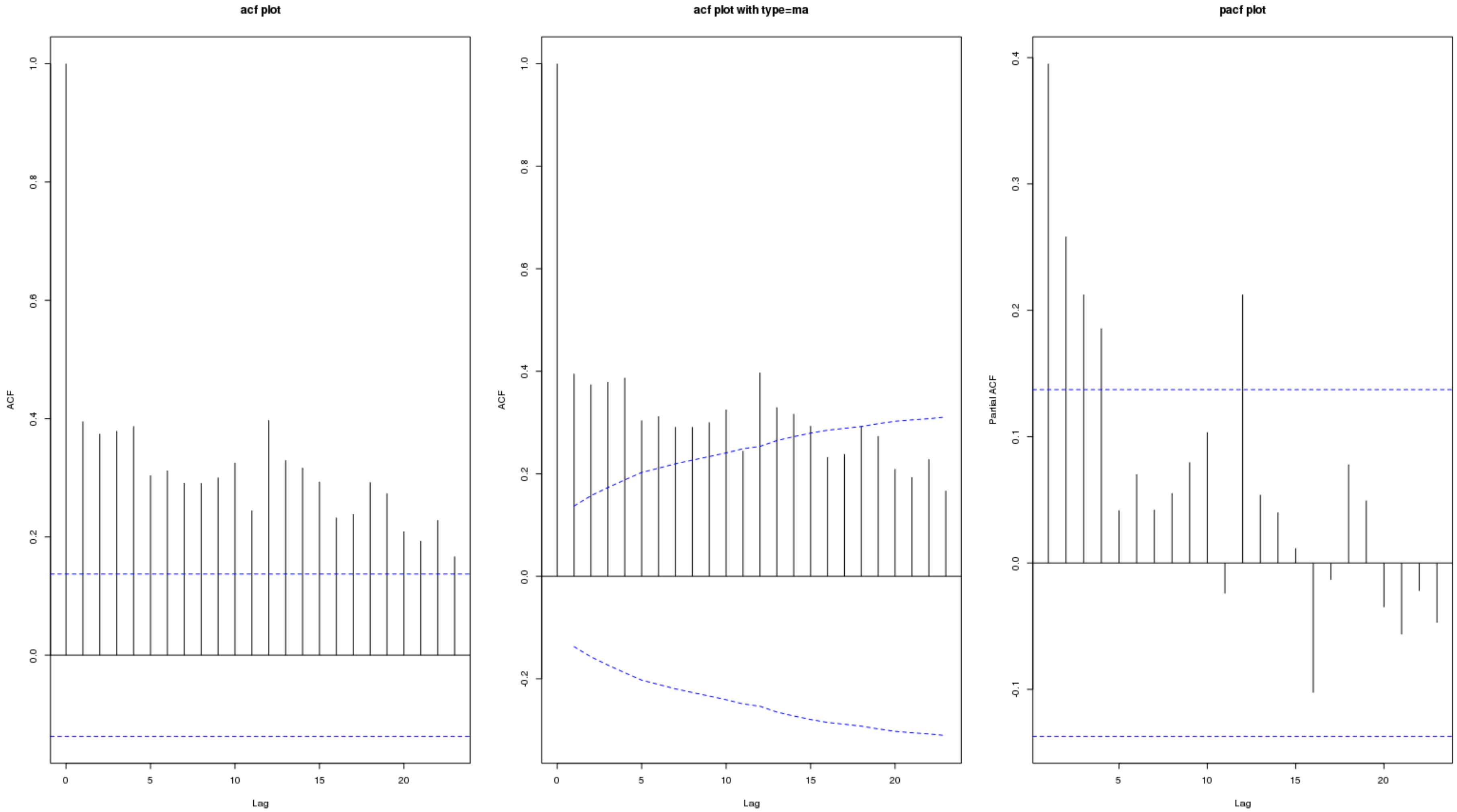
How to interpret these acf and pacf plots

Following are acf and pacf plots of a monthly data series. The second plot is acf with ci.type='ma':



The persistence of high values in acf plot probably represent a long term positive trend. The question is if this represent seasonal variation?

I tried to see different sites on this topic but I am not sure if these plots show seasonality.

