Forecasting

in Economics, Business, Finance and Beyond

CH 7B: Non-Stationary Models

Review: Classical Decomposition Time-Series Components

Time Series

Trend Component	Seasonal Component	Cyclical Component	Irregular /Random Component
Overall, persistent, long-term movement	Regular periodic fluctuations, usually within a 12-month period	Repeating swings or movements over more than one year; ARMA model	Erratic or residual fluctuations

2

Review: Classical Decomposition Time-Series Components

We have seen that the Classical Decomposition approach can be very useful.

It is also easy to interpret.

We have restricted ourselves to causal, invertible, stationary models for the cyclical component.

This method doesn't always work well, as we saw with the Dow Jones data.

There are other approaches. The Box-Jenkins approach uses a different method to model trend and seasonaity.

3

The Box-Jenkins Approach

The Box-Jenkins approach is one of the most widely used methodologies for the analysis of time-series data

It is popular because of its generality; it can handle any series, stationary or not, with or without seasonal elements, and it has well-documented computer programs

Box and Jenkins were neither the originators nor the most important contributors in the field of ARMA models, but they have popularized these models and made them readily accessible to everyone, so much that ARMA models are sometimes referred to as Box-Jenkins models.

4

The Box-Jenkins Approach

- The basic steps in the Box-Jenkins methodology are
  - (1) differencing the series so as to achieve stationarity,
  - (2) identification of a tentative model,
  - (3) estimation of the model,
  - (4) diagnostic checking (if the model is found inadequate, we go back to step 2), and
  - (5) using the model for forecasting and control.

5

The Box-Jenkins Approach

1. Differencing to achieve stationarity: How do we conclude whether a time series is stationary or not?
   - We can do this by studying the graphs of the ACF/PACF of the series.
   - The ACF/PACF of a stationary series drops off as k, the number of lags, becomes large, but this is not usually the case for a nonstationary series.
   - Thus the common procedure is to plot the ACF/PACF of the given series y_t and successive differences Δy, Δy, and so on, and look at the ACF/PACF at each stage.
   - We keep differencing until the ACF/PACF dampens

6

The Box-Jenkins Approach

2. Once we have used the differencing procedure to get a stationary time series, we examine the ACF/PACF to decide on the appropriate orders of the AR and MA components.
   - The ACF/PACF of a MA process is zero after a point.
   - That of an ARprocess declines geometrically. The ACF/PACF of ARMA processes show different patterns (but all dampen after a while).
   - Based on these, one arrives at a tentative ARMA model.
   - This step involves more of a judgmental procedure than the use of any clear-cut rules.

7

The Box-Jenkins Approach

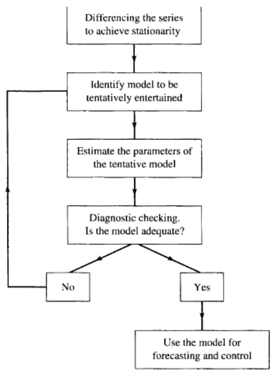
3. The next step is the estimation of the tentative ARMA model identified in step 2. We have discussed in the preceding section the estimation of ARMA models.

4. The next step is diagnostic checking to check the adequacy of the tentative model.

5. The final step is forecasting.

8

The Box-Jenkins Approach



9

Deterministic and stochastic trends

A trend is a long-term movement or tendency in the data.

A **deterministic trend** is a nonrandom function of time

(e.g. $y_t = t$, or $y_t = t^2$).

A **stochastic trend** is random and varies over time

An important example of a stochastic trend is a **random walk**:

$Y_t = Y_{t-1} + u_t$, where u_t is serially uncorrelated

If Y_t follows a random walk, then the value of Y tomorrow is the value of Y today, plus an unpredictable disturbance.

Two key features of a random walk:

$Y_{T+h|T} = Y_T$

- Your best prediction of the value of Y in the future is the value of Y today
- To a first approximation, log stock prices follow a random walk (more precisely, stock returns are unpredictable

$\text{var}(Y_{T+h|T} - Y_T) = h\,\sigma_u^2$

- The variance of your forecast error increases linearly in the horizon. The more distant your forecast, the greater the forecast uncertainty. (Technically this is the sense in which the series is "nonstationary")

11

A random walk with a drift

$Y_t = \beta_0 + Y_{t-1} + u_t$, where u_t is serially uncorrelated

The "drift" is β_0: If $\beta_0 \neq 0$, then Y_t follows a random walk around a linear trend. You can see this by considering the h- step ahead forecast: $Y_{T+h|T} = \beta_0 h + Y_T$

The random walk model (with or without drift) is a good description of stochastic trends in many economic time series.

