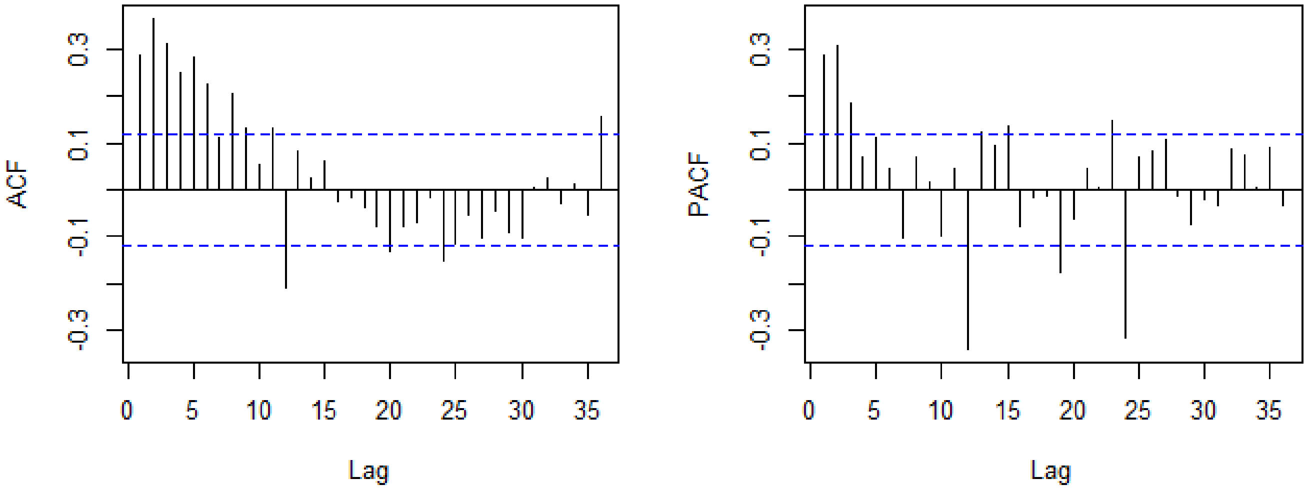
101 help

# Estimate ARMA coefficients through ACF and PACF inspection

I know that this is probably a question that's been asked plenty of times, but i haven't seen an answer that's both accurate and simple. How do you estimate the appropriate forecast model for a time series by visual inspection of the ACF and PACF plots? Which one, ACF or PACF, tells the AR or the MA (or do they both?) Which part of the graphs tell you the seasonal and non seasonal part for a seasonal ARIMA?

Take for instance these functions:




They show the ACF and PCF of a log transformed series that's been differenced twice, one simple difference and one seasonal.

How would you caracterize it? What model best fits it?

Thanks in advance!

**EDIT:** Added raw data

Original data: here

Log transformed data: here

**EDIT:** Corrected ACF and PACF functions (previous ones were overdifferentiated) r regression time-series machine-learning mathematical-statistics

edited Jul 7 '14 at 12:52

asked Jul 7 '14 at 8:38

4everlearning
75 1 2 8

## 2 Answers

My answer is really an abridgement of javlacelle's but is is too long for a simple comment but not too short to be useless.

While jvlacelle's response is technically correct at one level it "overly simplifies" as it premises certain "things" which are normally never true . It assumes that there is no deterministic structure required such as one or more time trends OR one or more level shifts or one or more seasonal pulses or one or more one-time pulses. Furthermore it assumes that the parameters of the identified model are invariant over time and the error process underlying the tentatively identified model is also invariant over time. Ignoring any of the above is often (always in my opinion !) a recipe for disaster or more precisely a "poorly identified model". A classic case of this is the unnecessary logarithmic transformation proposed for the airline series and for the series that the OP presents in his revised question. There is no need for any logarithmic transformation for his data as there are just a few "unusual" values at periods 198,207,218,219 and 256 which left untreated create the false impression that there is higher error variance with higher levels. Note that "unusual values" are identified taking into account any needed ARIMA structure which often escapes the human eye.Transformations are needed when the error variance is non-constant over time NOT when the variance of the the observed Y is non-constant over time. Primitive procedures still make the tactical error of selecting a transformation prematurely prior to any of the aforementioned remedies. One has to remember that the simple-minded ARIMA model identification strategy was developed in the early 60's BUT a lot of development/improvements have gone on since then.

