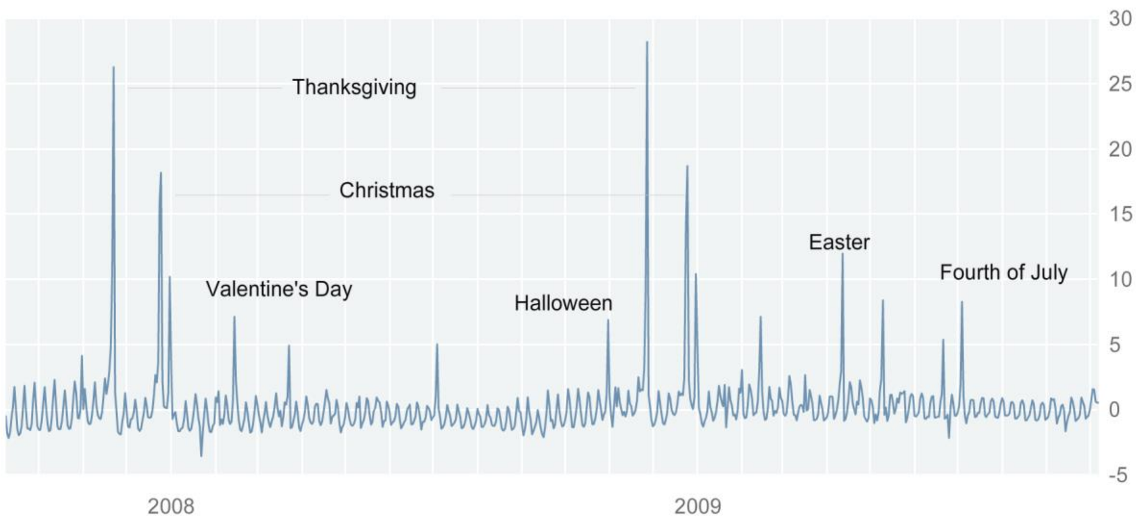
Business and Economics Forecasting
Econ 3342

Fall, 2019
Diego Escobari

Assignment 2 – Suggested Solutions

- Due Tuesday October 8 (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.

Is it a day to be happy? Facebook data team (www.facebook.com/data) decided to construct an index of happiness, and they call it The Facebook Global Happiness Index (FGHI). It is based on how many times Facebook users use words to covey joy (like "happy," "yay" and "awesome") and unhappiness ("sad," "doubt" and "tragic") in their profile updates. A time series graph of this index from September 9, 2007 to October 7, 2009 is presented below:[1]



From this graph we can eyeball that people appear to be happier during Thanksgiving, Christmas, Valentine's Day, Halloween, Easter and during the Fourth of July. Can we say these differences are statistically significant? Are there any recognizable seasonal patterns of happiness over the months of the year or over the days of the week? Is this series characterized by a linear trend?

The data set contains the variable 'happy' that is the FGHI over these 760 days. In addition, it has the following variables:

[1] The data comes from the article "Is It a Day to Be Happy? Check the Index," by Noam Cohen. October 12, 2009. The New York Times.

happy:	Facebook Global Happiness Index. For simplicity, let this index be measured in units of happiness.
mon:	dummy variable equal to 1 if Monday, zero otherwise.
tue:	dummy variable equal to 1 if Tuesday, zero otherwise.
wed:	dummy variable equal to 1 if Wednesday, zero otherwise.
thu:	dummy variable equal to 1 if Thursday, zero otherwise.
fri:	dummy variable equal to 1 if Friday, zero otherwise.
sat:	dummy variable equal to 1 if Saturday, zero otherwise.
sun:	dummy variable equal to 1 if Sunday, zero otherwise.
easter:	dummy variable equal to 1 if Easter, zero otherwise.
newyear:	dummy variable equal to 1 if New Year, zero otherwise.
thanks:	dummy variable equal to 1 if Thanksgiving, zero otherwise.
hallo:	dummy variable equal to 1 if Halloween, zero otherwise.
vale:	dummy variable equal to 1 if Valentine's Day, zero otherwise.
chris:	dummy variable equal to 1 if Christmas, zero otherwise.

a) Is there any weekly pattern or monthly pattern that you can identify as seasonal just by looking at the time series graph?

The time series graph shows a pronounced weekly pattern.

b) Do you think the data follows a trend? Estimate a model of 'happy' as a function of a constant and a trend. Use the AIC and the SIC to decide whether you model has no trend, a linear, quadratic, or cubic trend (maybe even to the fourth power). Explain.

No trend

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.260658	0.095175	2.738734	0.0063

R-squared	0.000000	Mean dependent var	0.260658
Adjusted R-squared	0.000000	S.D. dependent var	2.623782
S.E. of regression	2.623782	Akaike info criterion	4.768426
Sum squared resid	5225.134	Schwarz criterion	4.774522
Log likelihood	-1811.002	Hannan-Quinn criter.	4.770774
Durbin-Watson stat	1.189179		

Linear trend

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.015355	0.189934	-0.080846	0.9356
@TREND	0.000727	0.000433	1.678569	0.0936

R-squared	0.003703	Mean dependent var	0.260658
Adjusted R-squared	0.002389	S.D. dependent var	2.623782
S.E. of regression	2.620646	Akaike info criterion	4.767347
Sum squared resid	5205.783	Schwarz criterion	4.779540
Log likelihood	-1809.592	Hannan-Quinn criter.	4.772042
F-statistic	2.817595	Durbin-Watson stat	1.193599
Prob(F-statistic)	0.093648		

Quadratic trend

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.026972	0.284614	0.094769	0.9245
@TREND	0.000392	0.001732	0.226461	0.8209
@TREND^2	4.41E-07	2.21E-06	0.199793	0.8417