12

Practical Advice

If Y_t has a random walk trend, then ΔY_t is stationary and analysis should be undertaken using **ΔY_t instead of Y_t.**

Upcoming specifics that lead to this advice: the relationship between the random walk model and AR(1), AR(2), AR(*p*) ("unit autoregressive root")

13

Unit root in an AR(1)

Random walk (with drift): $Y_t = \beta_0 + Y_{t-1} + u_t$

AR(1): $Y_t = \beta_0 + \beta_1 Y_{t-1} + u_t$

- The random walk is an AR(1) with $\beta_1 = 1$.
- The special case of $\beta_1 = 1$ is called a unit root.
- When $\beta_1 = 1$, the AR(1) model becomes $\Delta Y_t = \beta_0 + u_t$

14

Unit roots in an AR(2)

AR(2): $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t$

Rearrange:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + u_t \\ &= \beta_0 + (\beta_1 + \beta_2) Y_{t-1} - \beta_2 Y_{t-1} + \beta_2 Y_{t-2} + u_t \\ &= \beta_0 + (\beta_1 + \beta_2) Y_{t-1} - \beta_2 (Y_{t-1} - Y_{t-2}) + u_t \end{aligned}$$

Subtract Y_{t-1} from both sides:

$$Y_t - Y_{t-1} = \beta_0 + (\beta_1 + \beta_2 - 1) Y_{t-1} - \beta_2 (Y_{t-1} - Y_{t-2}) + u_t$$

or

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + u_t,$$

where $\delta = \beta_1 + \beta_2 - 1$ and $\gamma_1 = -\beta_2$.

15

Unit roots in an AR(2), ctd.

Thus the AR(2) model can be rearranged as

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + u_t$$

where $\delta = \beta_1 + \beta_2 - 1$ and $\gamma_1 = -\beta_2$.

<u>Claim</u>: if $1 - \beta_1 z - \beta_2 z^2 = 0$ has a unit root, then $\beta_1 + \beta_2 = 1$

Thus, if there is a unit root, then $\delta = 0$ and the AR(2) model becomes $\Delta Y_t = \beta_0 + \gamma_1 \Delta Y_{t-1} + u_t$

If an AR(2) model has a unit root, then it can be written as an AR(1) in first differences.

16

Unit roots in the AR(p) model

AR(p): $Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \dots + \beta_p Y_{t-p} + u_t$

This can be rearranged as

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + u_t$$

where

$$\delta = \beta_1 + \beta_2 + \dots + \beta_p - 1$$
$$\gamma_1 = -(\beta_2 + \dots + \beta_p)$$
$$\gamma_2 = -(\beta_3 + \dots + \beta_p)$$
$$\dots$$
$$\gamma_{p-1} = -\beta_p$$

17

Unit roots in the AR(p) model, ctd.

The AR(p) model can be written as

$$\Delta Y_t = \beta_0 + \delta Y_{t-1} + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + u_t$$

where $\delta = \beta_1 + \beta_2 + \dots + \beta_p - 1$.

If there is a unit root in the AR(p) model, then $\delta = 0$ and the AR(p) model becomes an AR(p–1) model in first differences:

$$\Delta Y_t = \beta_0 + \gamma_1 \Delta Y_{t-1} + \gamma_2 \Delta Y_{t-2} + \dots + \gamma_{p-1} \Delta Y_{t-p+1} + u_t$$

18

ARIMA Models

We have already seen that appropriate differencing can remove trend & seasonality (Recall the DEATHS data).

The **AutoRegressive Integrated Moving Average (ARIMA)** model, is a broadening of the class of ARMA models to include differencing. A process $\{X_t\}$ is said to be an **ARIMA(p,d,q)** if $\{(1\text{-}B)^d X_t\}$ is a causal ARMA(p,q).

We write the model as:

$$f(B)(1\text{-}B)^d X_t = q(B) Z_t, \quad \{Z_t\} \sim WN(0,\sigma^2)$$

The process is stationary if and only if d=0. Differencing X_t d times at lag 1 results in an ARMA(p,q) with f(B) and q(B) as AR & MA polynomials.

19

The Dow Jones Data, reprised

Recall the problems we had trying to use the classical decomposition methods on the Dow Jones data set.

Let's try an ARIMA model.

Open the file DOWJ.tsm in ITSM.

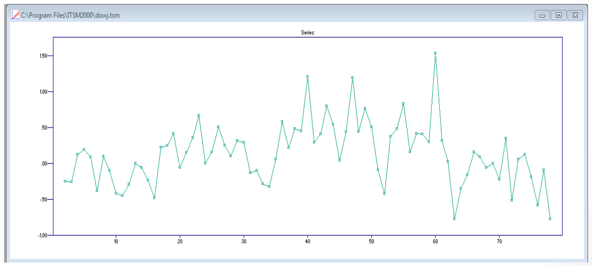
Instead of fitting a linear trend, lets "model" the trend by using the Box-Jenkins approach.

Click Transform -> Difference and leave the lag at the default lag of 1. Click OK.