Edited after data was posted :

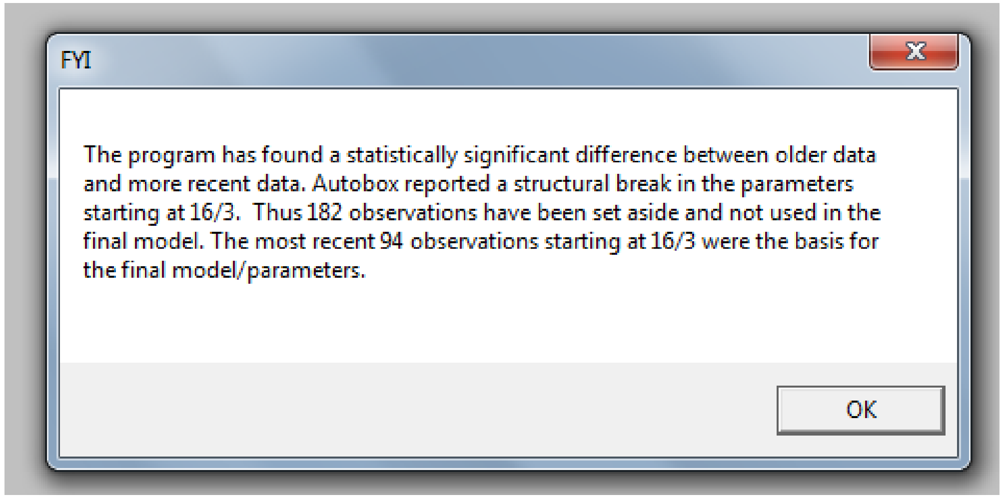
A reasonable model was identified using http://www.autobox.com/cms/ which is a piece of software that incorporates some of my aforementioned ideas as I helped develop it.

```
      MODEL COMPONENT         LAG    COEFF      STANDARD     P       T
 #                           (BOP)               ERROR     VALUE   VALUE

    Differencing               12
   1CONSTANT                          .949E+04   .359E+04   .0086    2.64
   2Autoregressive-Factor #  1   1    .399       .617E-01   .0000    6.46
   3                             2    .300       .607E-01   .0000    4.93
   4Autoregressive-Factor #  2  12   -.356       .596E-01   .0000   -5.97
```

# TESTING PARAMETER CONSTANCY

```
DIAGNOSTIC CHECK #4: THE CHOW PARAMETER CONSTANCY TEST
  The Critical value used for this test :      .01
  The minimum group or interval size was:       94

         F TEST TO VERIFY CONSTANCY OF PARAMETERS

CANDIDATE BREAKPOINT       F VALUE          P VALUE

     95    8/ 11           8.3012          .0000030209
    103    9/  7           5.7387          .0002083766
    111   10/  3           4.3902          .0019670310
    119   10/ 11           3.5675          .0076813814
    127   11/  7           5.6891          .0002263213
    135   12/  3           7.2626          .0000166498
    143   12/ 11           7.7626          .0000073072
    151   13/  7           4.4136          .0018919682
    159   14/  3           5.9658          .0001428191
    167   14/ 11           5.3783          .0003797101
    175   15/  7           5.6984          .0002228187
    183   16/  3           6.0916          .0001158625*
```

