Business and Economics Forecasting
Econ 3342

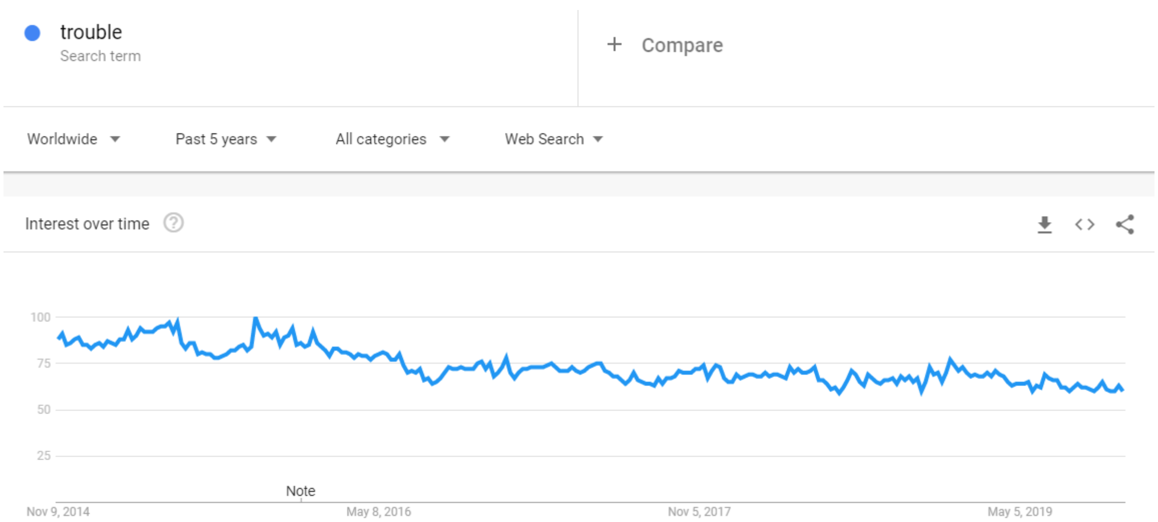
Spring, 2019

Diego Escobari

Assignment 3 – Suggested Solutions

- Due Thursday November 14 (before the beginning of the class).
- You can work in groups of up to three students.
- Send your PDF responses by email and make sure you copy all members when submitting your PDF file.
- Make sure your PDF file shows your work on EViews.

Go to Google Trends (http://www.google.com/trends) and type any keyword you think is interesting. Google Trends will give you the time series data of an index that captures how much people search for that specific term or keyword.[1] You can search, for example, for "trouble" to get:

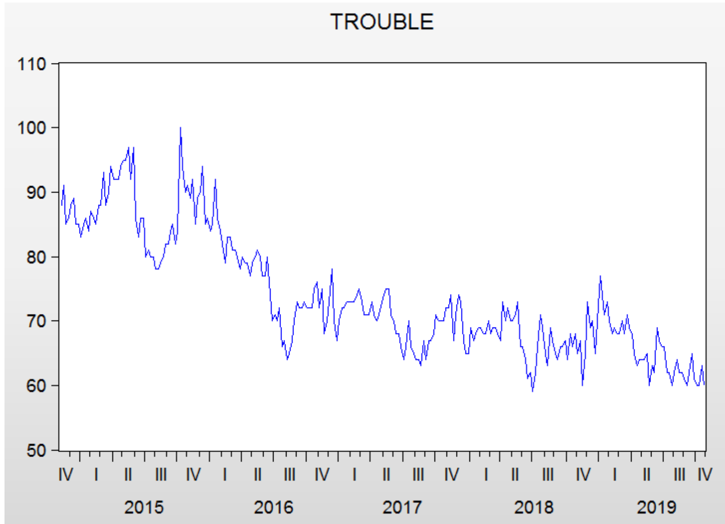


In this case, I selected "Worldwide" and "Past 5 years." You can try other keywords, such as "love", "presents", "travel", "vacation", etc. Use your imagination.

1. Select your keyword and download the data. I recommend "Past 5 years", but you can select a smaller sample if your EViews has restrictions. Export your data to EViews and obtain a time series graph.

Response: The time series graph of the series "trouble" is presented below.

[1] This source of data has been used by a professor in UC Berkeley to forecast unemployment.



2. Obtain the correlogram of the series. What can you say from the different autocorrelations and partial autocorrelations?