20

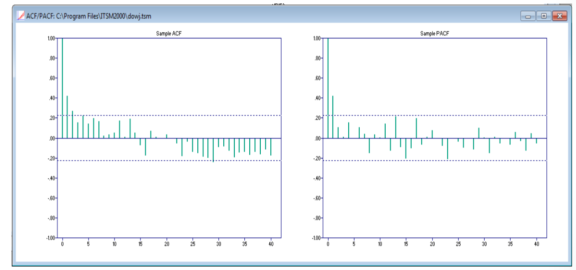
The Dow Jones Data, reprised

You should see this:



The Dow Jones Data, reprised

Check the ACF/PACF (Statistics -> ACF/PACF -> Sample):



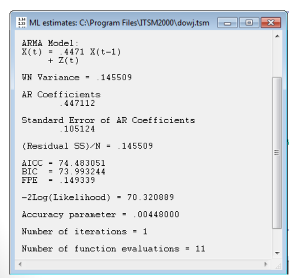
WOW! MUCH better! Looks like an AR(1) should work.

The Dow Jones Data, reprised

Click Model -> Estimation -> Preliminary, Click Yes to subtract the mean, enter 1 as the AR order, click Burg as the preliminary method, and click OK. Click Model -> Estimation -> Max likelihood, and click OK.

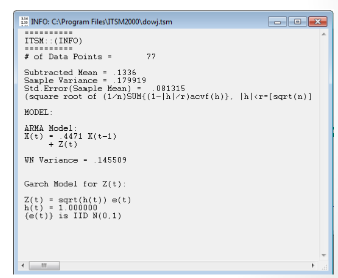
The Dow Jones Data, reprised

We get the report on the left below. Use Window (if necessary to return to the graph of the differenced data. Click the icon and select Info to get the report on the right below.



ML estimates: C:\Program Files\ITSM2000\dowj.tsm

```
ARMA Model:
X(t) = .4471 X(t-1)
     + Z(t)

WN Variance = .145509

AR Coefficients
       .447112

Standard Error of AR Coefficients
       .105124

(Residual SS)/N = .145509

AICC = 74.483051
BIC  = 73.993244
FPE  = .149339

-2Log(Likelihood) = 70.320889

Accuracy parameter = .00448000

Number of iterations = 1

Number of function evaluations = 11
```



INFO: C:\Program Files\ITSM2000\dowj.tsm

```
==========
ITSM::(INFO)
==========
# of Data Points =       77

Subtracted Mean = .1336
Sample Variance = .179919
Std.Error(Sample Mean) =  .081315
(square root of (1/n)SUM{(1-|h|/r)acvf(h)}, |h|<r=[sqrt(n)]

MODEL:

ARMA Model:
X(t) = .4471 X(t-1)
     + Z(t)

WN Variance = .145509


Garch Model for Z(t):

Z(t) = sqrt(h(t)) e(t)
h(t) = 1.000000
{e(t)} is IID N(0,1)
```

The Dow Jones Data, reprised

We have fit an ARIMA(1,1,0) model. Our ARMA model is:

$$Y_t = 0.1336 + 0.447112 * Y_{t-1} + Z_t$$

But $Y_t = X_t - X_{t-1}$, and so $Y_{t-1} = X_{t-1} - X_{t-2}$

Substituting for the Y's yields

$$X_t - X_{t-1} = 0.1336 + 0.447112 * (X_{t-1} - X_{t-2}) + Z_t$$

$$\Rightarrow \quad X_t = 0.1336 + X_{t-1} + 0.447112 * X_{t-1} - 0.447112 * X_{t-2} + Z_t$$

$$X_t = 0.1336 + 1.447112 * X_{t-1} - 0.447112 * X_{t-2} + Z_t$$

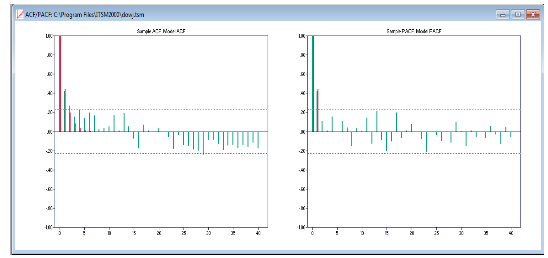
Note that this is a non-stationary AR(2) in X_t, but a stationary AR(1) in Y_t.

25

The Dow Jones Data, reprised

Our model is: $X_t = 0.1336 + 1.447112 * X_{t-1} - 0.447112 * X_{t-2} + Z_t$

Is it any good? Check the Sample/Model ACF/PACF:



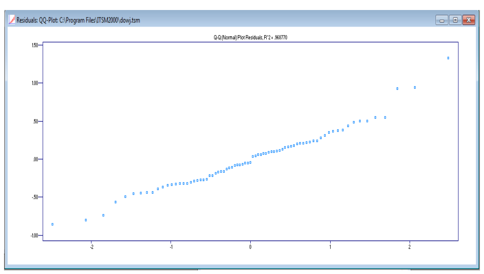
Looks good so far! Let's check the residuals.

26

The Dow Jones Data, reprised

Our model is: $X_t = 0.1336 + 1.447112 * X_{t-1} - 0.447112 * X_{t-2} + Z_t$

The Q-Q plot looks OK!



27

The Dow Jones Data, reprised

Our model is: $X_t = 0.1336 + 1.447112 * X_{t-1} - 0.447112 * X_{t-2} + Z_t$

The tests of randomness are OK. The Ljung-Box says the errors are not iid, but the other 4 tests say they are iid. The Jarque-Bera test rejects on normality, even though the normal Q-Q plot looked OK.

