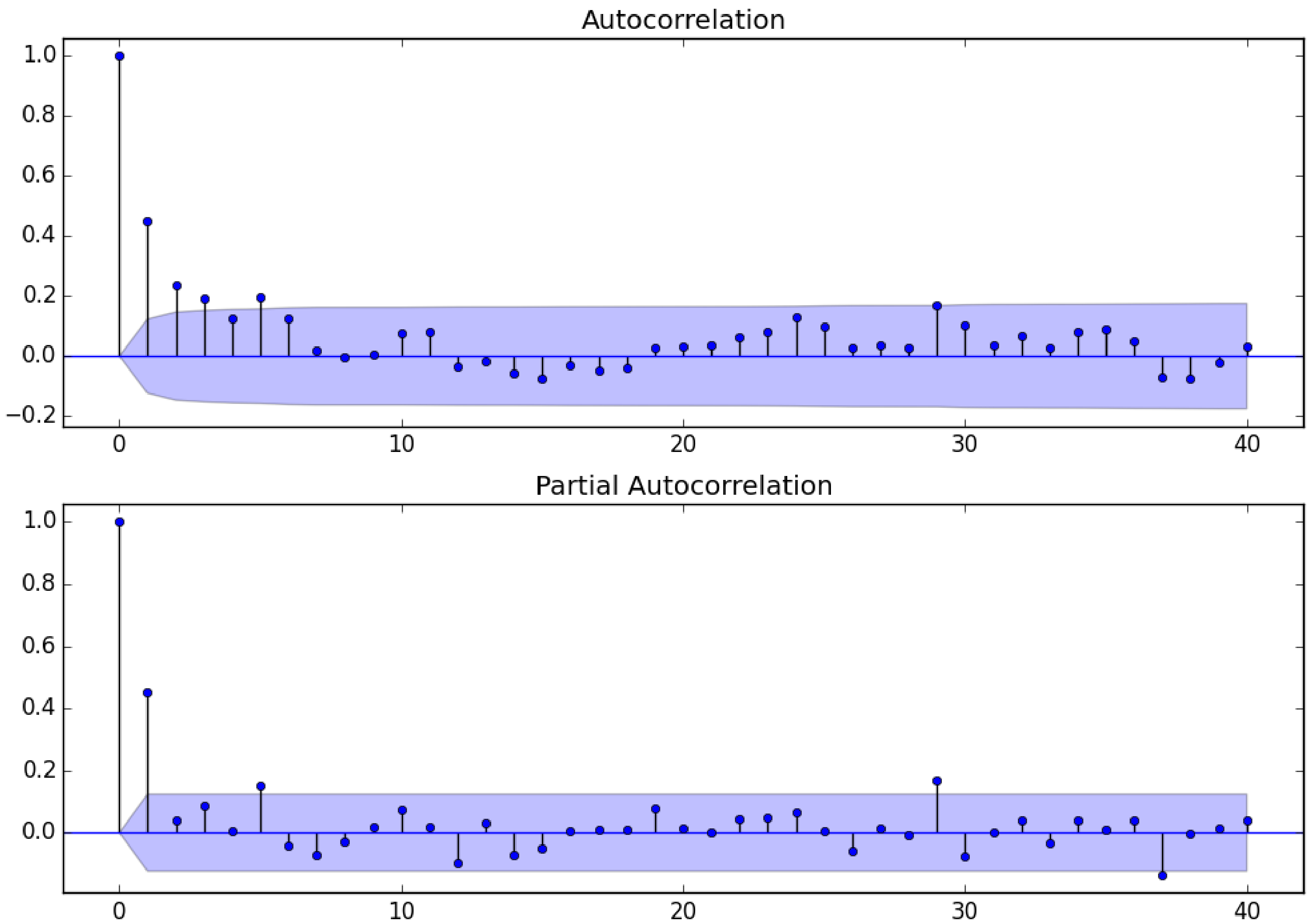
Analyse ACF and PACF plots

I want to see if I am on the right track analysing my ACF and PACF plots:



Background: (Reff: Philip Hans Franses, 1998)

1. As both ACF and PACF show significant values, I assume that an ARMA-model will serve my needs
2. The ACF can be used to estimate the MA-part, i.e q-value, the PACF can be used to estimate the AR-part, i.e. p-value

3. To estimate a model-order I look at a.) whether the ACF values die out sufficiently, b.) whether the ACF signals overdifferencing and c.) whether the ACF and PACF show any significant and easily interpretable peaks at certain lags
4. ACF and PACF might suggest not only one model but many from which I need to choose after considering other diagnostic tools

Having that in mind, I would go ahead and say that the most obvious model seems to be ARMA (4,2) as ACF values die out at lag 4 and PACF shows spikes at 1 and 2.

Another way to analyze would be an ARMA(2,1) as I see two significant spikes in my PACF and one significant spike in my ACF (after which the values die out starting from a much lower point (0.4)).

Looking at my in-sample-forecast results (using a simple Mean Absolute Percentage Error) ARMA (2,1) delivers much better results then ARMA(4,2). So I use ARMA(2,1)!

Can you confirm my method and findings of analyzing ACF and PACF plots?