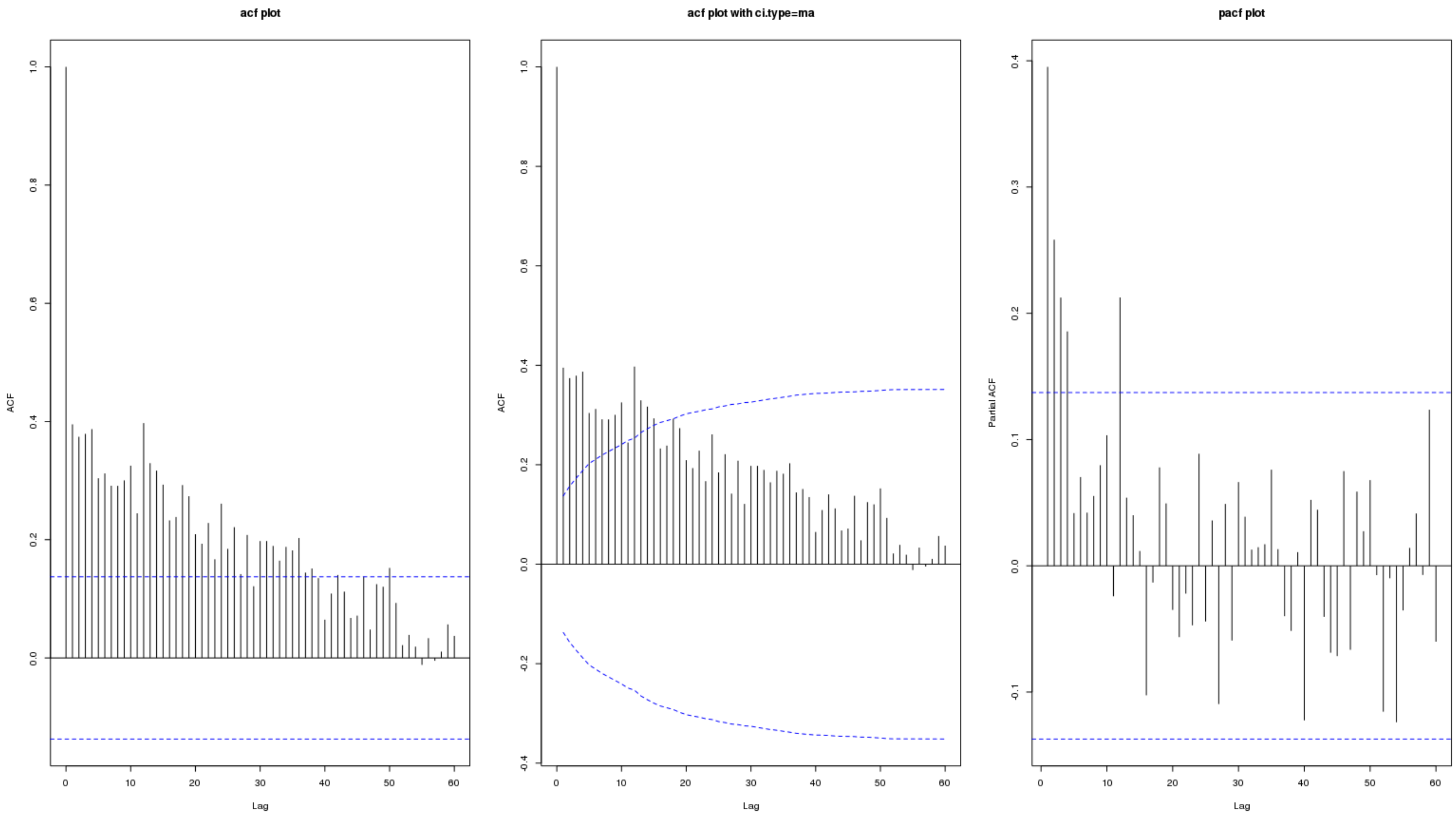
ACF and PACF plot analysis

Help interpreting ACF- and PACF-plots

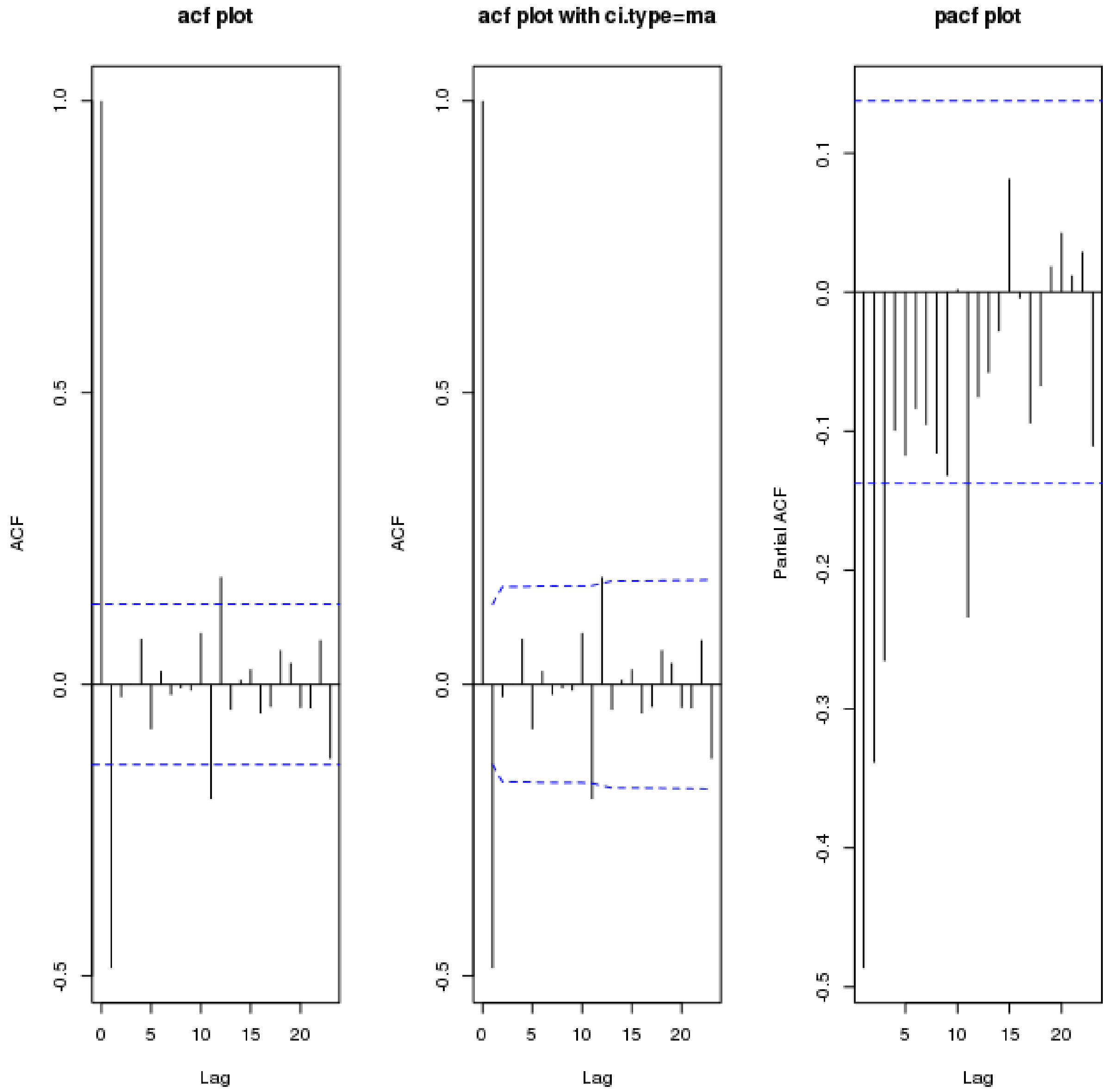
Help understanding the following picture of ACF

