284	0.355	0.067	0.188	0.185
12	33.70	12	0.0008	0.091	0.023	0.038	-0.081	-0.007	0.175

Back to diagnostic checking

Residual Analysis

- If the portmanteau test is insignificant, the model is adequate.
- If the portmanteau test is significant, the model is inadequate. Then we need to go back and consider other ARIMA models.
- The pattern of significant spikes in the ACF and PACF of the residuals may suggest how the model can be improved.
  – Significant spikes at the seasonal lags suggest adding a seasonal component to the chosen model.
  – Significant spikes at low lags suggest the nonseasonal AR or MA components of the model.

Example 1: Industry Sales

- Fitted model: ARIMA$(0,1,1)(0,1,1)_{12}$

```
                             Autocorrelation Plot of Residuals

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0      1843.781        1.00000    |                    |********************|             0
  1      -143.159        -.07764    |                . **|   .                |      0.096674
  2     86.261335        0.04679    |                .   |*  .                |      0.097255
  3       179.851        0.09754    |                .   |** .                |      0.097465
  4    -92.322566        -.05007    |                .  *|   .                |      0.098373
  5    -56.282458        -.03053    |                .  *|   .                |      0.098611
  6     18.823102        0.01021    |                .   |   .                |      0.098699
  7     68.411283        0.03710    |                .   |*  .                |      0.098709
  8    -63.962248        -.03469    |                .  *|   .                |      0.098839
  9       190.553        0.10335    |                .   |** .                |      0.098953
 10     49.229254        0.02670    |                .   |*  .                |      0.099957
 11     23.350217        0.01266    |                .   |   .                |      0.100023
 12    -79.112865        -.04291    |                .  *|   .                |      0.100038
 13    -14.682606        -.00796    |                .   |   .                |      0.100210
 14    -93.010226        -.05045    |                .  *|   .                |      0.100216
 15      -179.905        -.09757    |                . **|   .                |      0.100453
 16      -276.551        -.14999    |                .***|   .                |      0.101335
 17     80.953099        0.04391    |                .   |*  .                |      0.103389
 18      -162.092        -.08791    |                . **|   .                |      0.103563
 19      -179.815        -.09753    |                . **|   .                |      0.104258
 20      -129.861        -.07043    |                .  *|   .                |      0.105107
 21      -109.191        -.05922    |                .  *|   .                |      0.105548
 22    -50.957328        -.02764    |                .  *|   .                |      0.105858
 23    -25.346260        -.01375    |                .   |   .                |      0.105925
 24    -94.296602        -.05114    |                .  *|   .                |      0.105942

                              "." marks two standard errors
```

Example 1: Industry Sales *(continued)*

```
                    Partial Autocorrelations

 Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1       -0.07764     |               . **|   .               |
   2        0.04100     |               .   |*  .               |
   3        0.10502     |               .   |** .               |
   4       -0.03721     |               .  *|   .               |
   5       -0.04784     |               .  *|   .               |
   6       -0.00138     |               .   |   .               |
   7        0.05209     |               .   |*  .               |
   8       -0.02319     |               .   |   .               |
   9        0.09142     |               .   |** .               |
  10        0.03570     |               .   |*  .               |
  11        0.01743     |               .   |   .               |
  12       -0.06544     |               .  *|   .               |
  13       -0.01789     |               .   |   .               |
  14       -0.04222     |               .  *|   .               |
  15       -0.09073     |               . **|   .               |
  16       -0.17890     |               ****|   .               |
  17        0.03571     |               .   |*  .               |
  18       -0.06427     |               .  *|   .               |
  19       -0.10484     |               . **|   .               |
  20       -0.12942     |               .***|   .               |
  21       -0.05143     |               .  *|   .               |
  22       -0.00946     |               .   |   .               |
  23        0.00341     |               .   |   .               |
  24       -0.05927     |               .  *|   .               |
```

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	--------------------Autocorrelations--------------------					
6	2.38	4	0.6670	-0.078	0.047	0.098	-0.050	-0.031	0.010
12	4.28	10	0.9338	0.037	-0.035	0.103	0.027	0.013	-0.043
18	9.96	16	0.8687	-0.008	-0.050	-0.098	-0.150	0.044	-0.088
24	12.86	22	0.9370	-0.098	-0.070	-0.059	-0.028	-0.014	-0.051

Example 1: Industry Sales

(continued)

- Is the residual series white noise? Why?
- Is the fitted model ARIMA$(0,1,1)(0,1,1)_{12}$ adequate? Why?

Model Selection Criteria

- Akaike Information Criterion (AIC)

  $$AIC = -2 \ln(L) + 2k$$

- Schwartz Bayesian Criterion (SBC)

  $$SBC = -2 \ln(L) + k \ln(n)$$

  where L = likelihood function

  k = number of parameters to be estimated,

  n = number of observations.