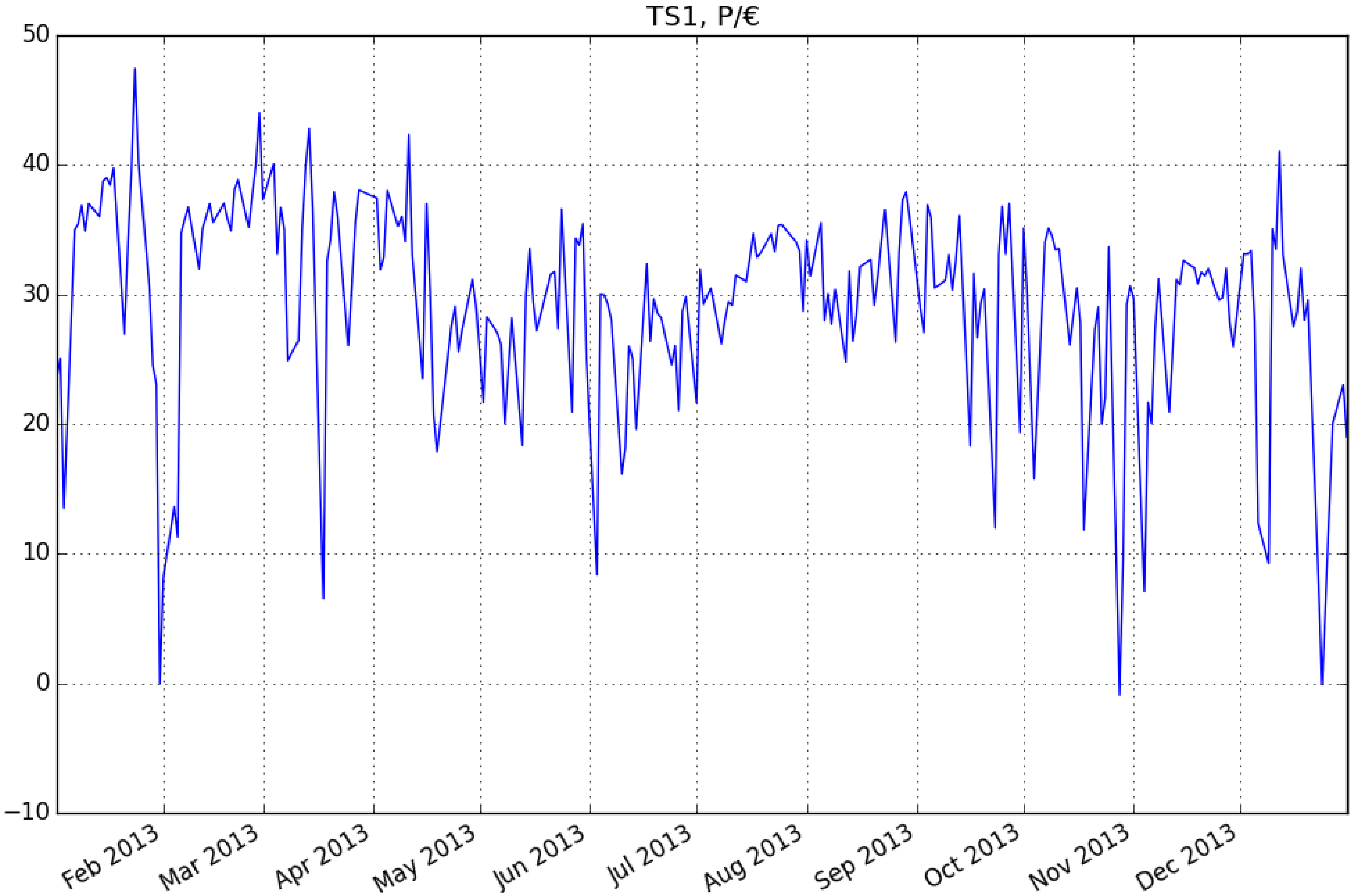
Help appreciated!

EDIT:

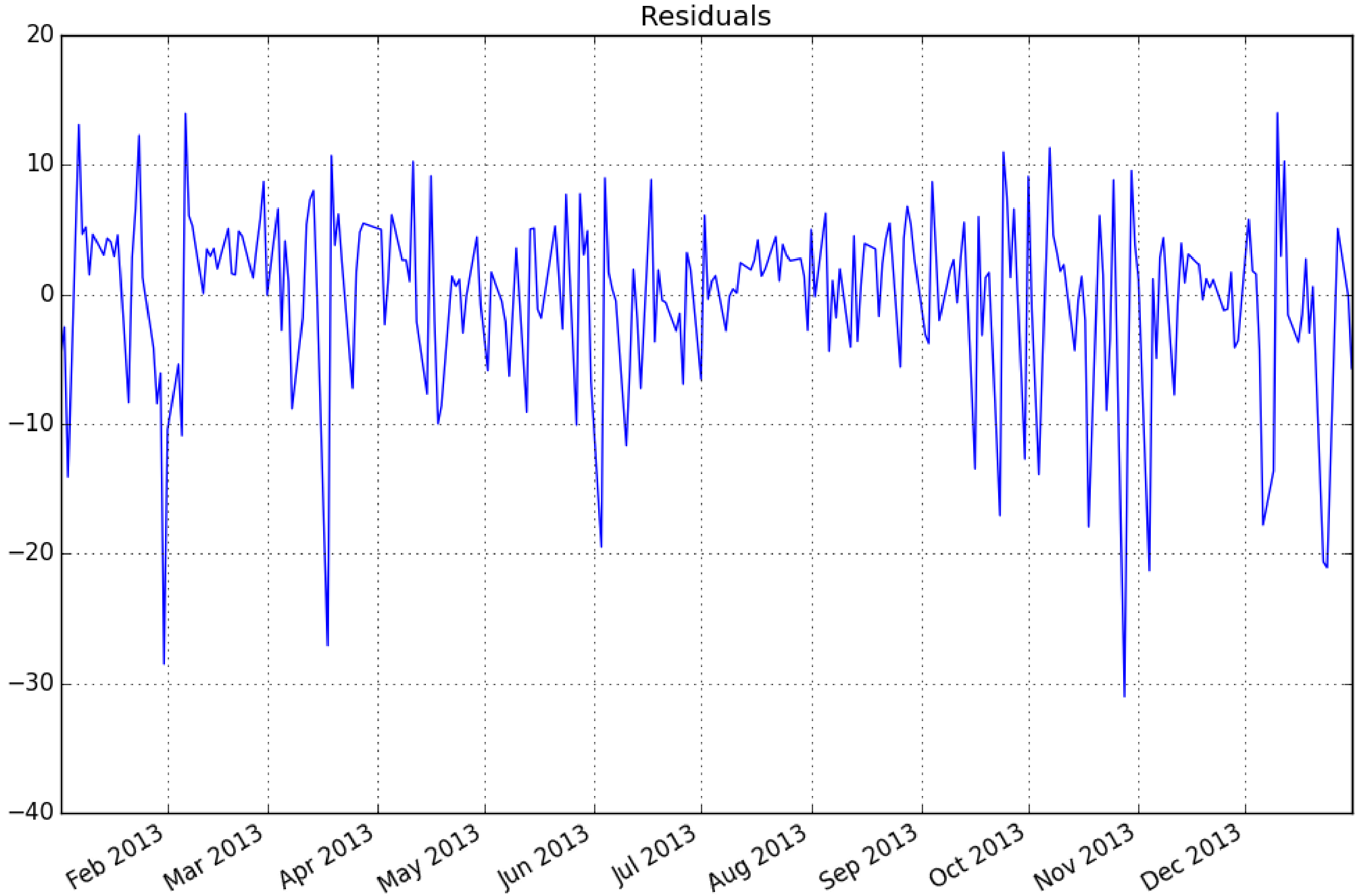
Descriptive Statistics:

```
count  252.000000
mean    29.576151
std      7.817171
min     -0.920000
25%     26.877500
50%     30.910000
75%     34.915000
max     47.430000

Skewness of endog_var: [-1.35798399]

Kurtsosis of endog_var: [ 5.4917757]

Augmented Dickey-Fuller Test for endog_var: (-3.76140904255411, 0.0033277703768345287,
{'5%': -2.8696473721448728, '1%': -3.4487489051519011, '10%': -2.5710891239349585}
```