Tests of randomness: C:\Program Files\ITSM2000\dowj.tsm

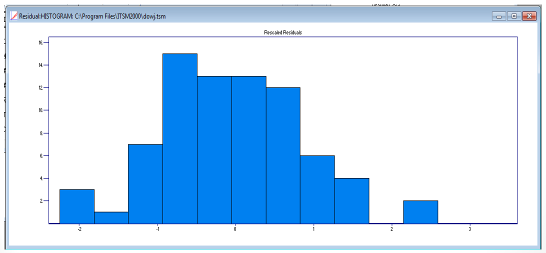
```
============================================
ITSM::(Tests of randomness on residuals)
============================================

Ljung - Box statistic = 32.507 Chi-Square ( 20 ), p-value = .03818

McLeod - Li statistic = 16.648 Chi-Square ( 21 ), p-value = .73223

# Turning points = 49.000~AN(50.000,sd = 3.6560), p-value = .78445

# Diff sign points = 37.000~AN(38.000,sd = 2.5495), p-value = .69489

Rank test statistic = .14950E+04~AN(.14630E+04,sd = .11368E+03), p-value = .77833

Jarque-Bera test statistic (for normality) = 8.5864 Chi-Square (2), p-value = .01366

Order of Min AICC YW Model for Residuals = 0
```

28

The Dow Jones Data, reprised

Our model is: $X_t = 0.1336 + 1.447112*X_{t-1} - 0.447112*X_{t-2} + Z_t$

The Jarque-Bera test rejects on normality, even though the normal Q-Q plot looked OK. The histogram below shows the problem; some skewness and outliers.

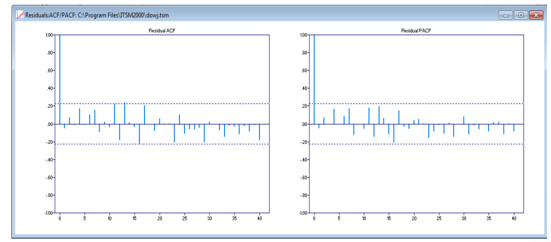


29

The Dow Jones Data, reprised

Our model is: $X_t = 0.1336 + 1.447112*X_{t-1} - 0.447112*X_{t-2} + Z_t$

The ACF/PACF of the residuals shows no significant values, although some are close.



30

The Dow Jones Data, reprised

Our model is: $X_t = 0.1336 + 1.447112*X_{t-1} - 0.447112*X_{t-2} + Z_t$

In general, this should be a useful model. YAY!

31

For next time…

Play with the Ind_Prod_Index.tsm data set. Try to find a model you like. You might try both the Classical Decomposition and Box-Jenkins approaches. Keep track of each step you do so that we can re-create your work (if we so desire!) in class.

You will need the PRO version of ITSM2000 to open the full data set; the student version will limit you to 250 observations (there are over 350 in the full data set).

Happy Modelling!

32

Summary of ARMA/ARIMA modeling procedures

1. Perform **preliminary transformations**
2. **Detrend** and **deseasonalize**
3. **Difference successively** (at lag 1)
4. **Examine sample ACF & PACF**
5. Obtain **preliminary estimates**
6. Obtain **maximum likelihood estimates**
7. **Choose the ML model with smallest AICC**
8. Use **Autofit**
9. **Fit subset models**
10. Check candidate models for **goodness-of-fit**

33

Summary of ARMA/ARIMA modeling procedures

1. Perform **preliminary transformations** (if necessary) to stabilize variance over time.

   This can often be achieved by the *Box-Cox transformation*:

   $f_\lambda(X_t) = (X_t^\lambda - 1)/\lambda$, if $X_t \geq 0$, and $\lambda > 0$,

   $f_\lambda(X_t) = \log(X_t)$, if $X_t > 0$, and $\lambda = 0$.

   In practice, $\lambda = 0$ or $\lambda = 0.5$ are often adequate.

34

Summary of ARMA/ARIMA modeling procedures

2. **Detrend** and **deseasonalize** the data (if necessary) to make the stationarity assumption look reasonable. (Trend and seasonality are also characterized by ACF's that are slowly decaying and nearly periodic, respectively).

   The primary methods for achieving this are ***classical decomposition***, and ***differencing***.

35

Summary of ARMA/ARIMA modeling procedures

3. If the data looks nonstationary without a well-defined trend or seasonality, an alternative to the above option is to **difference successively** (at lag 1).

   This may also need to be done after the above step anyway.

36

Summary of ARMA/ARIMA modeling procedures

4. **Examine sample ACF & PACF** to get an idea of potential p & q values.

   For an AR(p)/MA(q), the sample PACF/ACF cuts off after lag p/q.

Summary of ARMA/ARIMA modeling procedures

5. Obtain **preliminary estimates** of the coefficients for select values of p & q.

   For q=0 (pure AR), use *Burg*

   For p=0 (pure MA) use *Innovations*

   For $p\neq0$ & $q\neq0$ (ARMA) use *Hannan-Rissanen*.