Autocorrelation and partial autocorrelation interpretation

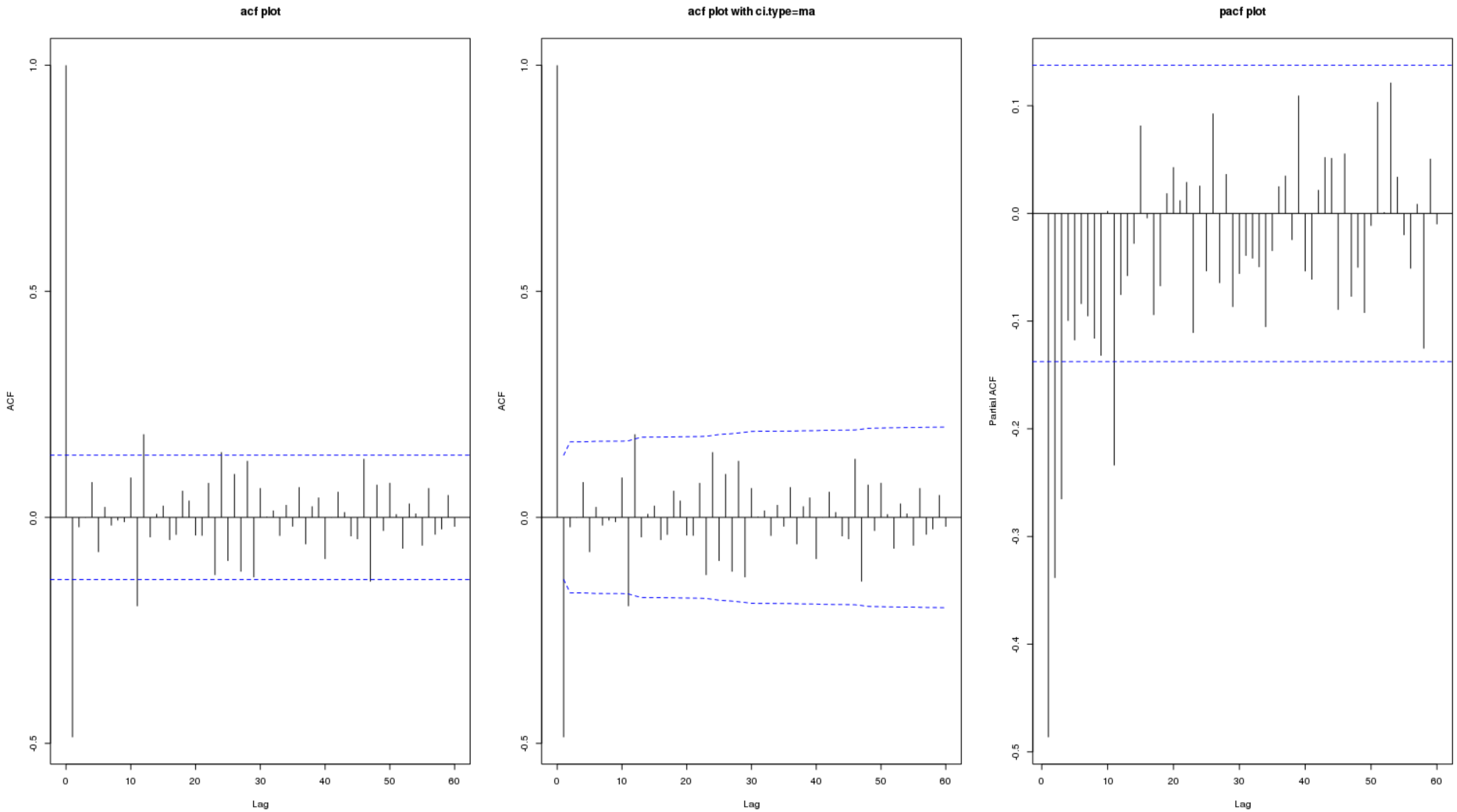
Edit: following is the graph for lag upto 60:



Following are plots of diff(my_series):



And upto lag 60:



Edit: This data is from: Is this an appropriate method to test for seasonal effects in suicide count data? Here the contributors did not consider acf and pacf plot of original or differenced series worth mentioning (so it must not be important). Only acf/pacf plots of residuals was referred to in a couple of places.

time-series

edited Apr 19 '15 at 10:40

asked Apr 18 '15 at 16:56

rnso
2,807 2 14 34

1 Can you add something about your data (eg, a basic plot)? Did you try anything like `stl()` ? – gung Apr 18 '15 at 17:01

I am trying to understand how to determine seasonality from acf and pacf plots. Is review of basic plot or stl necessary for this? Can we not determine something from these plots? – rnso Apr 18 '15 at 17:09

1 That would be fine. For clarity, your question isn't really about what is going on w/ your data, but is about what can be understood from these plots in isolation, is that right? – gung Apr 18 '15 at 17:11

1 Yes. I often need to determine if seasonality is present in my data so I want to understand what information can I derive from acf and pacf plots. The plots of stl function is reasonably easy to understand but not these plots. – rnso Apr 18 '15 at 17:18

Your data does indeed have some seasonality. Please see my response to @javlacalle . – IrishStat Apr 20 '15 at 15:36

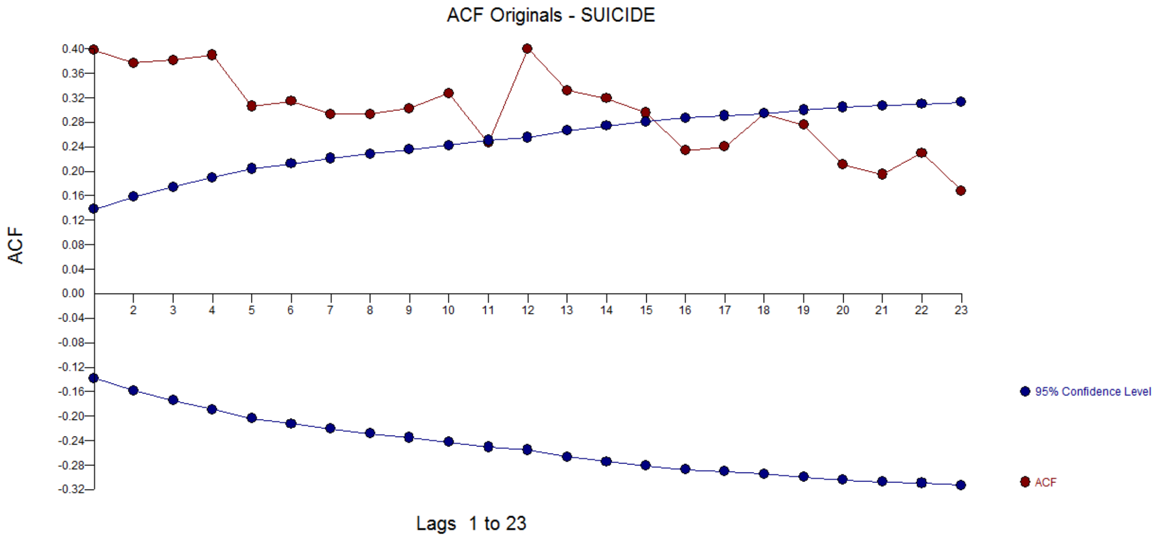
2 Answers

looking at plots in order to try to pigeonhole the data into a guessed arima model works well when 1: There are no outliers/pulses/level shifts, local time trends and no seasonal deterministic pulses in the data AND 2) when the arima model has constant parameters over time AND 3) when the error variance from the arima model has constant variance over time. When do these three things hold .... in most textbook data sets presenting the ease of arima modelling. When do 1 or more of the 3 not hold .... in every real world data set that I have ever seen . The simple answer to your question requires access to the original facts (the historical data) not the secondary descriptive information in your plots. But this is just my opinion!