Response: The correlogram is:

Sample: 11/09/2014 10/27/2019
Included observations: 260

	AC	PAC	Q-Stat	Prob
1	0.936	0.936	230.36	0.000
2	0.899	0.188	443.82	0.000
3	0.875	0.136	646.98	0.000
4	0.862	0.135	844.78	0.000
5	0.841	-0.003	1033.5	0.000
6	0.811	-0.068	1209.7	0.000
7	0.782	-0.031	1374.6	0.000
8	0.762	0.020	1531.5	0.000
9	0.741	-0.004	1680.5	0.000
10	0.717	-0.017	1820.4	0.000
11	0.703	0.083	1955.5	0.000
12	0.689	0.037	2086.1	0.000
13	0.671	-0.027	2210.2	0.000
14	0.660	0.068	2330.9	0.000
15	0.651	0.036	2448.8	0.000
16	0.633	-0.072	2560.8	0.000
17	0.619	0.006	2668.3	0.000
18	0.608	0.026	2772.4	0.000
19	0.621	0.204	2881.4	0.000
20	0.620	0.003	2990.5	0.000
21	0.609	-0.035	3096.3	0.000
22	0.599	-0.006	3198.8	0.000
23	0.590	-0.057	3298.9	0.000
24	0.580	-0.058	3396.1	0.000

The autocorrelation function shows positive and statistically significant correlations are various displacements. This is evidence of strong persistence in the series, that is likely to be captured with an AR process. The partial autocorrelation is also positive statistically significant for the first displacement.

3. Is your series White Noise? Explain.

Response: From the last column in the correlogram we can see the Ljung-Pierce Q-statistic. The p-values at different displacements are all less than 0.05. Hence, we reject the null hypothesis of White Noise. For our forecasting purposes this means that there is something in the dynamics of "trouble" that can be forecasted.

4. Estimate the most appropriate moving average model. What is the interpretation of the coefficient on the first lagged error term?

We need to estimate the different MA(q) models for various values of q. Then record the different Akaike information criteria (AIC) and Schwarz Bayesian information criteria (BIC), to then select the one with the smallest. A summary of the results in presented in the tables below:

AIC	MA(0)	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)	MA(7)	MA(8)
AR(0)	5.6376	5.2467	5.1025	5.0774	5.0663	5.0298	**4.9876**	4.9880	4.9878

BIC	MA(0)	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)	MA(7)	MA(8)
AR(0)	5.7061	5.3426	5.2120	5.2006	5.2033	5.1804	**5.1519**	5.1660	5.1795

Both selection criteria choose an MA(6):

Dependent Variable: TROUBLE
Method: ARMA Maximum Likelihood (BFGS)
Sample: 11/09/2014 10/27/2019
Included observations: 260
Convergence achieved after 12 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	86.83178	3.202246	27.11590	0.0000
@TREND	0.214361	0.161077	1.330799	0.1845
@TREND^2	-0.007146	0.002469	-2.894126	0.0041
@TREND^3	4.58E-05	1.42E-05	3.232261	0.0014
@TREND^4	-8.88E-08	2.70E-08	-3.292433	0.0011
MA(1)	0.573136	0.063735	8.992422	0.0000
MA(2)	0.382087	0.069881	5.467705	0.0000
MA(3)	0.203092	0.068955	2.945270	0.0035
MA(4)	0.262919	0.069828	3.765233	0.0002
MA(5)	0.363826	0.079281	4.589088	0.0000
MA(6)	0.251433	0.063129	3.982880	0.0001
SIGMASQ	7.795382	0.529201	14.73047	0.0000

R-squared	0.910848	Mean dependent var	73.76923
Adjusted R-squared	0.906893	S.D. dependent var	9.368920
S.E. of regression	2.858772	Akaike info criterion	4.987587
Sum squared resid	2026.799	Schwarz criterion	5.151926
Log likelihood	-636.3863	Hannan-Quinn criter.	5.053653
F-statistic	230.3419	Durbin-Watson stat	1.950394
Prob(F-statistic)	0.000000		

The coefficient on the first lagged error term says that *a shock* in TROUBLE last period increases how people search for the keyword TROUBLE on Google this period by 0.57. The effects is statistically significant.

5. Estimate the most appropriate autoregressive model. What is the interpretation of the coefficient on the first lagged autoregressive term?