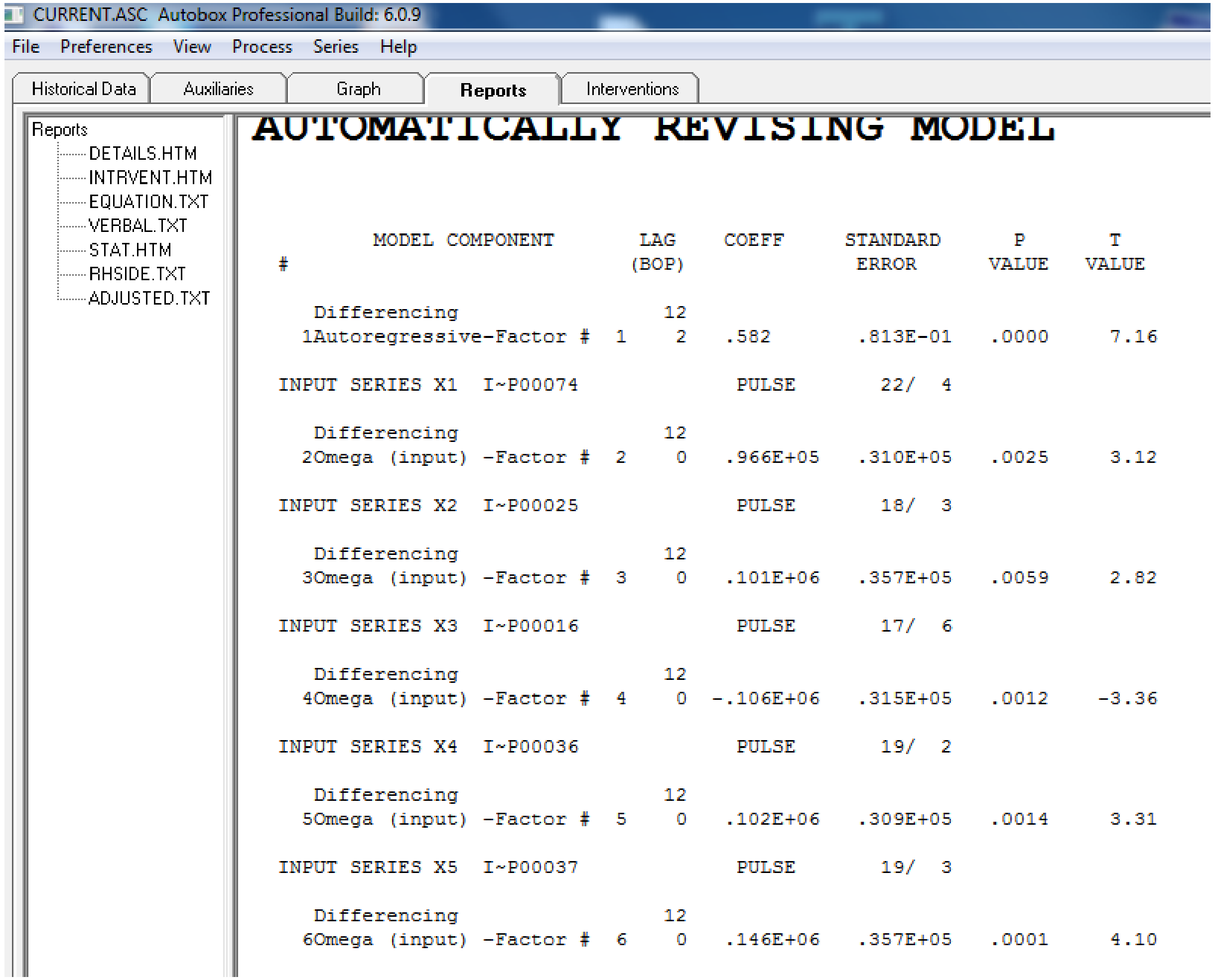
The Chow Test for parameter constancy suggested that the data be segmented and that the last 94 observations be used as model parameters had changed over time.