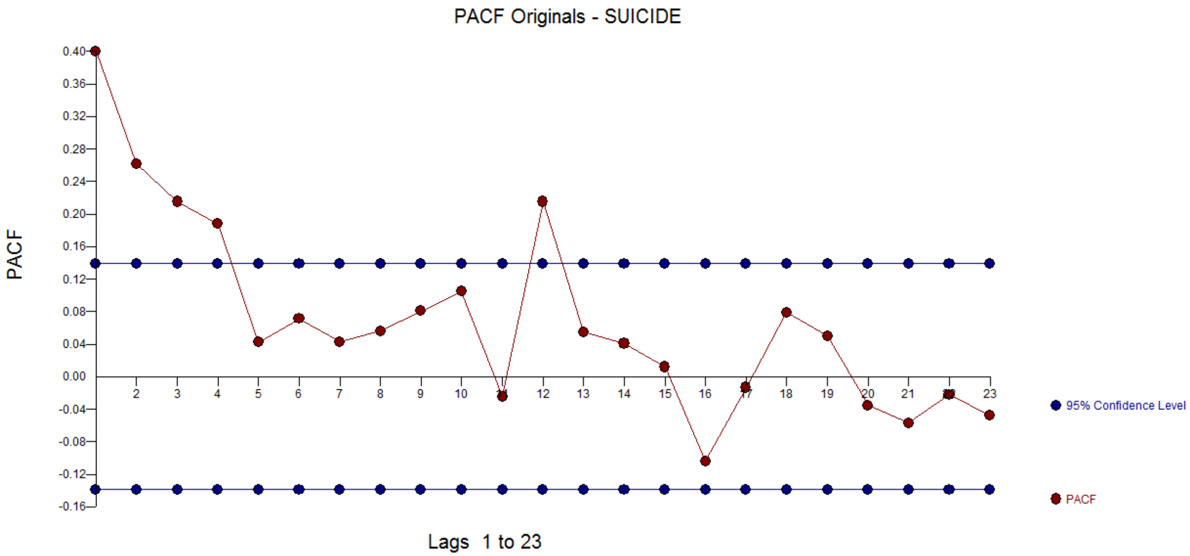
EDITED AFTER RECEIPT OF DATA:

I was on a Greek vacation (actually doing something other than time series analysis) and was unable to analyse the SUICIDE DATA but in conjunction with this post. It is now fitting and right that I submit an analysis to follow up/prove by example my comments about multi-stage model identification strategies and the failings of simple visual analysis of simple correlation plots as "the proof is in the pudding".

Here is the ACF of the original data



The PACF of the original series



. AUTOBOX http://www.autobox.com/cms/ a piece of software that I have helped developed uses heuristics to identify a starting model In this case the initially identified model was found to be

THE INITIAL MODEL

#	MODEL COMPONENT	LAG (BOP)	COEFF	STANDARD ERROR	P VALUE	T VALUE
1	CONSTANT		68.2			
2	Autoregressive-Factor # 1	1	.293			
3		2	.258			
4	Autoregressive-Factor # 2	12	.397			

. Diagnostic checking of the residuals from this model suggested some model augmentation using a level shift, pulses and a seasonal pulse Note that the Level Shift is detected at or about period 164 which is nearly identical to an earlier conclusion about period 176 from @forecaster. All roads do not lead to Rome but some can get you close !

```
#                            (BOP)                ERROR      VALUE    VALUE

  1CONSTANT                          39.7         5.24       .0000     7.56
  2Autoregressive-Factor #  1     1   .240        .713E-01   .0009     3.36
  3Autoregressive-Factor #  2    12   .213        .755E-01   .0052     2.83

INPUT SERIES X1   I~L00164            LEVEL       2008/  8

  4Omega (input) -Factor #  3     0  11.4         1.93       .0000     5.94

INPUT SERIES X2   I~P00042            PULSE       1998/  6

  5Omega (input) -Factor #  4     0 -25.0         6.75       .0003    -3.71

INPUT SERIES X3   I~P00201            PULSE       2011/  9

  6Omega (input) -Factor #  5     0  24.7         6.93       .0005     3.56

INPUT SERIES X4   I~P00048            PULSE       1998/ 12

  7Omega (input) -Factor #  6     0 -22.2         6.73       .0011    -3.30

INPUT SERIES X5   I~S00098            SEASP       2003/  2

  8Omega (input) -Factor #  7     0  -9.62        2.87       .0010    -3.35

INPUT SERIES X6   I~P00016            PULSE       1996/  4

  9Omega (input) -Factor #  8     0 -16.6         6.75       .0148    -2.46

INPUT SERIES X7   I~P00163            PULSE       2008/  7

 10Omega (input) -Factor #  9     0  27.8         6.99       .0001     3.97

INPUT SERIES X8   I~P00197            PULSE       2011/  5

 11Omega (input) -Factor # 10     0  21.2         6.90       .0024     3.07
```

. Testing for parameter constancy rejected parameter changes over time . Checking for deterministic changes in the error variance concluded that no deterministic changes were detected in the error variance.

```
***** RESULTS OF TSAY TEST *****

   REGION 1 VAR       REGION 2 VAR              F VAL       PROBABLITY

   53     50.189      151     49.494          .9861523       .4585205
   93     51.595      111     48.230          .9347913       .3684136
 * 133    45.119       71     57.280         1.2695264       .1245736
```

```
DIAGNOSTIC CHECK #5: THE TSAY VARIANCE CONSTANCY TEST

   The Critical value used for this test :     .01
   The minimum group or interval size was:      40
  There are 0 variance changes at this alpha level
```

.