We estimate all different AR models, from AR(0) to AR(8). The results for the AIC and BIC are as follows:

AIC	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)	AR(7)	AR(8)
MA(0)	5.6376	4.9992	4.994	4.9961	4.9898	4.9971	4.986	**4.981**	4.987

BIC	AR(0)	AR(1)	AR(2)	AR(3)	AR(4)	AR(5)	AR(6)	AR(7)	AR(8)
MA(0)	5.7061	**5.0951**	5.1036	5.1194	5.1268	5.1477	5.1504	5.1591	5.1788

The AIC selects model AR(7), while the BIC selects the AR(1). Following the parsimony principle, we go with the AR(1). The regression output is:

Dependent Variable: TROUBLE
Method: ARMA Maximum Likelihood (BFGS)
Sample: 11/09/2014 10/27/2019
Included observations: 260
Convergence achieved after 3 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	86.93049	3.474511	25.01949	0.0000
@TREND	0.208057	0.174246	1.194045	0.2336
@TREND^2	-0.007040	0.002677	-2.629517	0.0091
@TREND^3	4.51E-05	1.53E-05	2.943991	0.0035
@TREND^4	-8.75E-08	2.91E-08	-3.006625	0.0029
AR(1)	0.691352	0.044983	15.36906	0.0000
SIGMASQ	8.207009	0.518007	15.84343	0.0000

R-squared	0.906140	Mean dependent var	73.76923
Adjusted R-squared	0.903914	S.D. dependent var	9.368920
S.E. of regression	2.904149	Akaike info criterion	4.999212
Sum squared resid	2133.822	Schwarz criterion	5.095076
Log likelihood	-642.8975	Hannan-Quinn criter.	5.037751
F-statistic	407.0851	Durbin-Watson stat	2.149235
Prob(F-statistic)	0.000000		

The interpretation of the autoregressive coefficient is as follows. A one point increase in TROUBLE searches last periods increase the current TROUBLE value by about 0.69 points. The effect is statistically significant.

6. Estimate the most appropriate autoregressive moving average model.

We estimate all possible combinations of ARMA models. The results are:

AIC	MA(0)	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)	MA(7)	MA(8)
AR(0)	5.6376	5.2467	5.1025	5.0774	5.0663	5.0298	4.9876	4.9880	4.9878
AR(1)	4.9992	4.9900	4.9931	5.0007	4.9888	4.9738	4.9773	4.9838	4.9889
AR(2)	4.9940	4.9940	5.0008	4.9685	4.9784	4.9784	4.9836	4.9493	4.9838
AR(3)	4.9961	4.9996	4.9618	4.9748	4.9802	4.9861	4.9811	4.9575	4.9887
AR(4)	4.9898	4.9974	4.9940	5.0008	4.9325	4.9628	4.9418	4.9426	4.9570
AR(5)	4.9971	4.9366	4.9422	4.9888	4.9733	4.9470	4.9778	4.9512	4.9557
AR(6)	4.9860	4.9407	4.9471	4.9426	4.9434	4.9781	4.9487	4.9562	4.9541
AR(7)	4.9810	4.9446	4.9550	4.9352	4.9592	4.9484	4.9554	4.9625	4.9670
AR(8)	4.9870	4.9496	4.9599	4.9425	4.9486	4.9563	4.9606	4.9605	4.9728

BIC	MA(0)	MA(1)	MA(2)	MA(3)	MA(4)	MA(5)	MA(6)	MA(7)	MA(8)
AR(0)	5.7061	5.3426	5.2120	5.2006	5.2033	5.1804	5.1519	5.1660	5.1795
AR(1)	5.0951	5.0996	5.1163	5.1376	5.1395	5.1382	5.1554	5.1756	5.1944
AR(2)	5.1036	5.1173	5.1377	5.1191	5.1428	5.1564	5.1753	5.1548	5.2030
AR(3)	5.1194	5.1365	5.1125	5.1392	5.1582	5.1778	5.1865	5.1767	5.2215
AR(4)	5.1268	5.1480	5.1583	5.1788	5.1242	5.1682	5.1609	5.1754	5.2035
AR(5)	5.1477	5.1009	5.1203	5.1806	5.1788	5.1661	5.2106	5.1977	5.2159
AR(6)	5.1504	5.1187	5.1388	5.1481	5.1625	5.2109	5.1952	5.2164	5.2280
AR(7)	5.1591	5.1363	5.1605	5.1543	5.1920	5.1949	5.2156	5.2364	5.2546
AR(8)	5.1788	5.1551	5.1790	5.1754	5.1951	5.2165	5.2345	5.2481	5.2741

The results show that the AIC would select an ARMA(4,4), while the BIC selects an ARMA(1,0). The selection of the parsimonious model hence goes with the ARMA(1,0) = AR(1), the same that we selected on question 5 above.