Summary of ARMA/ARIMA modeling procedures

6. Starting from the preliminary estimates, obtain **maximum likelihood estimates** of the coefficients for the promising models found in step 5.

Summary of ARMA/ARIMA modeling procedures

7. From the fitted ML models above, **choose the one with smallest AICC**, taking into consideration also other candidate models whose AICC is *close* to the minimum (within about 2 units).

   The minimization of the AICC must be done one model at a time, but this search can be carried out systematically by examining all the pairs (p,q) such that p+q=1, 2, … , in turn.

   A quicker but rougher method: run through ARMA(p,p)'s, as p=1,2,…, in turn.

Summary of ARMA/ARIMA modeling procedures

8. Can bypass steps 4-7 by using the option **Autofit**. This automatically searches for the minimum AICC ARMA(p,q) model (based on ML estimates), for all values of p and q in the user-specified range.

   Drawbacks:
   a) can take a long time
   b) initial estimates for all parameters set at 0.001

   The resulting model should be checked via prelim. est. followed by ML est. to guard against the possibility of being trapped in a local maximum of the likelihood surface.

41

Summary of ARMA/ARIMA modeling procedures

9. Inspection of the standard errors of the coefficients at the ML estimation stage, may reveal that some of them are not significant.

   If so, **subset models** can be fitted by *constraining* these to be zero at a second iteration of ML estimation. Use a cutoff of between 1 (more conservative, use when few parameters in model) and 2 (less conservative) standard errors when assessing significance.

42

Summary of ARMA/ARIMA modeling procedures

10. Check the candidate models for **goodness-of-fit** by examining their *residuals*.

    This involves inspecting their ACF/PACF for departures from WN, carrying out the formal WN hypothesis tests, etc.

43

Summary of ARMA/ARIMA modeling procedures

1. Perform **preliminary transformations**
2. **Detrend** and **deseasonalize**
3. **Difference successively** (at lag 1)
4. **Examine sample ACF & PACF**
5. Obtain **preliminary estimates**
6. Obtain **maximum likelihood estimates**
7. **Choose the ML model with smallest AICC**
8. Use **Autofit**
9. **Fit subset models**
10. Check candidate models for **goodness-of-fit**

44

Playtime

We are halfway through the semester!

You need some practice before we move on.

Next time I will show you how to fit subset models, which just means we can set insignificant terms in high order ARMA(p,q) models to zero.

I also need to show you seasonal ARIMA, or SARIMA, models.

45

Playtime

In the meantime, here are two more practice data sets.

The Lake Huron Data - LAKE.tsm

The level in feet of Lake Huron (reduced by 570) in the years 1875–1972

The Wine Data – WINE.tsm

Monthly sales (in kiloliters) of red wine by Australian winemakers from January 1980 through October 1991

Happy Modeling!

46

Review: Backshift notation

A very useful notational device is the backward shift operator, B, which is used as follows:

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

In other words, 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**:

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

For monthly data, if we wish to shift attention to "the same month last year," then B^{12} is used, and the notation is $B^{12} y_t = y_{t-12}$.

47

Forecasting using R Backshift notation 3

Backshift notation for ARMA

ARMA model:

$$\begin{aligned} y_t &= c + \varphi_1 y_{t-1} + \cdots + \varphi_p y_{t-p} + e_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q} \\ &= c + \varphi_1 B y_t + \cdots + \varphi_p B^p y_t + e_t + \theta_1 B e_t + \cdots + \theta_q B^q e_t \end{aligned}$$

$$\varphi(B) y_t = c + \theta(B) e_t$$

where $\varphi(B) = 1 - \varphi_1 B - \cdots - \varphi_p B^p$

and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$.

48

Forecasting using R Backshift notation 5

Backshift notation

First difference: $1 - B$

Double difference: $(1 - B)^2$

dth-order difference: $(1 - B)^d$

Seasonal difference: $1 - B^s$

49

Backshift notation

Seasonal difference followed by a first difference: $(1 - B)(1 - B^s)$

Multiply terms together together to see the combined effect:

$$(1 - B)(1 - B^s)X_t = (1 - B - B^m + B^{s+1})X_t$$
$$= X_t - X_{t-1} - X_{t-s} + X_{t-s-1}$$

50

Backshift notation for ARIMA

ARIMA(1,**1**,1) model:

$$\underset{\text{AR(1)}}{\underbrace{(1 - \phi_1 B)}}\ \underset{\text{lag 1 difference}}{\underbrace{\mathbf{(1 - B)}}}y_t = c + \underset{\text{MA(1)}}{\underbrace{(1 + \theta_1 B)}}e_t$$

51

Seasonal ARIMA (SARIMA) Models

Last time we looked at ARIMA models for data with seasonality.