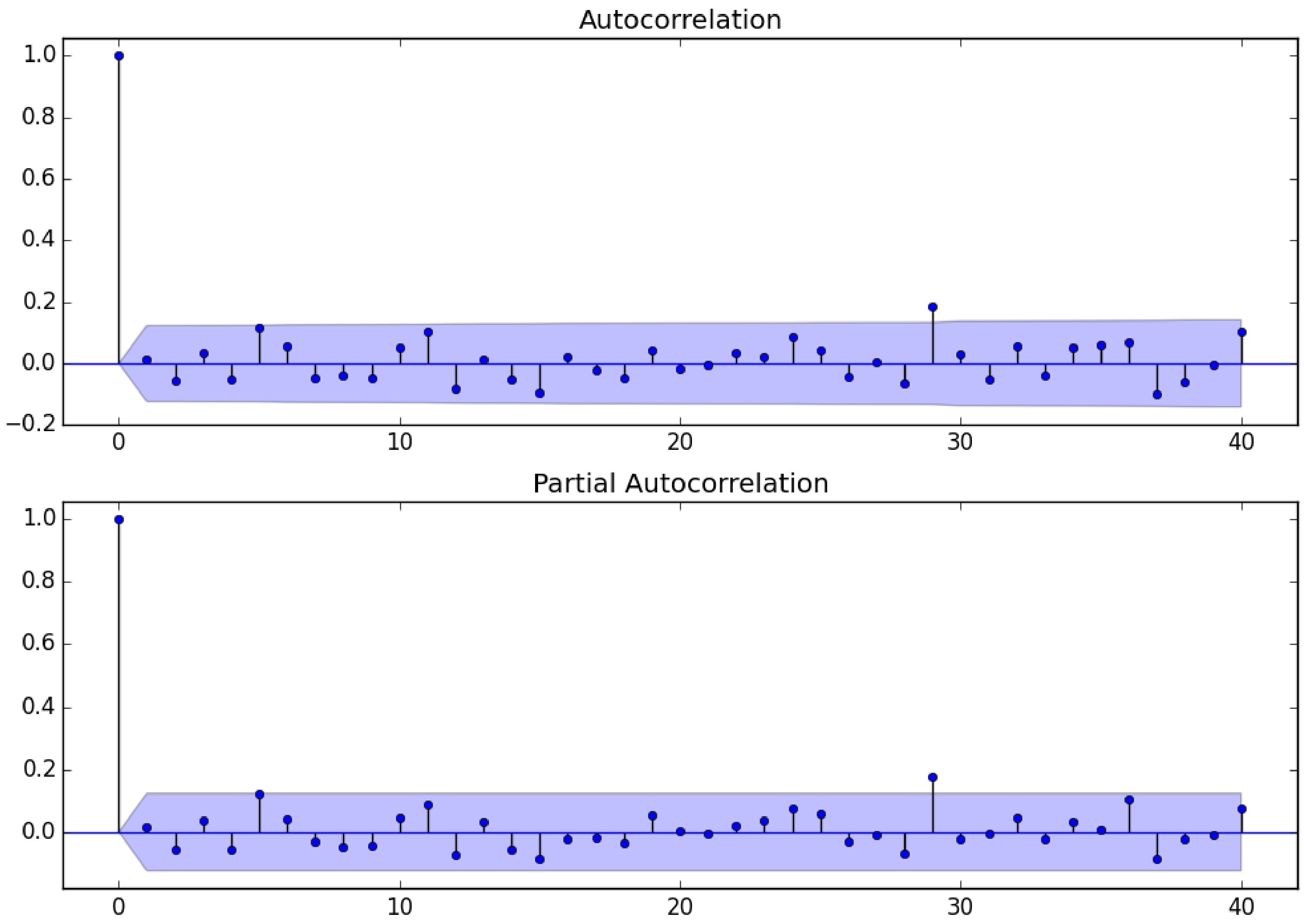
Time-Series:



Residuals (ARMA (2,1):



ACF/PACF of Residuals:


Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
−0.2
0
10
20
30
40
Partial Autocorrelation
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40

EDIT II:

Data:

```
14.37561
23.95561
25.41561
13.88561
23.31561
33.12561
35.30561
35.78561
37.21561
35.23561
37.34561
38.28561
39.03561
36.34561
39.08561
39.34561
38.80561
40.10561
34.13561
35.42561
27.29561
34.13561
39.89561
47.77561
40.57561
36.15561
33.66561
30.97561
24.90561
23.41561
0.31561
8.45561
37.36561
33.40561
13.97561
11.62561
35.07561
36.15561
37.09561
36.95561
37.85561
32.31561
35.41561
36.35561
37.34561
35.90561
37.40561
36.44561
37.37561
```

```
36.16561
35.24561
38.47561
39.18561
39.61561
29.55561
35.50561
38.05561
40.32561
44.39561
37.65561
46.27561
29.41561
40.41561
33.44561
37.04561
35.34561
25.24561
30.23561
15.40561
26.79561
35.38561
40.22561
43.14561
36.96561
41.93561
11.30561
6.87561
32.92561
34.54561
38.27561
36.40561
25.44561
37.26561
26.39561
31.13561
35.90561
38.41561
33.66561
33.16561
31.96561
30.34561
37.77561
32.25561
33.21561
38.37561
36.63561
40.78561
35.60561
```

```
36.37561
34.42561
42.67561
33.40561
31.49561
24.81561
23.82561
37.34561
30.73561
21.04561
18.20561
27.36561
18.49561
25.41561
27.92561
29.42561
25.91561
27.56561
28.69561
29.89561
31.47561
29.34561
25.35561
21.98561
28.61561
33.87561
20.07561
27.36561
26.48561
20.37561
22.33561
28.52561
21.24561
10.77561
18.69561
30.19561
33.89561
29.81561
27.55561
22.37561
20.32561
22.43561
31.89561
32.10561
27.67561
36.93561
36.51561
26.96561
21.27561
```

```
34.68561
34.13561
35.80561
25.38561
33.42561
9.28561
8.70561
30.36561
30.29561
29.56561
28.41561
33.40561
18.47561
16.48561
18.51561
26.35561
25.40561
19.92561
21.26561
10.90561
32.71561
26.71561
29.99561
28.87561
28.55561
14.07561
10.97561
24.92561
26.40561
21.40561
29.08561
30.18561
30.27561
16.15561
21.96561
32.29561
29.57561
30.24561
30.82561
28.83561
27.30561
26.53561
28.39561
29.76561
29.50561
31.81561
34.79561
24.14561
31.34561
```

```
33.14561
35.04561
33.20561
33.53561
35.28561
29.84561
35.02561
33.63561
35.65561
35.73561
35.35561
37.18561
27.38561
34.40561
33.69561
29.05561
34.55561
31.76561
30.91561
34.70561
35.87561
28.31561
30.39561
28.03561
30.72561
30.57561
23.93561
25.11561
32.15561
26.74561
28.76561
32.49561
34.79561
27.90561
33.05561
29.50561
31.67561
34.36561
36.88561
32.31561
26.24561
26.66561
33.59561
37.64561
38.26561
36.20561
33.27561
29.94561
29.19561
```

```
27.41561
37.24561
36.26561
30.84561
35.46561
32.24561
31.44561
33.40561
30.71561
33.03561
36.43561
33.44561
22.32561
18.65561
31.97561
27.00561
29.66561
30.76561
33.44561
29.19561
12.32561
33.41561
37.13561
33.43561
37.35561
40.17561
29.38561
19.70561
35.44561
30.48561
30.72561
16.09561
30.82561
30.55561
34.38561
35.45561
34.87561
33.78561
33.87561
29.83561
26.35561
26.44561
28.72561
30.85561
28.18561
12.18561
31.82561
18.01561
27.57561
```

```
29.38561
20.32561
22.36561
34.01561
34.40561
20.23561
-0.57439
9.87561
29.55561
31.01561
30.00561
28.12561
13.47561
7.42561
22.01561
20.38561
27.57561
31.54561
29.90561
16.40561
21.27561
26.22561
31.47561
31.11561
32.97561
32.34561
29.36561
32.40561
31.16561
32.05561
31.78561
32.34561
33.87561
31.80561
29.90561
30.09561
32.36561
28.15561
26.30561
15.32561
31.03561
33.47561
33.44561
33.71561
28.30561
12.70561
10.17561
43.96561
9.58561
```

```
35.38561
33.82561
41.37561
33.40561
33.64561
20.30561
27.85561
29.01561
32.36561
28.33561
29.90561
27.19561
0.39561
8.40561
0.24561
11.87561
29.15561
20.40561
0.42561
29.29561
23.39561
19.36561
```

time-series model-selection arma statsmodels

edited Jan 22 '15 at 22:57

asked Jan 22 '15 at 12:59

Peter Knutsen
82 1 1 5

Data look a bit left-skew, perhaps nonstationary. It looks to me like there's some potential issues with the residuals, perhaps even conditional heteroskedasticity. – Glen_b ♦ Jan 22 '15 at 16:13

In my opinion the skewness suggests anomalous values (pulses) which can only be confirmed by analysis of the original data. – IrishStat Jan 22 '15 at 21:05

3 Answers

Looking at your ACF and PACF is useful in the full context of your analysis as well. Your Ljung-Box Q-statistic; p-value; confidence interval, ACF and PACF should be viewed together. For instance the Q test here:

```
acf, ci, Q, pvalue = tsa.acf(res1.resid, nlags=4,confint=95,  qstat=True, unbiased=True)
```

Here - our Q test for autocorrelation is an overall gut check of our graphical interpretation.

Draft notes on Time Series analysis in Statsmodels: http://conference.scipy.org/proceedings/scipy2011/pdfs/statsmodels.pdf

answered Feb 20 '15 at 23:40

Andrew Owens
246 1 4

It looks to me like you're counting the spikes at lag 0.

Your PACF shows one reasonably large spike at lag 1, suggesting AR(1). This will of course induce a geometric-like decrease in the ACF (which, broadly speaking, you see). You seem to be trying to fit the same dependence twice - both as AR and MA.