- Ideally, the AIC and SBC will be as small as possible.
- Usually the model with the smallest AIC and SBC will have residuals which resemble white noise.

Example 1: Industry Sales

- AIC and SBC values

The ARIMA Procedure

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.84028	0.05527	15.20	<.0001	1
MA2,1	0.63569	0.10202	6.23	<.0001	12

Variance Estimate	1843.781
Std Error Estimate	42.93927
AIC	1117.827
SBC	1123.173
Number of Residuals	107

Identification Revisited

- Having estimated an ARIMA model, it is necessary to revisit the question of identification to see if the selected model can be improved.
  - Some of the estimated parameters may be insignificant. If so, a revised model with the insignificant terms omitted may be considered
  - The residual analysis can provide some guidance on whether the model is adequate and how it may be improved.
  - If there are more than one adequate model identified, use model selection criteria (AIC and SBC) to determine which of them is preferred.

Step Five: Forecasting

Forecasting

- After an adequate model has been found, forecasts for one period or several periods into the future can be made.

$$\hat{y}_{t+l}(t) = l\text{-step-ahead forecast of } y_{t+l} \text{ made at time } t$$

$$e_{t+l}(t) = l\text{-step-ahead forecast error}$$

$$= \; y_{t+l} - \hat{y}_{t+l}(t)$$

- Prediction intervals based on the forecasts can also be constructed. In general, for a given confidence level, the longer the forecast lead time, the larger the prediction interval.

Example 1: Industry Sales

- Fitted adequate model: ARIMA$(0,1,1)(0,1,1)_{12}$

$$(1 - B)(1 - B^{12})y_t = (1 - 0.840B)(1 - 0.636B^{12})\varepsilon_t$$

or

$$y_t = y_{t\text{-}1} + y_{t\text{-}12} - y_{t\text{-}13} + \varepsilon_t - 0.840\varepsilon_{t\text{-}1} - 0.636\varepsilon_{t\text{-}12} + 0.534\varepsilon_{t\text{-}13}$$

- Therefore, the predicted value is expressed as

$$\hat{y}_t = y_{t\text{-}1} + y_{t\text{-}12} - y_{t\text{-}13} + e_t - 0.840e_{t\text{-}1} - 0.636e_{t\text{-}12} + 0.534e_{t\text{-}13}$$

where e_t = the residual at time t.

  - As we forecast further ahead, there will be no empirical values for the e terms after a while, then their expected values will be zero.
  - When there is no known past values for y, the forecasted values will be used instead.

Example 1: Industry Sales *(continued)*

Obs	date	r	FORECAST	RESIDUAL	U95	L95
101	101	835.09	826.79	8.295	910.97	742.613
102	102	934.60	872.73	61.860	956.91	788.556
103	103	832.50	719.64	112.861	803.82	635.460
104	104	300.00	388.63	-88.632	472.81	304.453
105	105	791.44	797.36	-5.914	881.54	713.179
106	106	900.00	885.21	14.792	969.39	801.030
107	107	781.73	822.38	-40.652	906.56	738.203
108	108	880.00	867.20	12.797	951.38	783.026
109	109	875.02	887.88	-12.856	972.05	803.709
110	110	992.97	894.94	98.033	979.11	810.765
111	111	976.80	975.80	1.000	1059.97	891.635
112	112	968.70	901.49	67.210	985.66	817.319
113	113	871.68	889.01	-17.337	973.18	804.844
114	114	1006.85	950.87	55.981	1035.04	866.704
115	115	832.04	819.03	13.005	903.20	734.864
116	116	345.59	405.19	-59.598	489.35	321.018
117	117	849.53	843.54	5.983	927.71	759.378
118	118	913.87	940.50	-26.629	1024.67	856.333
119	119	868.75	851.70	17.041	935.87	767.538
120	120	993.73	922.86	70.872	1007.03	838.695

Example 1: Industry Sales

(continued)

- Forecast for period 121

$$\hat{y}_{121} = y_{120} + y_{109} - y_{108} + \hat{e}_{121} - 0.840e_{120} - 0.636e_{109} + 0.534e_{108}$$
$$= 993.73 + 875.02 - 880.00 + 0 - 0.840(70.87) - 0.636(-12.86) + 0.534(12.80)$$
$$= 944.22$$

- Forecast for period 122

$$\hat{y}_{122} = \hat{y}_{121} + y_{110} - y_{109} + \hat{e}_{122} - 0.840\hat{e}_{121} - 0.636e_{110} + 0.534e_{109}$$
$$= 944.22 + 992.97 - 875.02 + 0 - 0.840(0) - 0.636(98.03) + 0.534(-12.86)$$
$$= 992.99$$

Example 1: Industry Sales

(continued)