.These last 94 values yielded an equation

```
  AUTOMATIC FORECASTING SYSTEMS
         HATBORO PA 19040
           215-675-0652
      VERSION: 06/26/2014 06:15



 MODELLING OUTPUT SERIES:SERIES



 [(1-B**12)]Y(T) = +[X1(T)][(1-B**12)][(+  96555.    )]   :PULSE           22/  4   256
                   +[X2(T)][(1-B**12)][(+  .10068E+06)]   :PULSE           18/  3   207
                   +[X3(T)][(1-B**12)][(-  .10595E+06)]   :PULSE           17/  6   198
                   +[X4(T)][(1-B**12)][(+  .10228E+06)]   :PULSE           19/  2   218
                   +[X5(T)][(1-B**12)][(+  .14632E+06)]   :PULSE           19/  3   219
                  +      [(1-  .582B** 2)]**-1  [A(T)]
```

with all coefficients being significant.


CURRENT.ASC Autobox Professional Build: 6.0.9

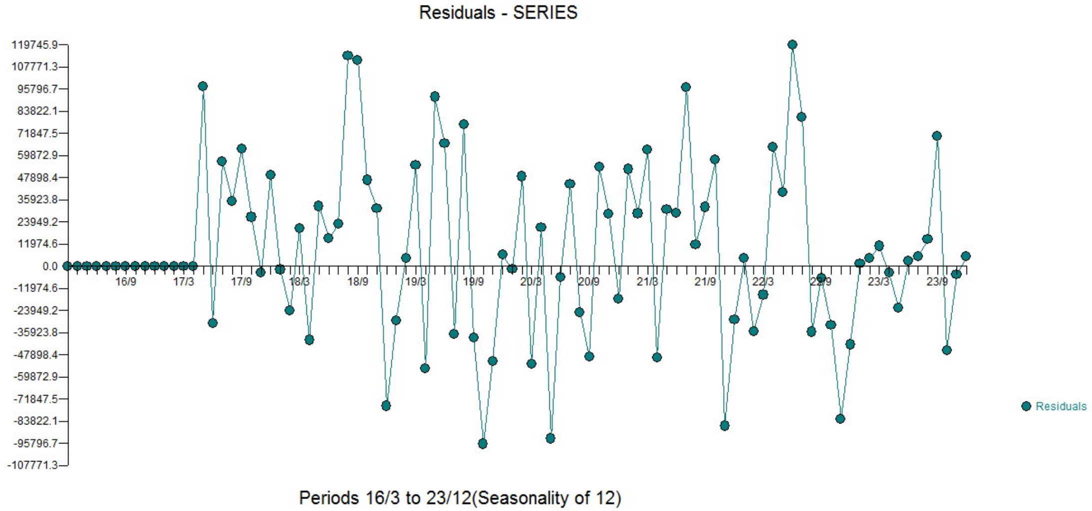
File Preferences View Process Series Help