I'd have just tried AR(1) on that to start with and seen if there was anything left worth worrying over.

edited Jan 22 '15 at 13:45

answered Jan 22 '15 at 13:34

Glen_b ♦
155k 20 257 530

Peter; my answer had a typo in it (I had AR(1) correct in the last para, but typed MA(1) in the second paragraph), which is fixed now. – Glen_b ♦ Jan 22 '15 at 13:48

Thanks for your answer. Counting from lag 0 is of course a cardinal mistake! I tried AR(1) and the result was not as good as ARMA(2,1)! – Peter Knutsen Jan 22 '15 at 14:26

It may well be the case that it's not as good - nevertheless, the AR(1) would be the place to start. What did the PACF of residuals look like, for example? What does the original series look like? There's much that might be going on that can't be gleaned easily from an ACF and PACF of the data. – Glen_b ♦ Jan 22 '15 at 14:39

Thanks. I posted some additional information which might lead to new insights. – Peter Knutsen Jan 22 '15 at 15:02

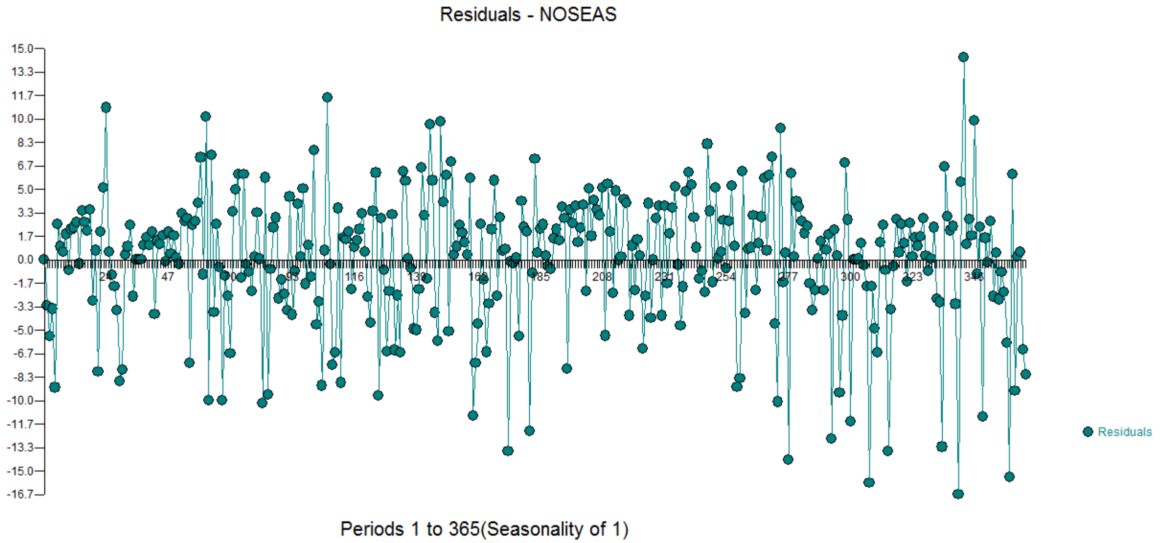
The sole reliance on the ACF and PACF using tools suggested in the mid 60's is sometimes but seldomly correct except for simulated data. Model Identification tools like AIC/BIC almost never correctly identify a useful model but rather show what happens when you don't read the small print regarding the assumptions. I would suggest that you start as simply as possible BUT not too simply and estimate a tentative model ; AR(1) as suggested by Glen_b . The residuals/analysis from this tentative model can be used to compute yet another ACF and PACF suggesting potential model augmentation or model simplification. Note that interpretation ala your references REQUIRE that the current series/residuals are free of any deterministic structure i.e. Pulses, Level Shifts, Local Time Trends and Seasonal Pulses and furthermore that the series has constant error variance and that the parameters of the tentative model are invariant over time. If you wish you can post your data and I will attempt to help you form a useful model.

EDIT AFTER DATA WAS REPORTED :