Obs	date	r	FORECAST	RESIDUAL	U95	L95
101	101	835.09	826.79	8.295	910.97	742.613
102	102	934.60	872.73	61.860	956.91	788.556
103	103	832.50	719.64	112.861	803.82	635.460
104	104	300.00	388.63	-88.632	472.81	304.453
105	105	791.44	797.36	-5.914	881.54	713.179
106	106	900.00	885.21	14.792	969.39	801.030
107	107	781.73	822.38	-40.652	906.56	738.203
108	108	880.00	867.20	12.797	951.38	783.026
109	109	875.02	887.88	-12.856	972.05	803.709
110	110	992.97	894.94	98.033	979.11	810.765
111	111	976.80	975.80	1.000	1059.97	891.635
112	112	968.70	901.49	67.210	985.66	817.319
113	113	871.68	889.01	-17.337	973.18	804.844
114	114	1006.85	950.87	55.981	1035.04	866.704
115	115	832.04	819.03	13.005	903.20	734.864
116	116	345.59	405.19	-59.598	489.35	321.018
117	117	849.53	843.54	5.983	927.71	759.378
118	118	913.87	940.50	-26.629	1024.67	856.333
119	119	868.75	851.70	17.041	935.87	767.538
120	120	993.73	922.86	70.872	1007.03	838.695
121	121	.	944.22	.	1028.38	860.056
122	122	.	992.99	.	1078.21	907.761
123	123	.	1028.54	.	1114.82	942.259
124	124	.	978.25	.	1065.57	890.929
125	125	.	928.14	.	1016.49	839.790
126	126	.	1018.48	.	1107.84	929.111
127	127	.	865.29	.	955.66	774.922
128	128	.	423.67	.	515.03	332.303
129	129	.	891.98	.	984.33	799.629
130	130	.	976.44	.	1069.76	883.119

Forecasting

- If the pattern of the series appears to be changing over time, the new data may be used to re-estimate the model parameters or, if necessary, to develop an entirely new model.
  - If the magnitudes of the most recent errors tend to be consistently larger than previous errors
  - If the recent forecast errors tend to be consistently positive or negative

Example: Electricity

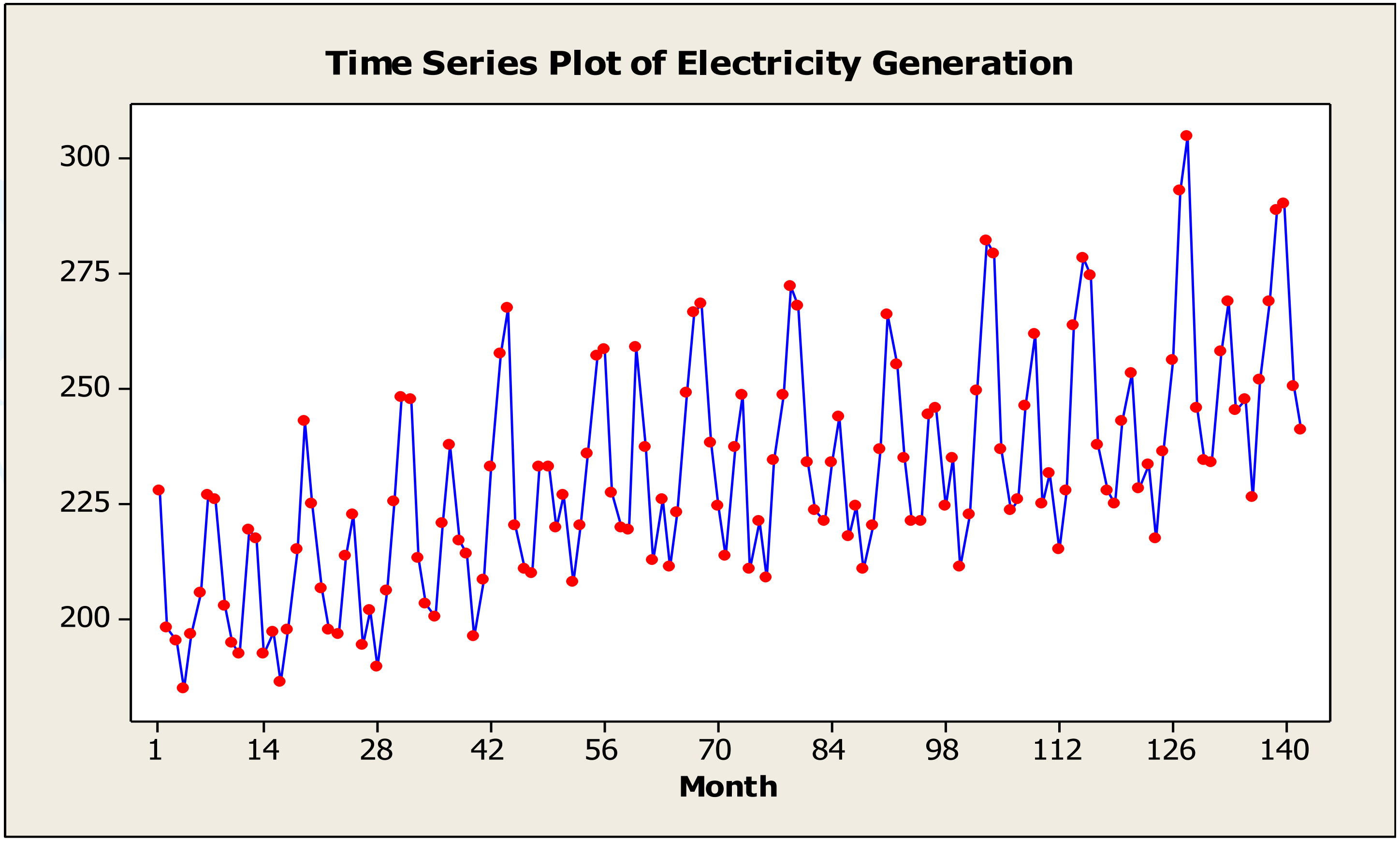
This example studies the total net generation of electricity by the US electric industry (monthly for the period 1985-1996). In general there are two peaks per year: in mid-summer and mid-winter.