Historical Data | Auxiliaries | Graph | **Reports** | Interventions

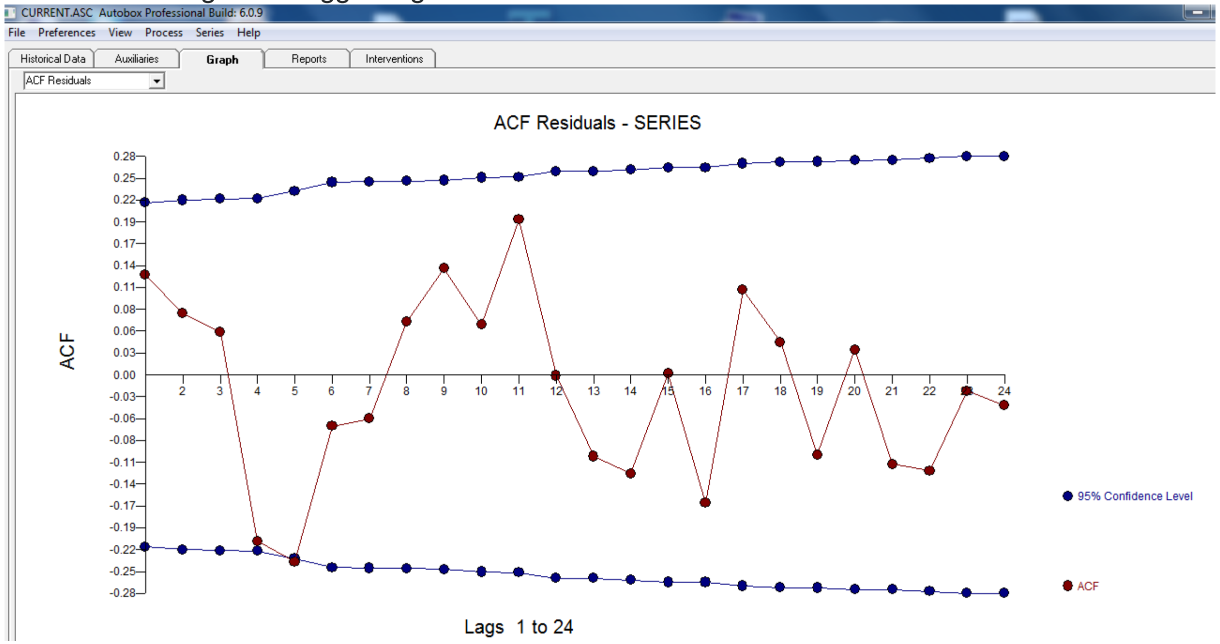
Reports
- DETAILS.HTM
- INTRVENT.HTM
- EQUATION.TXT
- VERBAL.TXT
- STAT.HTM
- RHSIDE.TXT
- ADJUSTED.TXT

```
AUTOMATICALLY REVISING MODEL

          MODEL COMPONENT          LAG     COEFF      STANDARD    P       T
  #                               (BOP)                ERROR     VALUE   VALUE

     Differencing                   12
    1Autoregressive-Factor #  1      2    .582        .813E-01   .0000    7.16

  INPUT SERIES X1   I~P00074              PULSE        22/  4

     Differencing                   12
    2Omega (input) -Factor #  2      0    .966E+05    .310E+05   .0025    3.12

  INPUT SERIES X2   I~P00025              PULSE        18/  3

     Differencing                   12
    3Omega (input) -Factor #  3      0    .101E+06    .357E+05   .0059    2.82

  INPUT SERIES X3   I~P00016              PULSE        17/  6

     Differencing                   12
    4Omega (input) -Factor #  4      0   -.106E+06    .315E+05   .0012   -3.36

  INPUT SERIES X4   I~P00036              PULSE        19/  2

     Differencing                   12
    5Omega (input) -Factor #  5      0    .102E+06    .309E+05   .0014    3.31

  INPUT SERIES X5   I~P00037              PULSE        19/  3

     Differencing                   12
    6Omega (input) -Factor #  6      0    .146E+06    .357E+05   .0001    4.10
```