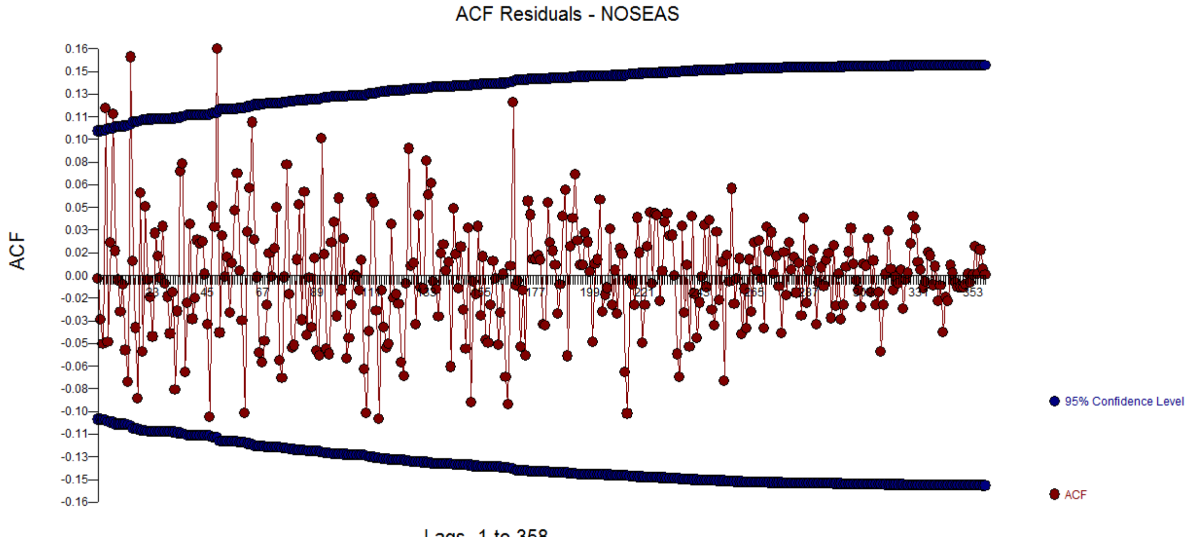
365 values were delivered and analyzed, yielding the following AR(1) model with identified Pulses and 2 Level Shifts .

```
      215-675-0652
   VERSION: 01/22/2015 06:40


MODELLING OUTPUT SERIES:NOSEAS


Y(T) =  35.148                              NOSEAS
       +[X1(T)][(- 27.6842)]              :PULSE          356
       +[X2(T)][(- 30.6971)]              :PULSE           31
       +[X3(T)][(- 26.8650)]              :PULSE          301
       +[X4(T)][(- 22.7138)]              :PULSE          358
       +[X5(T)][(- 26.1459)]              :PULSE          362
       +[X6(T)][(+ 2.9238)]               :LEVEL SHIFT    194
       +[X7(T)][(- 8.0312)]               :LEVEL SHIFT    105
       +[X8(T)][(- 28.2898)]              :PULSE           77
       +[X9(T)][(- 26.0516)]              :PULSE           32
       +[X10(T)[(- 23.2761)]              :PULSE           36
       +[X11(T)[(- 26.1168)]              :PULSE           76
       +[X12(T)[(- 14.3966)]              :PULSE          308
       +[X13(T)[(- 20.5078)]              :PULSE           35
       +[X14(T)[(- 20.4270)]              :PULSE          154
       +[X15(T)[(- 14.0550)]              :PULSE            4
       +[X16(T)[(- 20.6135)]              :PULSE          153
       +[X17(T)[(- 18.6130)]              :PULSE          357
       +[X18(T)[(- 15.3091)]              :PULSE           69
       +[X19(T)[(- 18.7496)]              :PULSE          302
       +[X20(T)[(- 11.5644)]              :PULSE          132
       +[X21(T)[(- 26.8954)]              :PULSE          343
       +[X22(T)[(- 16.1270)]              :PULSE          167
       +[X23(T)[(- 11.0503)]              :PULSE          174
       +[X24(T)[(- 18.7255)]              :PULSE          341
       +[X25(T)[(- 18.5606)]              :PULSE          266
       +[X26(T)[(- 17.8317)]              :PULSE          291
      +     [(1-  .383B** 1)]**-1  [A(T)]
```

. note that this had been a popular guess . The residuals from this model are plotted here



. There is a suggestion of variance hetero-scedasticity but this is a symptom and one needs to find the correct cure which we will ultimately find. Proceeding the acf of the residuals shown here



exhibits a suggestion of model inadequacy. A closer look at the table of the acf of the residuals is here

LAG	ACF VALUE	STND. ERROR	T- RATIO	CHI-SQUARE & PROBABILITY		PACF VALUE	STND. ERROR	T- RATIO
1	-.003	.052	-.05	.0	NA	-.003	.052	-.05
2	-.032	.052	-.61	.4	NA	-.032	.052	-.61
3	-.049	.052	-.93	1.3	NA	-.049	.052	-.94
4	.119	.053	2.27	6.5	NA	.118	.052	2.26
5	-.047	.053	-.89	7.3	NA	-.051	.052	-.98
6	.023	.053	.44	7.5	NA	.029	.052	.56
7	.115	.053	2.15	12.5	NA	.126	.052	2.40
8	.017	.054	.32	12.6	NA	-.002	.052	-.03
9	-.003	.054	-.06	12.6	NA	.019	.052	.36
10	-.026	.054	-.48	12.8	NA	-.021	.052	-.41
11	-.006	.054	-.12	12.9	NA	-.033	.052	-.63
12	-.053	.054	-.99	13.9	NA	-.047	.052	-.89
13	-.076	.054	-1.40	16.1	NA	-.091	.052	-1.73
14	.156	.055	2.85	25.4	NA	.149	.052	2.86

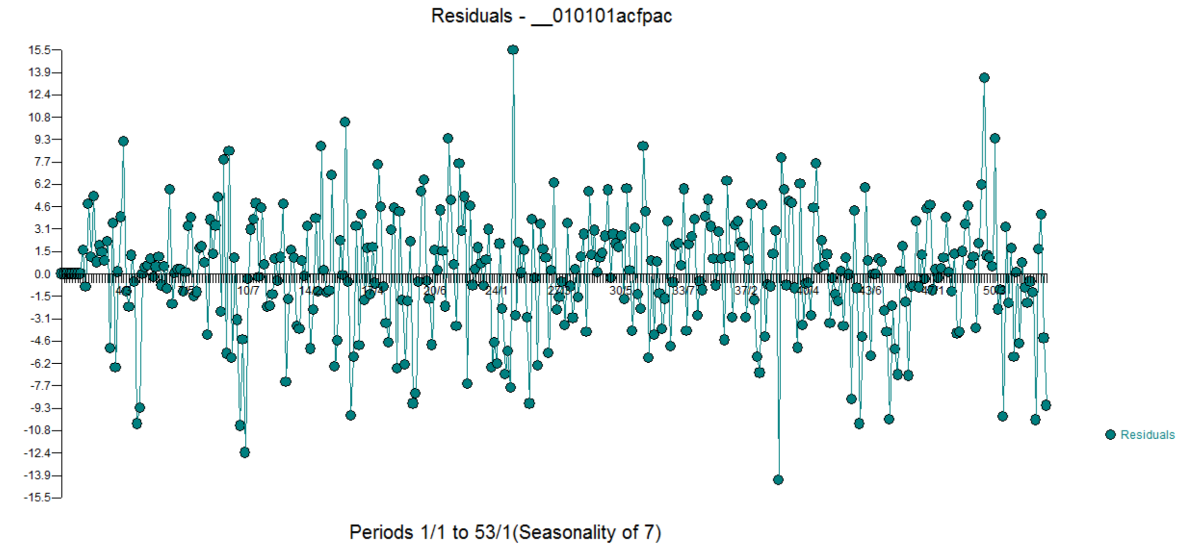
suggesting structure at lags 7 and 14. Putting the the two clues together (sample size of 365 and significant weekly i.e. lag 7 structure) I decided to investigate whether or not this was indeed daily data. New users often omit very important information when they define their data on the mistaken premise that the computer should be smart enough to figure everything out. Note that the lag 7 and lag 14 clues were swamped in the OP'S ACF and PACF plots. The presence of deterministic structure in the residuals increase the error variance thus suppressing the acf. Once the outliers/pulses/level shifts were identified the acf revealed the presence of an autoregressive structure /daily indicators which then needed to be accounted for.