- Dataset: elecnew.dat
- SAS program: ARIMA_elec.sas

Example: Electricity

(continued)

- Steps 1 and 2: Stationarity and Identification



Questions

- Is the series stationary in the variance?
- Do the data need transformation?

Example: Electricity

(continued)

- Steps 1 and 2: Stationarity and Identification

```
                                  The ARIMA Procedure

                                Name of Variable = r

                          Mean of Working Series     231.0894
                          Standard Deviation         24.28651
                          Number of Observations          142


                                    Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0       589.835       1.00000     |                    |********************|              0
   1       397.900       0.67460     |                 .  |*************       |       0.083918
   2       162.695       0.27583     |               .    |******              |       0.115982
   3     13.410138       0.02274     |               .    |    .               |       0.120513
   4     60.179109       0.10203     |               .    |**  .               |       0.120543
   5       226.535       0.38407     |               .    |********            |       0.121150
   6       274.250       0.46496     |               .    |*********           |       0.129441
   7       231.566       0.39260     |              .     |********            |       0.140712
   8     58.025227       0.09838     |              .     |**   .              |       0.148225
   9     -4.049811       -.00687     |              .     |     .              |       0.148684
  10       116.024       0.19671     |              .     |**** .              |       0.148686
  11       314.046       0.53243     |              .     |***********         |       0.150508
  12       469.851       0.79658     |             .      |****************    |       0.163234
  13       310.204       0.52592     |            .       |***********         |       0.188633
  14     92.586524       0.15697     |            .       |***    .            |       0.198691
  15    -50.421283       -.08548     |            .     **|       .            |       0.199562
  16    -11.409408       -.01934     |            .       |       .            |       0.199820
  17       134.261       0.22762     |            .       |*****  .            |       0.199833
  18       185.093       0.31380     |            .       |****** .            |       0.201651
  19       156.763       0.26577     |            .       |*****  .            |       0.205061
  20      4.533345       0.00769     |            .       |       .            |       0.207473
  21    -56.377598       -.09558     |            .     **|       .            |       0.207475
  22     47.541000       0.08060     |            .       |**     .            |       0.207785
  23       227.166       0.38514     |            .       |********            |       0.208005
  24       364.853       0.61857     |           .        |************        |       0.212967

                          "." marks two standard errors
```

Example: Electricity

(continued)

- Steps 1 and 2: Stationarity and Identification

```
                         Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1        0.67460    |                 .  |*************       |
  2       -0.32894    |             *******|  .                 |
  3       -0.00544    |                 .  |  .                 |
  4        0.35960    |                 .  |*******             |
  5        0.34262    |                 .  |*******             |
  6       -0.16634    |                 ***|  .                 |
  7        0.14221    |                 .  |***                 |
  8       -0.23973    |               *****|  .                 |
  9        0.22089    |                 .  |****                |
 10        0.36505    |                 .  |*******             |
 11        0.32137    |                 .  |******              |
 12        0.35715    |                 .  |*******             |
 13       -0.49097    |          **********|  .                 |
 14       -0.05402    |                 . *|  .                 |
 15       -0.02264    |                 .  |  .                 |
 16       -0.12131    |                 .**|  .                 |
 17       -0.20504    |                ****|  .                 |
 18       -0.00929    |                 .  |  .                 |
 19        0.03688    |                 .  |* .                 |
 20        0.02961    |                 .  |* .                 |
 21       -0.00615    |                 .  |  .                 |
 22        0.06615    |                 .  |* .                 |
```

Questions

- What pattern can you see from the plots?
- Are the data stationary?
- If not, find an appropriate differencing which yields stationary data.