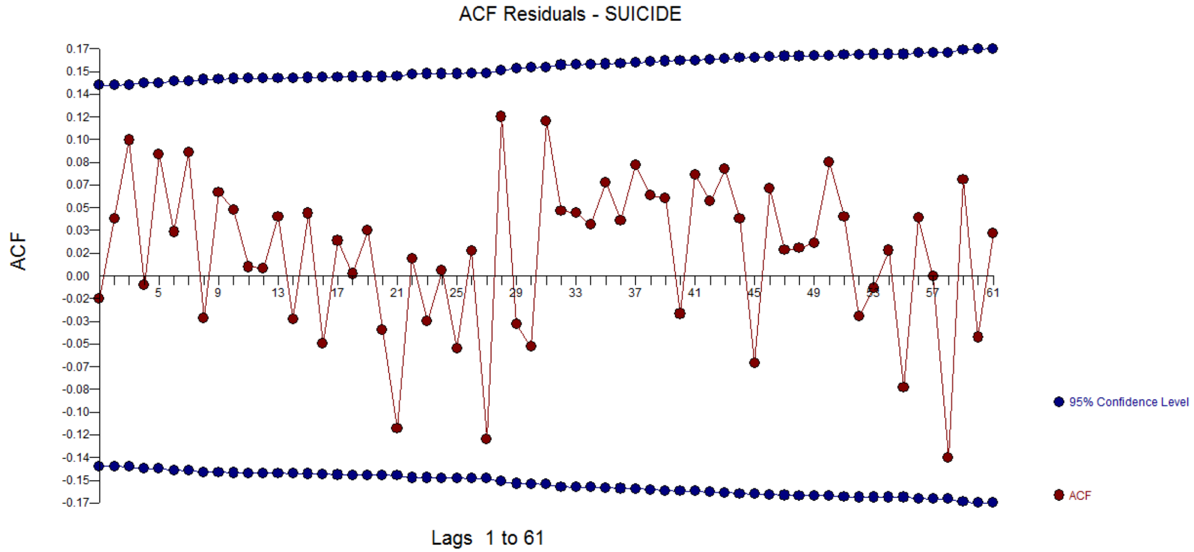
The Box-Cox test for the need for a power transform was positive with the conclusion that a logarithmic transform was necessary.

```
 EVALUATING POWER TRANSFORMATIONS VIA THE BOX-COX TEST.
           DIAGNOSTIC CHECK #6: THE BOX-COX POWER TRANSFORMATION TEST



MODEL STAGE:  81    12EST    2

      BOX-COX:   TRIAL   CURRENT    PREV BEST SOS  TRANSFORMATION

                  1       .95E+04       .95E+04     1.0
                  2       .94E+04       .95E+04      .0

                   THE OPTIMAL TRANSFORM TO DECOUPLE THE FIRST AND
                   SECOND MOMENT (POWER TRANSFORM/BOX-COX TEST) IS =   .00

                                1.0  =  NO TRANSFORM
                                 .5  =  SQUARE ROOT
                                 .0  =  LOGARITHMS
                               - .5  =  RECIPROCAL SQUARE ROOT
                               -1.0  =  RECIPROCAL
```