If one seasonal cycle lasts for s measurements, then if we difference at lag s we might* remove the seasonality.

$$Y_t = \nabla_s X_t = X_t - X_{t-s} = \left(1 - B^s\right) X_t$$

*actually, only the average seasonality is removed; seasonal variability across each set of lag s observations might remain

Differencing seasonally D times is denoted

$$Y_t = \nabla_s^D X_t = \left(1 - B^s\right)^D X_t$$

52

Seasonal ARIMA (SARIMA) Models

A **multiplicative** seasonal autoregressive integrated moving average model is denoted SARIMA (p,d,q) x (P,D,Q)$_s$

$$\phi(B)\Phi(B^s)\nabla^d\nabla_s^D X_t = \theta(B)\Theta(B^s)Z_t$$

p P d D q Q

p: non-seasonal AR(p)

q: non-seasonal MA(q)

P: seasonal AR(P)

Q: seasonal MA(Q)

d: lag 1 difference d times

D: lag s difference D times

53

SARIMA Models

e.g. SARIMA (1,1,1) x (1,1,1)$_{12}$ (without constant)

$$(1-\phi_1 B)(1-\Phi_1 B^{12})(1-B)(1-B^{12})X_t = (1-\theta_1 B)(1-\Theta_1 B^{12})Z_t$$

Note: $(1-\phi_1 B)(1-\Phi_1 B^{12})X_t = (1-\phi_1 B - \Phi_1 B^{12} + \phi_1\Phi_1 B^{13})$

This is NOT the same as fitting an AR(13) and setting coefficients for terms 2 – 11 to 0! Here, we are constraining the model so that the coefficient on the lag 13 term is the product of the coefficients of the lag 1 and lag 12 terms.

The same is true for the MA polynomial.

54

But…….why?

Earlier we looked at the DEATHS.tsm data. We looked at a classical decomposition model of the form

$$X_t = m_t + s_t + Y_t$$

This model assumes that the seasonal component s_t repeats itself precisely the same way cycle after cycle.

This might not be true.

SARIMA models allow for randomness in the seasonal pattern from one cycle to the next. We can decide whether or not to incorporate the multiplicative constraints.

55

The DEATHS.tsm Data

The deaths.tsm data set gives the monthly accidental deaths in the US for 1973 – 1978.

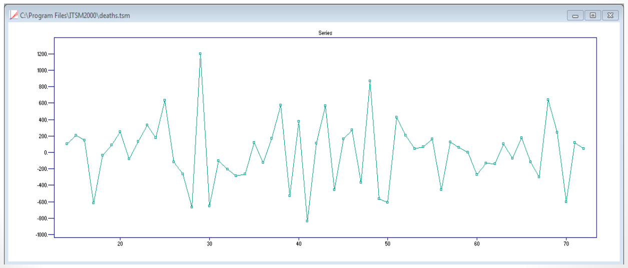
A classical decomposition with seasonality at lag 12 (monthly data) and quadratic trend leads to (after subtracting the mean of 0.6821×10^{-12}) and ARMA(2,2) model with AICC = 983.730. There are problems with the sample/model ACF/PACF, but the residuals look OK.

Let's try SARIMA models.

56

The DEATHS.tsm Data

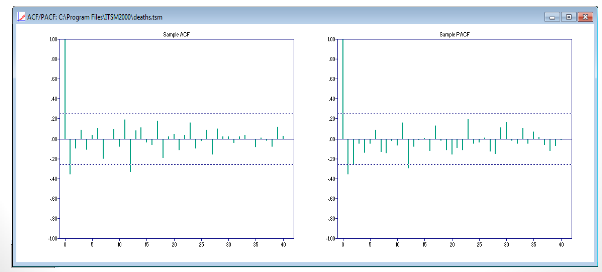
Difference at lag 12, and then at lag 1. This should remove seasonality at lag 12 and the quadratic trend. Subtract the mean of 28.8305.



57

The DEATHS.tsm Data

The ACF/PACF: Both show significant values at lag 12; this suggests that the seasonality changes from year to year. Lag 24 is not significant, so we might model this using a MA(1) in lag 12. We also note a significant value at lag 1; this shows short term variability within each year, which can be modeled with an MA(1) at lag 1. We also note that these values are both about -0.3.



58

The DEATHS.tsm Data

Recall from earlier work that we look at the ACF for MA models and the PACF for AR models. We have significant values in both; so why did I choose MA's rather than AR's to model this pattern?

Also recall that AR models damp slowly in the ACF, while MA models ALWAYS dampen completely in the ACF after the longest lag in the model. We see no evidence of slow damping in the ACF.

So, we will try a SARIMA(0,1,1)x(0,1,1)$_{12}$ model; in other words, we will fit a constrained MA(13) in lag 12, with the coefficient at lag 13 constrained to be the product of the coefficients at lag 1 and lag 12.

59

The DEATHS.tsm Data

We have already differenced at lags 12 and 1 (D = 1, d = 1, s = 12).

Choose **Model -> Specify**.

Enter an **MA(13)** model with $\theta_1 = \theta_{12} = -0.3$, $\theta_{13} = 0.09$, and all other coefficients set to 0. (This step replaces the preliminary estimation step we usually employ.) Click **OK**.

Choose **Model -> Estimation -> Max likelihood** and click on the **Constrain optimization** button. Specify **1 multiplicative relation**, and enter **1 x 12 = 13** in the boxes provided. (Note that the MA button automatically was selected; for a SARIMA model with both AR and MA terms you need to be careful to select the appropriate side to constrain.) Click **OK**. Click **OK** in the second dialog box also.