Example: Electricity

(continued)

- Steps 1 and 2: Stationarity and Identification

```
                                 Name of Variable = r

                 Period(s) of Differencing                           12
                 Mean of Working Series                        4.624538
                 Standard Deviation                            8.460615
                 Number of Observations                             130
                 Observation(s) eliminated by differencing          12


                                   Autocorrelations

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0     71.582005        1.00000     |                    |********************|             0
   1     26.304756        0.36748     |                .   |*******             |      0.087706
   2      7.265402        0.10150     |                .   |** .                |      0.098842
   3      7.004010        0.09785     |                .   |** .                |      0.099641
   4      6.273813        0.08765     |                .   |** .                |      0.100377
   5      8.126128        0.11352     |                .   |** .                |      0.100964
   6      9.153861        0.12788     |                .   |***.                |      0.101941
   7      6.696492        0.09355     |                .   |** .                |      0.103168
   8      4.512418        0.06304     |                .   |*  .                |      0.103819
   9      8.892881        0.12423     |                .   |** .                |      0.104113
  10      9.374951        0.13097     |                .   |***.                |      0.105247
  11     -8.022806        -.11208     |                . **|   .                |      0.106493
  12    -21.151168        -.29548     |              ******|   .                |      0.107397
  13     -1.095840        -.01531     |               .    |    .               |      0.113478
  14     -0.648280        -.00906     |               .    |    .               |      0.113494
  15     -2.745953        -.03836     |               .   *|    .               |      0.113499
  16      2.438672        0.03407     |               .    |*   .               |      0.113599
  17     -8.293145        -.11586     |               .  **|    .               |      0.113678
  18     -3.462361        -.04837     |               .   *|    .               |      0.114582
  19      1.602588        0.02239     |               .    |    .               |      0.114739
  20      2.210203        0.03088     |               .    |*   .               |      0.114773
  21      1.457980        0.02037     |               .    |    .               |      0.114837
  22     -2.006987        -.02804     |               .   *|    .               |      0.114864
  23     -0.707524        -.00988     |               .    |    .               |      0.114917
  24     -7.886328        -.11017     |               .  **|    .               |      0.114924

                              "." marks two standard errors
```

Example: Electricity

(continued)

- Steps 1 and 2: Stationarity and Identification

```
                              Partial Autocorrelations

  Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

    1        0.36748    |                .   |*******             |
    2       -0.03878    |                .  *|   .                |
    3        0.08493    |                .   |** .                |
    4        0.03031    |                .   |*  .                |
    5        0.07959    |                .   |** .                |
    6        0.06386    |                .   |*  .                |
    7        0.02135    |                .   |   .                |
    8        0.01192    |                .   |   .                |
    9        0.09636    |                .   |** .                |
   10        0.04482    |                .   |*  .                |
   11       -0.22836    |               *****|   .                |
   12       -0.25702    |               *****|   .                |
   13        0.18900    |                .   |****                |
   14       -0.07104    |                .  *|   .                |
   15       -0.02374    |                .   |   .                |
   16        0.08149    |                .   |** .                |
   17       -0.13454    |                .***|   .                |
   18        0.10480    |                .   |** .                |
   19        0.02690    |                .   |*  .                |
   20        0.04732    |                .   |*  .                |
   21        0.11518    |                .   |** .                |
   22       -0.05185    |                .  *|   .                |
   23       -0.11444    |                . **|   .                |
   24       -0.19591    |                ****|   .                |
```

Example: Electricity

(continued)

- Steps 1 and 2: Stationarity and Identification

```
Period(s) of Differencing                               1,12
Mean of Working Series                              0.130078
Standard Deviation                                  9.458743
Number of Observations                                   129
Observation(s) eliminated by differencing                 13

                                  Autocorrelations

Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

  0     89.467819       1.00000     |                    |********************|             0
  1    -25.912552       -.28963     |              ******|   .                |      0.088045
  2    -17.734845       -.19823     |                ****|   .                |      0.095145
  3      0.253296       0.00283     |                .   |   .                |      0.098294
  4     -2.920514       -.03264     |                .  *|   .                |      0.098295
  5      1.853153       0.02071     |                .   |   .                |      0.098379
  6      4.338437       0.04849     |                .   |*  .                |      0.098412
  7     -2.317332       -.02590     |                .  *|   .                |      0.098597
  8     -6.085861       -.06802     |                .  *|   .                |      0.098650
  9      3.793443       0.04240     |                .   |*  .                |      0.099013
 10     18.131192       0.20266     |                .   |****                |      0.099154
 11     -5.367522       -.05999     |                .  *|   .                |      0.102314
 12    -32.125589       -.35907     |             *******|   .                |      0.102587
 13     19.313824       0.21587     |                .   |****                |      0.111906
 14      2.622177       0.02931     |               .    |*   .               |      0.115089
 15     -7.326501       -.08189     |               .  **|    .               |      0.115147
 16     16.982155       0.18981     |               .    |****.               |      0.115597
 17    -15.698164       -.17546     |               .****|    .               |      0.117989
 18     -0.772249       -.00863     |               .    |    .               |      0.119994
 19      6.521073       0.07289     |               .    |*   .               |      0.119999
 20     -0.522156       -.00584     |               .    |    .               |      0.120342
 21      2.782310       0.03110     |               .    |*   .               |      0.120344
 22     -5.033370       -.05626     |               .   *|    .               |      0.120406
 23      8.879535       0.09925     |               .    |**  .               |      0.120610
 24    -15.200440       -.16990     |               . ***|    .               |      0.121242

                          "." marks two standard errors
```