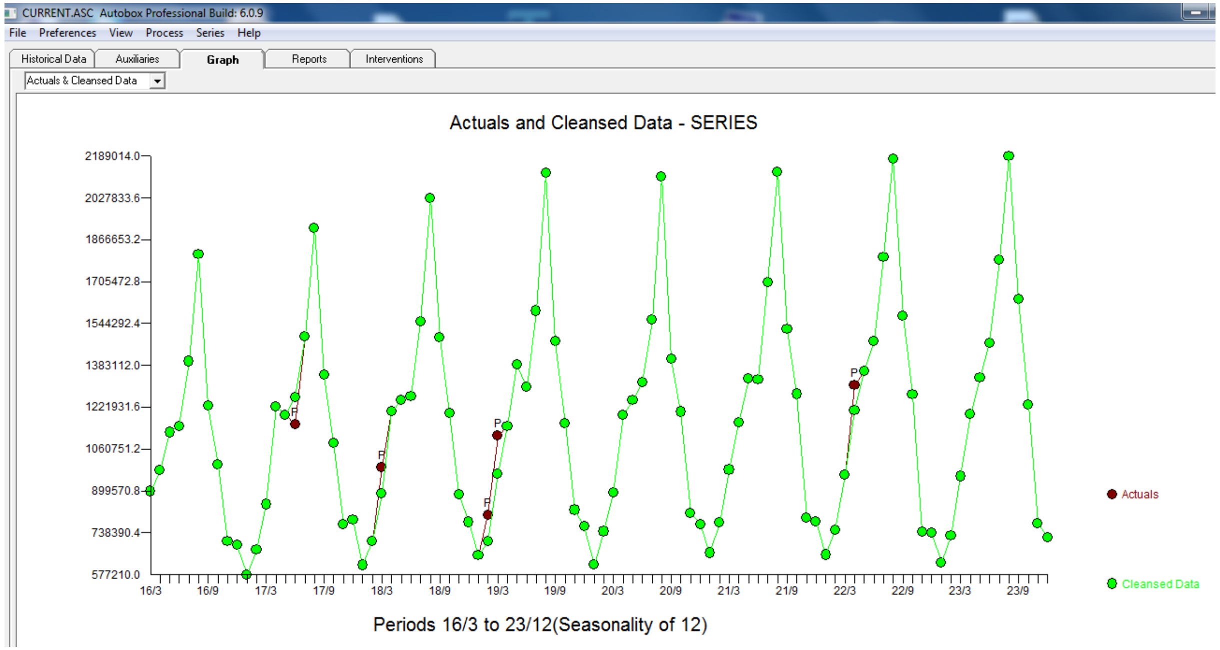
. The plot of the residuals suggests a reasonable scatter



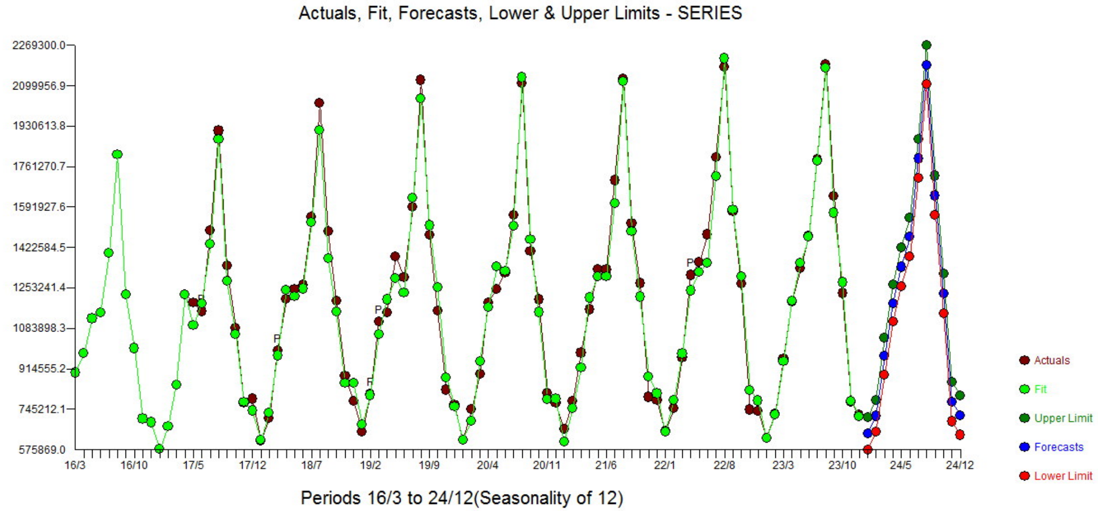

with the following ACF suggesting randomness




. THe Actual and Cleansed graph is illuminating as it shows the subtle BUT significant outliers.