I then analyzed the data allowing the software to proceed with the clue that it was daily data. With only 365 values it is not possible to properly construct models containing seasonal/holiday predictors BUT that is possible with more than 1 year of data.

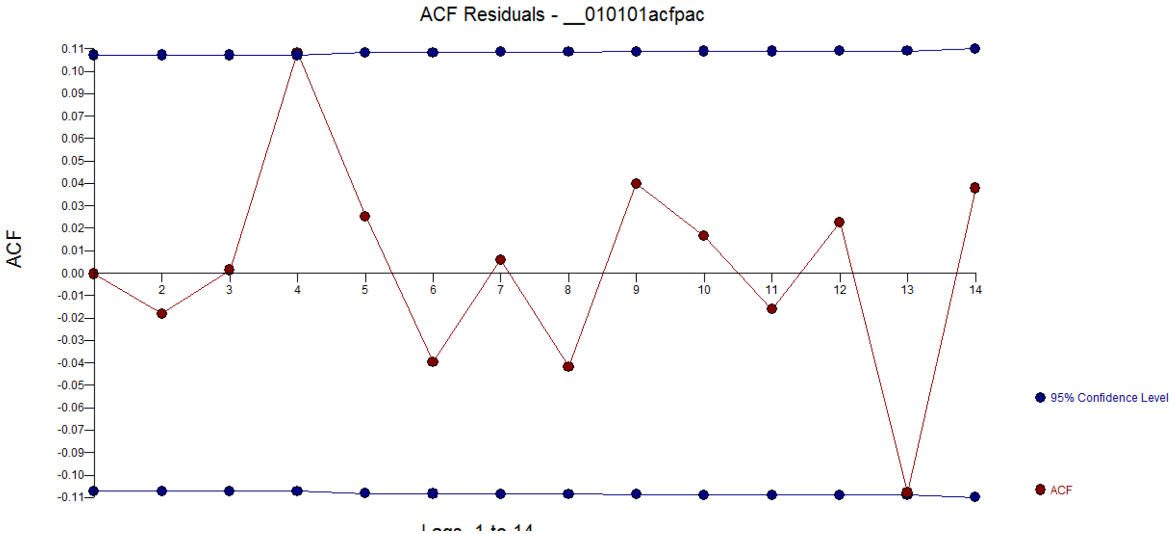
The model that was found is presented here

```
Y(T) =  32.169                                  __010101acfpac
       +[X1(T)][(+ 3.4875)]                                FIXED_EFF_N10107
       +[X2(T)][(+ 4.0895)]                                FIXED_EFF_N10207
       +[X3(T)][(+ 4.8971)]                                FIXED_EFF_N10307
       +[X4(T)][(+ 4.7409)]                                FIXED_EFF_N10407
       +[X5(T)][(+ 5.7827)]                                FIXED_EFF_N10507
       +[X6(T)][(- 33.2875)]                               :PULSE          01/31/01     5/  3    31
       +[X7(T)][(- 8.8435)]                                :LEVEL SHIFT    04/14/01    15/  6   104
       +[X8(T)][(+ 3.5890)]                                :LEVEL SHIFT    07/03/01    27/  2   184
       +[X9(T)][(- 28.3727)]                               :PULSE          12/28/01    52/  5   362
       +[X10(T)[(- 27.2964)]                               :PULSE          02/01/01     5/  4    32
       +[X11(T)[(- 25.5713)]                               :PULSE          10/28/01    43/  7   301
       +[X12(T)[(- 23.6382)]                               :PULSE          12/22/01    51/  6   356
       +[X13(T)[(- 22.4020)]                               :PULSE          02/05/01     6/  1    36
       +[X14(T)[(- 23.1492)]                               :PULSE          03/18/01    11/  7    77
       +[X15(T)[(- 27.1761)]                               :PULSE          12/24/01    52/  1   358
       +[X16(T)[(- 19.4441)]                               :PULSE          10/29/01    44/  1   302
       +[X17(T)[(- 28.9743)]                               :PULSE          12/07/01    49/  5   341
       +[X18(T)[(- 19.3526)]                               :PULSE          10/18/01    42/  4   291
       +[X19(T)[(- 11.8816)]                               :PULSE          11/04/01    44/  7   308
       +[X20(T)[(- 20.4205)]                               :PULSE          03/17/01    11/  6    76
       +[X21(T)[(- 11.3145)]                               :PULSE          06/22/01    25/  5   173
       +[X22(T)[(- 14.6556)]                               :PULSE          06/03/01    22/  7   154
       +[X23(T)[(- 18.1125)]                               :PULSE          02/04/01     5/  7    35
       +[X24(T)[(- 18.8571)]                               :PULSE          12/06/01    49/  4   340
       +[X25(T)[(- 15.1483)]                               :PULSE          06/02/01    22/  6   153
       +[X26(T)[(- 17.0799)]                               :PULSE          12/23/01    51/  7   357
       +[X27(T)[(- 15.4893)]                               :PULSE          10/04/01    40/  4   277
       +[X28(T)[(- 16.8814)]                               :PULSE          11/30/01    48/  5   334
       +[X29(T)[(- 26.5779)]                               :PULSE          12/09/01    49/  7   343
       +[X30(T)[(- 17.7628)]                               :PULSE          12/25/01    52/  2   359
       +[X31(T)[(- 14.2254)]                               :PULSE          03/23/01    12/  5    82
      +     [(1-  .411B** 1)(1-  .200B** 7)]**-1  [A(T)]
```