Example: Electricity

(continued)

- Steps 1 and 2: Stationarity and Identification

```
                              Partial Autocorrelations

 Lag     Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

   1       -0.28963     |              ******|   .                |
   2       -0.30794     |              ******|   .                |
   3       -0.19470     |                ****|   .                |
   4       -0.20488     |                ****|   .                |
   5       -0.14309     |                .***|   .                |
   6       -0.06941     |                .  *|   .                |
   7       -0.07629     |                . **|   .                |
   8       -0.13707     |                .***|   .                |
   9       -0.07819     |                . **|   .                |
  10        0.19033     |                .   |****                |
  11        0.14969     |                .   |***.                |
  12       -0.28764     |              ******|   .                |
  13       -0.00810     |                .   |   .                |
  14       -0.03952     |                .  *|   .                |
  15       -0.14995     |                .***|   .                |
  16        0.08908     |                .   |** .                |
  17       -0.12660     |                .***|   .                |
  18       -0.03932     |                .  *|   .                |
  19       -0.02959     |                .  *|   .                |
  20       -0.10390     |                . **|   .                |
  21        0.07577     |                .   |** .                |
  22        0.10523     |                .   |** .                |
  23        0.15359     |                .   |***.                |
  24       -0.25199     |               *****|   .                |
```

Question

- What model(s) do you think is(are) appropriate to fit the data?

Example: Electricity

(continued)

- Steps 1 and 2: Stationarity and Identification
  - At the seasonal level, ACF cuts off after lag 12 and PACF dies down $\Rightarrow$ seasonal MA(1) component
  - At the nonseasonal level, ACF cuts off after second lag and PACF dies down $\Rightarrow$ nonseasonal MA(2) component
  - Plausible model: ARIMA$(0,1,2)(0,1,1)_{12}$

Should the constant term be included?

Example: Electricity

(continued)

- Step 3: Estimation
  - ARIMA$(0,1,2)(0,1,1)_{12}$

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.57573	0.08309	6.93	<.0001	1
MA1,2	0.22962	0.08368	2.74	0.0061	2
MA2,1	0.80524	0.09450	8.52	<.0001	12

Example: Electricity

(continued)

- Step 4: Diagnostic Checking
  - Residual analysis for ARIMA$(0,1,2)(0,1,1)_{12}$

```
                            Autocorrelation Plot of Residuals

 Lag    Covariance    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1      Std Error

   0     41.755689        1.00000    |                    |********************|             0
   1      0.563880        0.01350    |                .   |   .                |      0.088045
   2      0.844696        0.02023    |                .   |   .                |      0.088061
   3     -1.380084        -.03305    |                .  *|   .                |      0.088097
   4     -0.509446        -.01220    |                .   |   .                |      0.088193
   5     -1.073637        -.02571    |                .  *|   .                |      0.088206
   6      3.428447        0.08211    |                .   |** .                |      0.088264
   7     -0.026933        -.00065    |                .   |   .                |      0.088855
   8     -3.012537        -.07215    |                .  *|   .                |      0.088855
   9      2.614205        0.06261    |                .   |*  .                |      0.089308
  10      5.139207        0.12308    |                .   |** .                |      0.089647
  11     -3.807877        -.09119    |                . **|   .                |      0.090948
  12      3.123751        0.07481    |                .   |*  .                |      0.091654
  13      4.498787        0.10774    |                .   |** .                |      0.092126
  14      1.030376        0.02468    |                .   |   .                |      0.093097
  15     -3.953567        -.09468    |                . **|   .                |      0.093148
  16      5.239907        0.12549    |                .   |***.                |      0.093891
  17     -4.653677        -.11145    |                . **|   .                |      0.095183
  18     -0.689724        -.01652    |                .   |   .                |      0.096189
  19      2.483783        0.05948    |                .   |*  .                |      0.096211
  20     -2.139162        -.05123    |                .  *|   .                |      0.096495
  21      2.515779        0.06025    |                .   |*  .                |      0.096706
  22     -0.345529        -.00828    |                .   |   .                |      0.096997
  23      3.292574        0.07885    |                .   |** .                |      0.097002
  24     -3.576530        -.08565    |                . **|   .                |      0.097498

                                 "." marks two standard errors
```

Example: Electricity

(continued)