. Finally a plot of actual,fit and forecast summarizes our work ALL WITHOUT TAKING LOGARITHMS




. It is well known but often forgotten that power transforms are like drugs .... unwarranted usage can harm you. Finally notice that the model has an AR(2) BUT not an AR(1) structure.

edited Jul 7 '14 at 22:21 answered Jul 7 '14 at 11:01

IrishStat
**14.8k** 1 16 31

Very nice analysis, thanks. – javlacalle Jul 8 '14 at 0:28

+1 excellent as always – forecaster Jul 8 '14 at 3:35

---

Just to clear up concepts, by visual inspection of the ACF or PACF you can choose (not estimate) a tentative ARMA model. Once a model is selected you can estimate the model by maximizing the likelihood function, minimizing the sum of squares or, in the case of the AR model, by means of the method of moments.

An ARMA model can be chosen upon inspection of the ACF and PACF. This approach relies on the following facts: 1) the ACF of a stationary AR process of order p goes to zero at an exponential rate, while the PACF becomes zero after lag p. 2) For an MA process of order q the theoretical ACF and PACF exhibit the reverse behaviour (the ACF truncates after lag q and the PACF goes to zero relatively quickly).

It is usually clear to detect the order of an AR or MA model. However, with processes that include both an AR and MA part the lag at which they are truncated may be blurred because both the ACF and PACF will decay to zero.

One way to proceed is to fit first an AR or MA model (the one that seems more clear in the ACF and PACF) of low order. Then, if there is some further structure it will show up in the residuals, so the ACF and PACF of the residuals is checked to determine if additional AR or MA terms are necessary.

Usually you will have to try and diagnose more than one model. You can also compare them by looking at the AIC.

The ACF and PACF that you posted first suggested an ARMA(2,0,0)(0,0,1), that is, a regular AR(2) and a seasonal MA(1). The seasonal part of the model is determined similarly as the regular part but looking at the lags of seasonal order (e.g. 12, 24, 36,... in monthly data). If you are using R it is recommended to increase the default number of lags that are displayed, `acf(x, lag.max = 60)`.

The plot that you show now reveals suspicious negative correlation. If this plot is based on the same that as the previous plot you may have taken too many differences. See also this post.

You can get further details, among other sources, here: Chapter 3 in Time Series: Theory and Methods by Peter J. Brockwell and Richard A. Davis and here.

edited Jul 7 '14 at 11:14

answered Jul 7 '14 at 10:40

javlacalle
**6,558** 9 27

1 Thank you very much. I added the data to the question. – 4everlearning Jul 7 '14 at 10:56

You're right. I may have taken one difference too many. I have one doubt though. I did a simple difference (i.imgur.com/1MjLzlX.png) and a seasonal (12) one (i.imgur.com/E64Sd7p.png) both on the log data. Which one do should i look at, the seasonal one right? – 4everlearning Jul 7 '14 at 11:45

1 @4everlearning Right, after taking seasonal differences the ACF and PACF looks closer to what we could expect for a stationary process. You can start by fitting an ARIMA(2,0,0)(0,1,1), in R `arima(x, order = c(2,0,0), seasonal = list(order = c(0,1,1)))` , and displaying the ACF and PACF of the residuals. Be also aware of further issues raised by IrishStat that you should be concerned with in the analysis. – javlacalle Jul 7 '14 at 11:59

Thanks. How would i go about finding those AR and MA orders? Plus, the Akaike Information Criterion yields a negative value for my model. I understand that this is not important although i'm not really sure how to compare it to other models, say AIC=-797.74 and AIC=-800.00. Which is preferable? – 4everlearning Jul 7 '14 at 12:49

You can determine the orders following the idea given in the answer above. If you see that the ACF goes to zero relatively quickly and the PACF truncates after lag 2 then it is probably that an AR(2) structure is present in the data. The reverse idea applies to detect an MA. As a general recommendation, start with a model of low order and inspect the residuals looking for AR or MA structures to be added to the initial model. – javlacalle Jul 7 '14 at 13:02