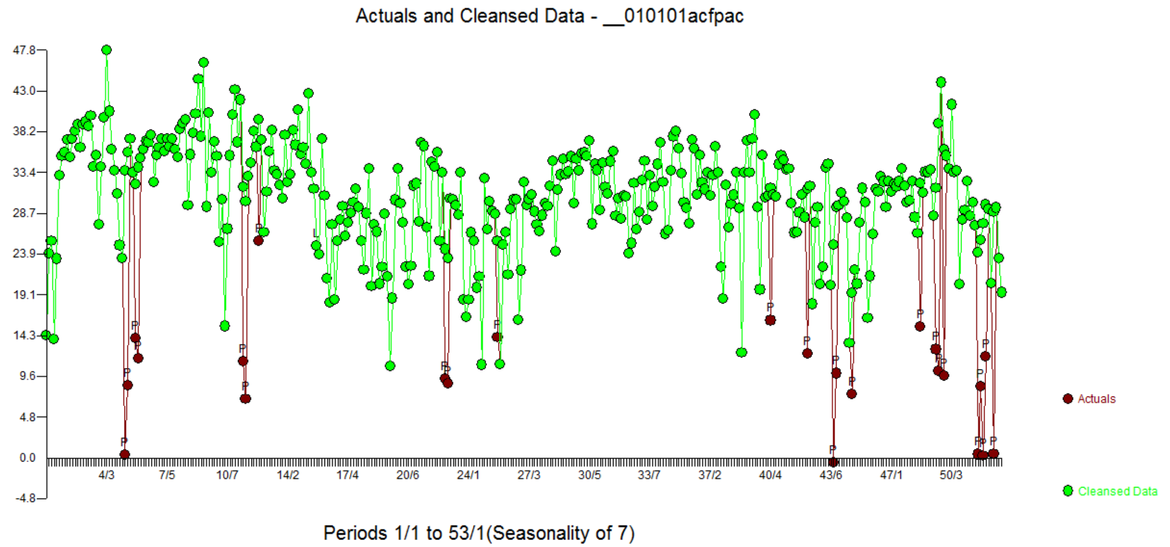
containing 5 daily dummies , two Level Shifts , a number of pulses and an arima model of the form (1,0,0)(1,0,0) . The plot of the residuals no longer evidences the non-constancy structure as a better model is in place.



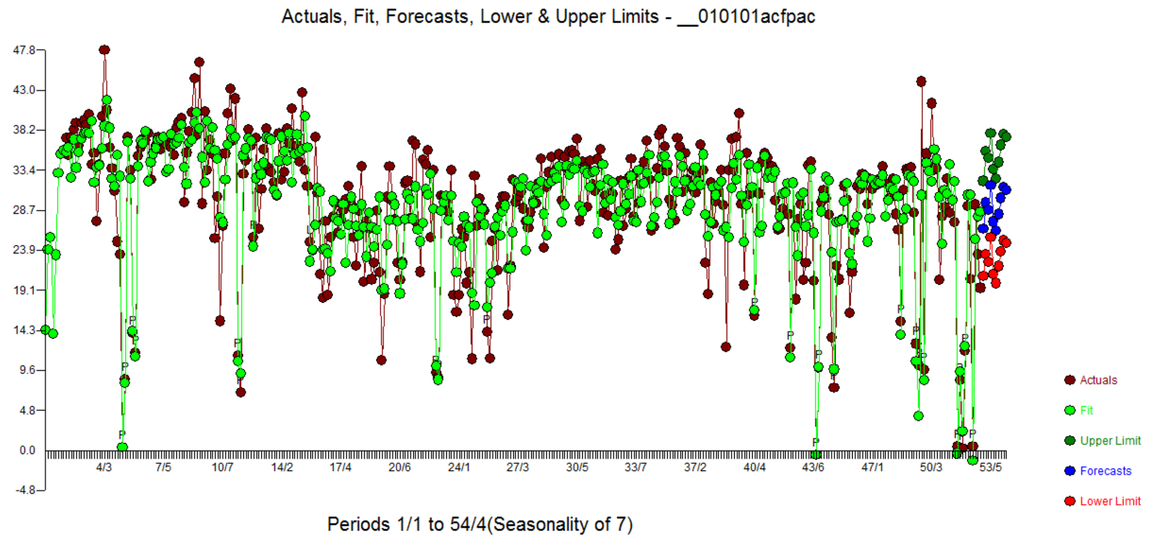
. Th



e acf of the residuals is much cleaner . The Actual /Cleansed graph highlights the unusual pulse points.



. THe lesson here is that when one analyzed the data without the critical piece of information that it was a daily time series there were a ton of pulses reflecting an inadequate representation (or perhaps the advanced knowledge of the daily clue) . The Actual/Fit and Forecast is presented here



.

It would be interesting to see what others would do with the same data set. Note that all analyses were conducted in a hands-free mode using software that is commercially available.

edited Jan 24 '15 at 16:53

answered Jan 22 '15 at 13:43

IrishStat
14.8k 1 16 31

1 early morning mis read ... Don't normally see the lag(0) in my graphs – IrishStat Jan 22 '15 at 13:45

1 It tricked me at first as well. – Glen_b ♦ Jan 22 '15 at 13:48

Thanks for your answer. As someone without experience in the field of time-series forecasting it is hard to fully understand the procedure of choosing the right model as there is no officially right way to go. Unfortunately i am not allowed to post my raw data. I hope that the additional information is useful (see 'EDIT:') – Peter Knutsen Jan 22 '15 at 14:57

You can scale/mask your data before you present it. Looking at the plot it appears there might be some unusual values which if untreated downwards biases the acf and the pacf incorrectly suggesting sufficiency. There is a visual suggestion of a downwards trend followed by no trend but that is just a guess at this moment. – IrishStat Jan 22 '15 at 17:52

i just added some data which you might use.. – Peter Knutsen Jan 22 '15 at 22:58