- Step 4: Diagnostic Checking
  - Residual analysis for ARIMA$(0,1,2)(0,1,1)_{12}$

```
                       Partial Autocorrelations

Lag    Correlation    -1 9 8 7 6 5 4 3 2 1 0 1 2 3 4 5 6 7 8 9 1

  1       0.01350     |               .   |   .               |
  2       0.02005     |               .   |   .               |
  3      -0.03361     |               .  *|   .               |
  4      -0.01172     |               .   |   .               |
  5      -0.02407     |               .   |   .               |
  6       0.08236     |               .   |** .               |
  7      -0.00279     |               .   |   .               |
  8      -0.07796     |               . **|   .               |
  9       0.07061     |               .   |*  .               |
 10       0.12821     |               .   |***.               |
 11      -0.10311     |               . **|   .               |
 12       0.06908     |               .   |*  .               |
 13       0.12551     |               .   |***.               |
 14       0.02721     |               .   |*  .               |
 15      -0.11616     |               . **|   .               |
 16       0.12214     |               .   |** .               |
 17      -0.07669     |               . **|   .               |
 18      -0.03117     |               .  *|   .               |
 19       0.03104     |               .   |*  .               |
 20      -0.05387     |               .  *|   .               |
 21       0.10981     |               .   |** .               |
 22      -0.07512     |               . **|   .               |
 23       0.06219     |               .   |*  .               |
 24      -0.04060     |               .  *|   .               |
```

Example: Electricity

(continued)

- Step 4: Diagnostic Checking
  - Residual analysis for ARIMA(0,1,2)(0,1,1)$_{12}$

Autocorrelation Check of Residuals

To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.26	3	0.7383	0.014	0.020	-0.033	-0.012	-0.026	0.082
12	6.69	9	0.6693	-0.001	-0.072	0.063	0.123	-0.091	0.075
18	14.07	15	0.5201	0.108	0.025	-0.095	0.125	-0.111	-0.017
24	17.77	21	0.6633	0.059	-0.051	0.060	-0.008	0.079	-0.086

Questions:

Is this model adequate? Why?

Example: Electricity

(continued)

- Step 4: Diagnostic Checking

Fitted adequate model: ARIMA$(0,1,2)(0,1,1)_{12}$

$$(1 - B)(1 - B^{12})y_t = (1 - 0.576B - 0.230B^2)(1 - 0.805B^{12})\,\varepsilon_t$$

or

$$y_t = y_{t-1} + y_{t-12} - y_{t-13} + \varepsilon_t - 0.576\varepsilon_{t-1} - 0.230\varepsilon_{t-2} - 0.805\varepsilon_{t-12} + 0.464\varepsilon_{t-13} + 0.185\varepsilon_{t-14}$$

Maximum Likelihood Estimation

Parameter	Estimate	Standard Error	t Value	Approx Pr > |t|	Lag
MA1,1	0.57573	0.08309	6.93	<.0001	1
MA1,2	0.22962	0.08368	2.74	0.0061	2
MA2,1	0.80524	0.09450	8.52	<.0001	12

Variance Estimate	41.75569
Std Error Estimate	6.461864
AIC	864.0822
SBC	872.6616
Number of Residuals	129

Example: Electricity

(continued)

- Step 5: Forecasting
  - Forecast the next 24 months of generation of electricity by the US electric industry.