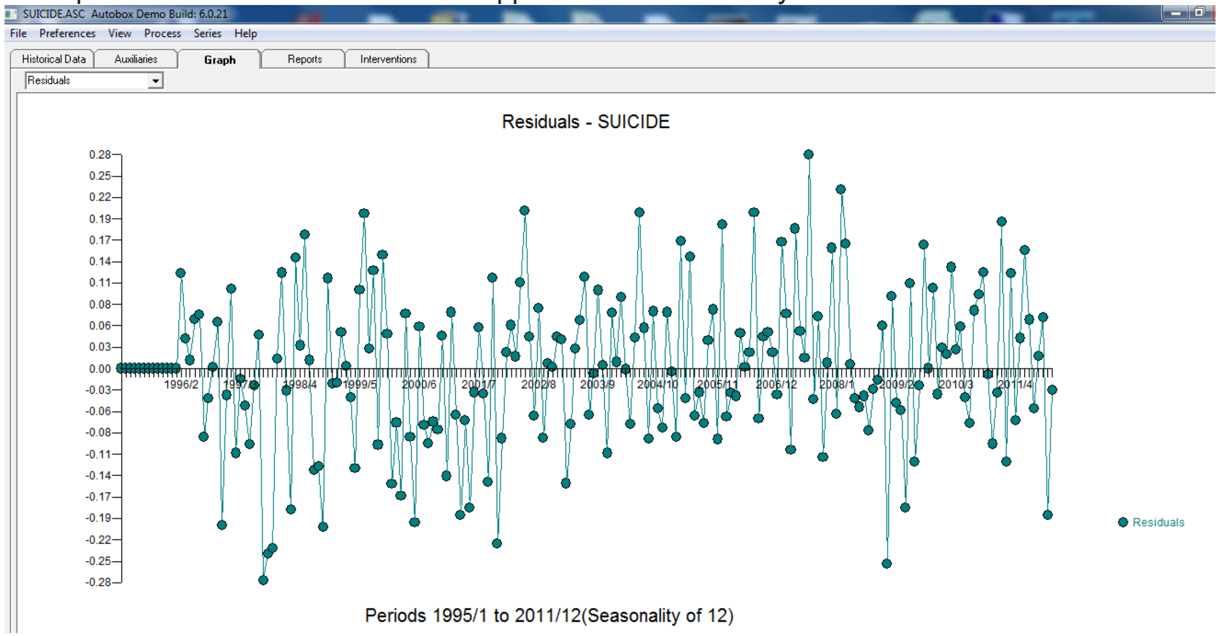
. The final model is here

```
                         LOGARITHMIC TRANSFORMATION

        MODEL COMPONENT        LAG    COEFF      STANDARD      P       T
 #                            (BOP)               ERROR      VALUE   VALUE
   Lambda Value                        .000
  1CONSTANT                            2.39       .318       .0000    7.53
  2Autoregressive-Factor #  1     1    .268       .706E-01   .0002    3.79
  3Autoregressive-Factor #  2    12    .219       .731E-01   .0030    3.00

 INPUT SERIES X1  I~L00164              LEVEL     2008/  8

  4Omega (input) -Factor #  3     0    .161       .296E-01   .0000    5.42

 INPUT SERIES X2  I~P00042              PULSE     1998/  6

  5Omega (input) -Factor #  4     0  -.437        .989E-01   .0000   -4.42

 INPUT SERIES X3  I~P00201              PULSE     2011/  9

  6Omega (input) -Factor #  5     0    .268       .101       .0088    2.65

 INPUT SERIES X4  I~P00048              PULSE     1998/ 12

  7Omega (input) -Factor #  6     0  -.404        .986E-01   .0001   -4.09

 INPUT SERIES X5  I~S00098              SEASP     2003/  2

  8Omega (input) -Factor #  7     0  -.148        .422E-01   .0006   -3.50

 INPUT SERIES X6  I~P00016              PULSE     1996/  4

  9Omega (input) -Factor #  8     0  -.274        .988E-01   .0062   -2.77

 INPUT SERIES X7  I~P00163              PULSE     2008/  7

 10Omega (input) -Factor #  9     0    .353       .102       .0006    3.48

 INPUT SERIES X8  I~P00197              PULSE     2011/  5
```

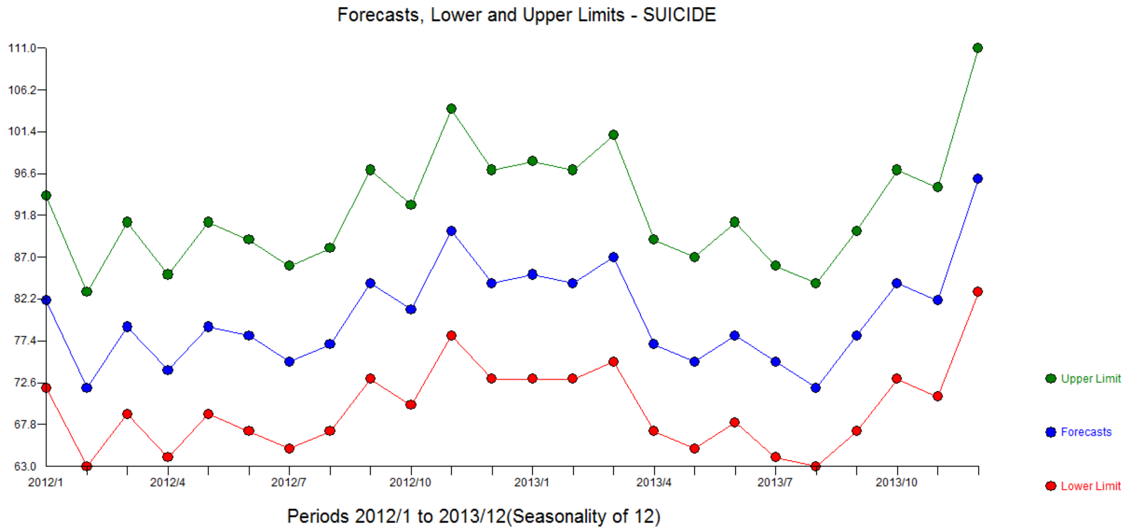
. The residuals from the final model appear to be free of any autocorrelation



. The plot of the final models residuals appears to be free of any Gaussian Violations



. The plot of Actual/Fit/Forecasts is here


Forecasts, Lower and Upper Limits - SUICIDE
111.0
106.2
101.4
96.6
91.8
87.0
82.2
77.4
72.6
67.8
63.0
2012/1
2012/4
2012/7
2012/10
2013/1
2013/4
2013/7
2013/10
Periods 2012/1 to 2013/12(Seasonality of 12)
Upper Limit
Forecasts
Lower Limit