Repeat the Max likelihood estimation step as often as needed to get the estimates to "settle".

60

The DEATHS.tsm Data

Note that the standard Error for the lag 13 term is 0. This is because it is NOT estimated; it is computed from the estimates for the lag 1 and lag 12 terms.

Our model is $\nabla\nabla_{12}X_t = 28.8305 + (1 - 0.478397B)(1 - 0.590952B^{12})Z_t$

with $Z_t \sim WN(0, 94253.9)$

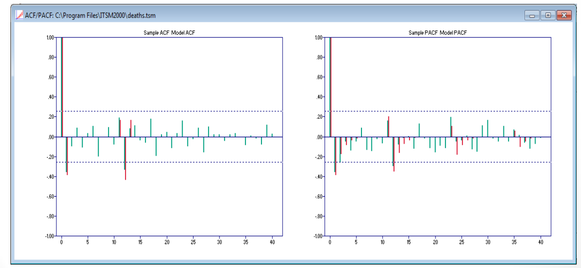
ML estimates: C:\Program Files\ITSM2000\deaths.tsm

```
WN Variance = .942539E+05

MA Coefficients
     -.478397      .000000      .000000      .000000
      .000000      .000000      .000000      .000000
      .000000      .000000      .000000     -.590952
      .282710

Standard Error of MA Coefficients
      .134650      .000000      .000000      .000000
      .000000      .000000      .000000      .000000
      .000000      .000000      .000000      .205710
      .000000

(Residual SS)/N = .942539E+05

AICC = .855534E+03
BIC  = .850983E+03

-2Log(Likelihood) = .849097E+03

Accuracy parameter = .00263000

Number of iterations = 1

Number of function evaluations = 19

Uncertain minimum.
```


The DEATHS.tsm Data

The AICC for our model is 855.534.

The sample/model ACF/PACF looks OK.



The DEATHS.tsm Data

The residual tests look OK. This appears to be a reasonable model.

Tests of randomness: C:\Program Files\ITSM2000\deaths.tsm

```
============================================
ITSM::(Tests of randomness on residuals)
============================================

Ljung - Box statistic = 19.765 Chi-Square ( 20 ), p-value = .47269

McLeod - Li statistic = 19.365 Chi-Square ( 22 ), p-value = .62268

# Turning points = 36.000~AN(38.000,sd = 3.1885), p-value = .53050

# Diff sign points = 33.000~AN(29.000,sd = 2.2361), p-value = .07364

Rank test statistic = .81300E+03~AN(.85550E+03,sd = 76.459), p-value = .57831

Jarque-Bera test statistic (for normality) = 2.0636 Chi-Square (2), p-value = .35636

Order of Min AICC YW Model for Residuals = 0
```


The DEATHS.tsm Data

Of course, we could have just fit an ARMA model to the differenced and mean-corrected data $\nabla\nabla_{12}X_t - 28.8305 = Y_t$

Using Autofit (both orders set to 15) suggests an MA(1) model.

ML estimates: C:\Program Files\ITSM2000\deaths.tsm

```
Method: Maximum Likelihood


ARMA Model:
X(t) = Z(t) - .5240 Z(t-1)

WN Variance = .123419E+06

MA Coefficients
     -.524017

Standard Error of MA Coefficients
      .126869

(Residual SS)/N = .123419E+06

AICC = .863647E+03
BIC  = .862741E+03

-2Log(Likelihood) = .859433E+03

Accuracy parameter = .00525000

Number of iterations = 1

Number of function evaluations = 11

Uncertain minimum.
```

The DEATHS.tsm Data

The AICC for this model is 863.867, which is not as good as our SARIMA model.

Note that this model can be written using SARIMA notation as SARIMA(0,1,1)x$(0,1,0)_{12}$.

Let's try a SARIMA(0,1,13)x$(0,1,0)_{12}$ without the multiplicative structure.

After differencing (lags 12 and 1) and mean correcting, fit an MA(13) model.

65

The DEATHS.tsm Data

We note that the coefficients for terms 2, 3, 8, 10, and 11 are smaller in magnitude than their standard errors, so we constrain them to 0 and try again.

```
ML estimates: C:\Program Files\ITSM2000\deaths.tsm
WN Variance = .791332E+05

MA Coefficients
     -.516513      -.015528      -.000491      -.222435
      .126968      -.128064      -.194148      -.078579
      .221366      -.010813       .051683      -.537173
      .318321

Standard Error of MA Coefficients
      .123417       .121913       .121744       .121736
      .121769       .122459       .120976       .122459
      .121769       .121736       .121744       .121913
      .123417

(Residual SS)/N = .791332E+05

AICC = .878941E+03
BIC  = .863048E+03

-2Log(Likelihood) = .841395E+03

Accuracy parameter = .000770000

Number of iterations = 1

Number of function evaluations = 205

Uncertain minimum.
```