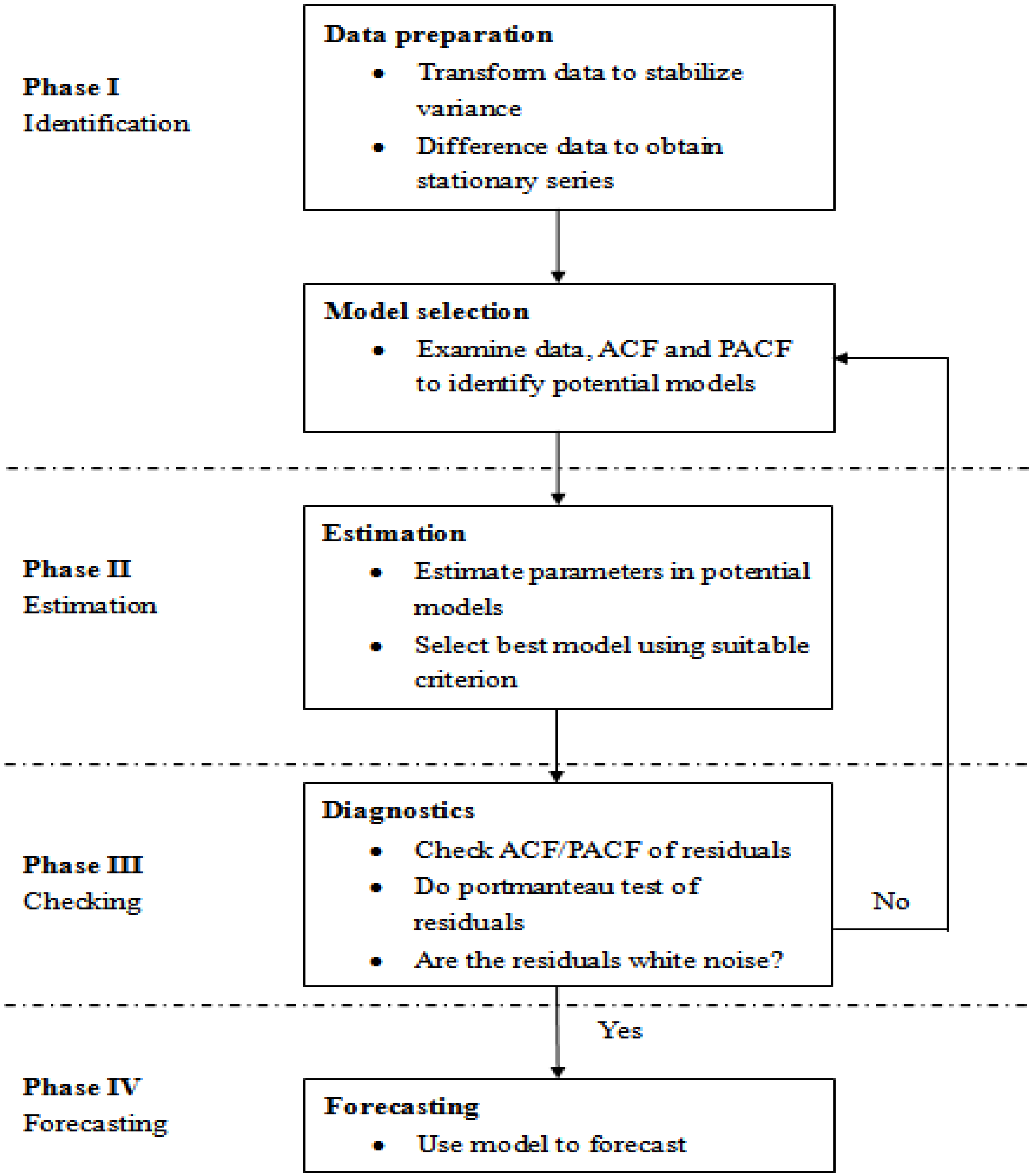
Obs	Forecast	Std Error	95% Confidence Limits	
143	240.3627	6.4619	227.6977	253.0277
144	262.6872	7.0194	248.9294	276.4449
145	270.3573	7.1312	256.3804	284.3342
146	244.0965	7.2413	229.9039	258.2892
147	249.9720	7.3497	235.5668	264.3771
148	232.8138	7.4566	218.1992	247.4284
149	249.4648	7.5619	234.6437	264.2859
150	270.8399	7.6658	255.8152	285.8646
151	295.6695	7.7683	280.4439	310.8951
152	295.8185	7.8695	280.3947	311.2424
153	257.2000	7.9694	241.5804	272.8197
154	246.5065	8.0680	230.6935	262.3195
155	245.1110	8.4513	228.5467	261.6752
156	267.2616	8.6392	250.3291	284.1940
157	274.9317	8.7689	257.7450	292.1184
158	248.6709	8.8967	231.2337	266.1082
159	254.5464	9.0228	236.8621	272.2307
160	237.3882	9.1470	219.4603	255.3161
161	254.0392	9.2697	235.8710	272.2074
162	275.4143	9.3907	257.0089	293.8197
163	300.2439	9.5102	281.6043	318.8836
164	300.3930	9.6282	281.5221	319.2639
165	261.7745	9.7448	242.6751	280.8738
166	251.0809	9.8600	231.7558	270.4061

Example: Electricity

(continued)

- Based on the following information, give forecast for period 143

date	r	FORECAST	RESIDUAL
120	242.91	249.752	-6.8422
121	253.08	253.866	-0.7856
122	228.13	227.621	0.5088
123	233.68	235.112	-1.4322
124	217.38	218.261	-0.8812
125	236.38	231.598	4.7823
126	256.08	256.924	-0.8439
127	292.83	278.779	14.0508
128	304.71	281.851	22.8589
129	245.57	255.050	-9.4805
130	234.41	234.955	-0.5450
131	234.12	234.937	-0.8175
132	258.17	256.343	1.8275
133	268.66	264.279	4.3815
134	245.31	238.745	6.5649
135	247.47	247.316	0.1544
136	226.25	229.714	-3.4637
137	251.67	242.705	8.9647
138	268.79	269.725	-0.9353
139	288.94	293.249	-4.3093
140	290.16	291.521	-1.3614
141	250.69	253.524	-2.8338
142	240.80	241.743	-0.9431



Schematic representation of the Box-Jenkins methodology for time series modeling