SUICIDE.ASC Autobox Demo Build: 6.0.21

File Preferences View Process Series Help

Historical Data | Auxiliaries | Graph | Reports | Interventions

Forecast Data

Period/Major,Minor	Lower 80%	SUICIDE	Upper 80%
205 2012/1	72.00	82.00	94.00
206 2012/2	63.00	72.00	83.00
207 2012/3	69.00	79.00	91.00
208 2012/4	64.00	74.00	85.00
209 2012/5	69.00	79.00	91.00
210 2012/6	67.00	78.00	89.00
211 2012/7	65.00	75.00	86.00
212 2012/8	67.00	77.00	88.00
213 2012/9	73.00	84.00	97.00
214 2012/10	70.00	81.00	93.00
215 2012/11	78.00	90.00	104.00
216 2012/12	73.00	84.00	97.00
217 2013/1	73.00	85.00	98.00
218 2013/2	73.00	84.00	97.00
219 2013/3	75.00	87.00	101.00
220 2013/4	67.00	77.00	89.00
221 2013/5	65.00	75.00	87.00
222 2013/6	68.00	78.00	91.00
223 2013/7	64.00	75.00	86.00
224 2013/8	63.00	72.00	84.00
225 2013/9	67.00	78.00	90.00
226 2013/10	73.00	84.00	97.00
227 2013/11	71.00	82.00	95.00
228 2013/12	83.00	96.00	111.00

with forecasts here

edited Apr 19 '15 at 13:04

answered Apr 18 '15 at 18:43

IrishStat
14.8k 1 16 31

Thanks for your answer. Are these assumptions so important and always so flouted in real world data that acf and pacf plots can almost never be interpreted in isolation? – rnso Apr 19 '15 at 2:06

I hate to say never BUT the assumptions that I laid out would severely complicate the visual identification process if violated. Your data set clearly (to my old eyes) is an example of this. Identifying an initial model , estimating and re-identifying based upon residual diagnostics is a multi-stage process not a one and done EXCEPT in trivial cases. – IrishStat Apr 19 '15 at 9:47

To reiterate following my friend stats.stackexchange.com/users/48766/javlacalle : Checking for the presence of pulses and level shifts AND seasonal pulses AND local time trends AND error variance constancy is also necessary. – IrishStat Apr 19 '15 at 10:25

(+1) Nice analysis of the data. However, what about the original question? can seasonality be identified in the data? Maybe it can be inferred from the output that you show, but I couldn't figure it out. – javlacalle Apr 19 '15 at 23:04

seasonality is present in the AR(12) term in the ARIMA model and in the seasonal pulse starting at period 98 (2003/2) – IrishStat Apr 20 '15 at 1:19

|

Interpretation of the ACF and PACF

The slow decay of the autocorrelation function suggests the data follow a long-memory process. The duration of shocks is relatively persistent and influence the data several observations ahead. This is probably reflected by a smooth trending pattern in the data.

The ACF and PACF of order 12 are beyond the significance confidence bands. However, this does not necessarily mean the presence of an identifiable seasonal pattern. The ACF and PACF of other seasonal orders (24, 36, 48, 60) are within the confidence bands. From the graphic, it is not possible to conclude whether the significance of the ACF and PACF of order 12 is due to seasonality or transitory fluctuations.

The persistence of the ACF mentioned before suggests that first differences may be needed to render the data stationary. However, the ACF/PACF of the differenced series look suspicious, negative correlation may have been induced by the differencing filter and may not be actually appropriate. See this post for some details.

Determine if seasonality is present

The analysis of the ACF and PACF should be complemented with other tools, for example:

- Spectrum (a view to the ACF in the frequency domain), may reveal the periodicity of cycles that explain most of the variability in the data.
- Fit the basic structural time series model and check if the variance of the seasonal component is close to zero relatively to the other parameters (in R function `stats::StructTS` and package stsm).

- Tests for seasonality, based on seasonal dummies, seasonal cycles or those described and implemented in X-12.
- Checking for the presence of pulses and level shifts as mentioned by IrishStat is also necessary since they can distort the conclusions from the previous methods (in R the package tsoutliers can be useful to this end).

edited Apr 19 '15 at 8:16

answered Apr 18 '15 at 18:02

javlacalle
6,558 9 27

I have added plot upto lag 60. What would be the R command for getting "differenced series"? I will add plots for diff(my_series). – rnso Apr 18 '15 at 18:16

@mso I have added major changes to my previous answer. The command for the differenced series is the function `diff` that you used. – javlacalle Apr 18 '15 at 19:53

javlacalle -- there were two very similar paragraphs and @rnso attempted to help by removing one. I have removed what I think is the one you wanted to replace. Could you please check that the correct paragraph was removed? – Glen_b ♦ Apr 19 '15 at 1:35

@Glen_b thanks for the editing, I've made some changes. – javlacalle Apr 19 '15 at 8:18

@ javlacalle, @IrishStat : please see the edit in my question re original data. – rnso Apr 19 '15 at 10:41