66

The DEATHS.tsm Data

Now the coefficients for terms 4, 5, and 7 are less in magnitude than their standard errors, so we set them to 0 and try again.

```
ML estimates: C:\Program Files\ITSM2000\deaths.tsm
WN Variance = .644348E+05

MA Coefficients
     -.598909       .000000       .000000      -.150348
      .146284      -.425384      -.098376       .000000
      .205929       .000000       .000000      -.726690
      .525961

Standard Error of MA Coefficients
      .201497       .000000       .000000       .171892
      .154970       .285762       .199146       .000000
      .158108       .000000       .000000       .193877
      .274054

(Residual SS)/N = .644348E+05

AICC = .862577E+03
BIC  = .846697E+03

-2Log(Likelihood) = .840904E+03

Accuracy parameter = .00141000

Number of iterations = 1

Number of function evaluations = 88

Uncertain minimum.
```

67

The DEATHS.tsm Data

The coefficient for term 9 is just barely larger in magnitude than its' standard error, so lets zero it and see if AICC gets better. Our current model has AICC = 856.485.

```
ML estimates: C:\Program Files\ITSM2000\deaths.tsm
WN Variance = .640111E+05

MA Coefficients
     -.629246       .000000       .000000       .000000
      .000000      -.473307       .000000       .000000
      .161532       .000000       .000000      -.738588
      .544338

Standard Error of MA Coefficients
      .145071       .000000       .000000       .000000
      .000000       .192802       .000000       .000000
      .127087       .000000       .000000       .178905
      .233197

(Residual SS)/N = .640111E+05

AICC = .856485E+03
BIC  = .839128E+03

-2Log(Likelihood) = .842869E+03

Accuracy parameter = .00228000

Number of iterations = 1

Number of function evaluations = 49

Uncertain minimum.
```

68

The DEATHS.tsm Data

This model has AICC = 855.605, and all coefficients are fairly large compared to their standard errors (maybe not lag 6).

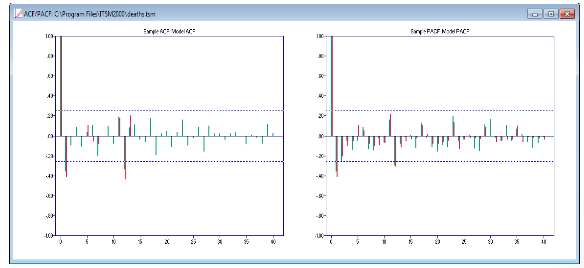
ML estimates: C:\Program Files\ITSM2000\deaths.tsm

```
WN Variance = .712778E+05

MA Coefficients
     -.595661        .000000        .000000        .000000
      .000000       -.406240        .000000        .000000
      .000000        .000000        .000000       -.685311
      .459061

Standard Error of MA Coefficients
      .147234        .000000        .000000        .000000
      .000000        .299156        .000000        .000000
      .000000        .000000        .000000        .246033
      .288229

(Residual SS)/N = .712778E+05

AICC = .855605E+03
BIC  = .841878E+03

-2Log(Likelihood) = .844473E+03

Accuracy parameter = .00308000

Number of iterations = 1

Number of function evaluations = 38

Uncertain minimum.
```

69

The DEATHS.tsm Data

The sample/model ACF/PACF looks good.



70

The DEATHS.tsm Data

The residual tests are OK except for the # Diff sign points test.

Tests of randomness: C:\Program Files\ITSM2000\deaths.tsm

```
============================================
ITSM::(Tests of randomness on residuals)
============================================

Ljung - Box statistic = 18.083 Chi-Square ( 20 ), p-value = .58194

McLeod - Li statistic = 24.053 Chi-Square ( 24 ), p-value = .45859

# Turning points = 36.000~AN(38.000,sd = 3.1885), p-value = .53050

# Diff sign points = 34.000~AN(29.000,sd = 2.2361), p-value = .02535

Rank test statistic = .80100E+03~AN(.85550E+03,sd = 76.459), p-value = .47597

Jarque-Bera test statistic (for normality) = 1.4591 Chi-Square (2), p-value = .48213

Order of Min AICC YW Model for Residuals = 0
```

71

The DEATHS.tsm Data

The AICC's for our two best models:

SARIMA(0,1,1)x(0,1,1)$_{12}$: AICC = 855.534

SARIMA(0,1,13)x(0,1,0)$_{12}$ (subset): AICC = 856.485

The multiplicative model is slightly better, but not much. Also, it had better looking residuals.

72

The DEATHS.tsm Data

If we get rid of the lag 6 term and optimize we get a model with AICC = 857.364; not quite as good. The sample/model ACF/PACF and residual tests are almost the same.

The model is:

$$\nabla\nabla_{12}X_t = 28.8305 + \left(1 - 0.459509B - 0.653828B^{12} + 0.127346B^{13}\right)Z_t$$

which is NOT the same as our multiplicative model:

$$\nabla\nabla_{12}X_t = 28.8305 + \left(1 - 0.478397B\right)\left(1 - 0.590952B^{12}\right)Z_t$$

73

So… is it all worth it?

Sometimes!

No guarantees!

Have a great break!

74