7. Are the residuals from your ARMA model White Noise? What would that mean?

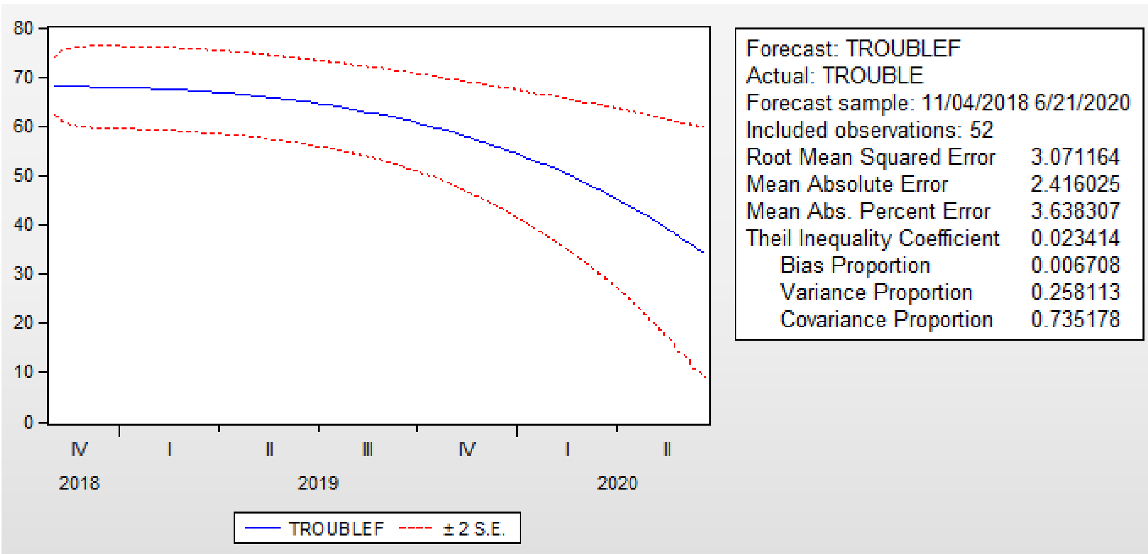
The correlogram of the residuals of the ARMA(1,0) = AR(1) is:

Sample: 11/09/2014 10/27/2019
Included observations: 260

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.075	-0.075	1.4912	0.222
		2	0.014	0.008	1.5427	0.462
		3	-0.000	0.001	1.5427	0.672
		4	0.132	0.133	6.1934	0.185
		5	0.140	0.163	11.420	0.044
		6	0.029	0.055	11.646	0.070
		7	-0.046	-0.044	12.213	0.094
		8	0.022	-0.009	12.347	0.136
		9	0.010	-0.032	12.376	0.193
		10	-0.103	-0.147	15.257	0.123
		11	-0.017	-0.046	15.341	0.167
		12	-0.018	-0.013	15.432	0.219

The residuals are White Noise because when looking at the Q-statistics we cannot reject the null hypothesis of White Noise at none of the displacements. That is, all p-values are above 0.05. This means that the AR(1) model is appropriate for forecasting as there is nothing left if the error term that can be predicted.

8. Provide an out of sample forecast and interpret it.



We extended the workfile to add more observations and then we obtained the forecast presented in the figure above. Our forecast goes up intil 6/21/2020. The forecast shows that TRUBLE will be decreasing in the following six months.

Note that you need to control for trend and seasonality if your series has them.

Response: We did not find any monthly seasonal component (i.e., people are not more or less likely to search for trouble in any particular month). When looking for the most appropriate trend, we selected a fourth power trend model as this is the one with the smallest BIC. The AIC selected the eight-power trend model but following the parsimony principle we go with the model selected by the BIC.

Model	AIC	BIC
No Trend	7.3165	7.3302
Linear Trend	6.0370	6.0644
Quadratic Trend	5.8406	5.8817
Cubic Trend	5.8482	5.9030
Fourth Power	5.6376	**5.7061**
Firth Power	5.6317	5.7139
Sixth Power	5.6383	5.7341
Seventh Power	5.6300	5.7395
Eight Power	**5.6184**	5.7416

The selected model is:

Dependent Variable: TROUBLE
Method: Least Squares
Sample: 11/09/2014 10/27/2019
Included observations: 260

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	86.63044	1.217184	71.17284	0.0000
@TREND	0.224641	0.065385	3.435641	0.0007
@TREND^2	-0.007296	0.001029	-7.086945	0.0000
@TREND^3	4.66E-05	5.98E-06	7.797313	0.0000
@TREND^4	-9.03E-08	1.14E-08	-7.886525	0.0000

R-squared	0.819075	Mean dependent var	73.76923
Adjusted R-squared	0.816237	S.D. dependent var	9.368920
S.E. of regression	4.016227	Akaike info criterion	5.637607
Sum squared resid	4113.171	Schwarz criterion	5.706081
Log likelihood	-727.8888	Hannan-Quinn criter.	5.665134
F-statistic	288.6064	Durbin-Watson stat	0.612419
Prob(F-statistic)	0.000000		