R-squared	0.003756	Mean dependent var	0.260658
Adjusted R-squared	0.001124	S.D. dependent var	2.623782
S.E. of regression	2.622308	Akaike info criterion	4.769926
Sum squared resid	5205.509	Schwarz criterion	4.788215
Log likelihood	-1809.572	Hannan-Quinn criter.	4.776969
F-statistic	1.426972	Durbin-Watson stat	1.193662
Prob(F-statistic)	0.240680		

Cubic trend

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.812672	0.376379	2.159185	0.0311
@TREND	-0.012071	0.004297	-2.808923	0.0051
@TREND^2	4.15E-05	1.32E-05	3.154572	0.0017
@TREND^3	-3.61E-08	1.14E-08	-3.165420	0.0016

R-squared	0.016787	Mean dependent var	0.260658
Adjusted R-squared	0.012886	S.D. dependent var	2.623782
S.E. of regression	2.606823	Akaike info criterion	4.759391
Sum squared resid	5137.418	Schwarz criterion	4.783777
Log likelihood	-1804.569	Hannan-Quinn criter.	4.768781
F-statistic	4.302611	Durbin-Watson stat	1.209495
Prob(F-statistic)	0.005072		

Fourth power

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.099212	0.467379	0.212273	0.8320
@TREND	0.006841	0.008544	0.800653	0.4236
@TREND^2	-7.08E-05	4.58E-05	-1.544613	0.1229
@TREND^3	1.94E-07	9.07E-08	2.139920	0.0327
@TREND^4	-1.52E-10	5.93E-11	-2.557667	0.0107

R-squared	0.025233	Mean dependent var	0.260658
Adjusted R-squared	0.020069	S.D. dependent var	2.623782
S.E. of regression	2.597321	Akaike info criterion	4.753395
Sum squared resid	5093.288	Schwarz criterion	4.783878
Log likelihood	-1801.290	Hannan-Quinn criter.	4.765134
F-statistic	4.886028	Durbin-Watson stat	1.219967
Prob(F-statistic)	0.000680		

Fifth power

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.194747	0.552567	-2.162177	0.0309
@TREND	0.058461	0.014745	3.964790	0.0001
@TREND^2	-0.000548	0.000121	-4.545479	0.0000
@TREND^3	1.87E-06	4.03E-07	4.646057	0.0000
@TREND^4	-2.64E-09	5.86E-10	-4.509047	0.0000
@TREND^5	1.31E-12	3.07E-13	4.271510	0.0000

R-squared	0.048264	Mean dependent var	0.260658
Adjusted R-squared	0.041953	S.D. dependent var	2.623782
S.E. of regression	2.568156	Akaike info criterion	4.732116
Sum squared resid	4972.949	Schwarz criterion	4.768695
Log likelihood	-1792.204	Hannan-Quinn criter.	4.746202
F-statistic	7.647268	Durbin-Watson stat	1.249444
Prob(F-statistic)	0.000001		

Sixth power

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.467460	0.642065	-2.285532	0.0226
@TREND	0.073753	0.023522	3.135477	0.0018
@TREND^2	-0.000750	0.000271	-2.771669	0.0057
@TREND^3	2.94E-06	1.34E-06	2.192028	0.0287
@TREND^4	-5.28E-09	3.22E-09	-1.641457	0.1011
@TREND^5	4.37E-12	3.68E-12	1.187942	0.2352
@TREND^6	-1.34E-15	1.61E-15	-0.834509	0.4043

R-squared	0.049143	Mean dependent var	0.260658
Adjusted R-squared	0.041567	S.D. dependent var	2.623782
S.E. of regression	2.568673	Akaike info criterion	4.733824
Sum squared resid	4968.354	Schwarz criterion	4.776499
Log likelihood	-1791.853	Hannan-Quinn criter.	4.750257
F-statistic	6.486225	Durbin-Watson stat	1.250599
Prob(F-statistic)	0.000001		

Seventh power

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1.476246	0.730705	-2.020303	0.0437
@TREND	0.074413	0.035178	2.115328	0.0347
@TREND^2	-0.000762	0.000540	-1.410490	0.1588
@TREND^3	3.03E-06	3.68E-06	0.821765	0.4115
@TREND^4	-5.59E-09	1.29E-08	-0.435243	0.6635
@TREND^5	4.97E-12	2.39E-11	0.207850	0.8354
@TREND^6	-1.91E-15	2.25E-14	-0.084824	0.9324
@TREND^7	2.14E-19	8.46E-18	0.025244	0.9799

R-squared	0.049144	Mean dependent var	0.260658
Adjusted R-squared	0.040293	S.D. dependent var	2.623782
S.E. of regression	2.570379	Akaike info criterion	4.736454
Sum squared resid	4968.350	Schwarz criterion	4.785226
Log likelihood	-1791.853	Hannan-Quinn criter.	4.755235
F-statistic	5.552334	Durbin-Watson stat	1.250600
Prob(F-statistic)	0.000003		

The model with the smallest Akaike information criterion and Schwarz criterion is the model with a fifth power trend.

c) To characterize the weekly ‘seasonal’ pattern, estimate the following model:

$$Happy_t = \beta_1 mon_t + \beta_2 tue_t + \beta_3 wed_t + \beta_4 thu_t + \beta_5 fri_t + \beta_6 sat_t + \beta_7 sun_t + \varepsilon_t$$

Dependent Variable: HAPPY
Method: Least Squares
Sample (adjusted): 9/09/2007 10/07/2009
Included observations: 760 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MON	-0.603670	0.243914	-2.474929	0.0135
TUE	-0.545872	0.243914	-2.237968	0.0255
WED	-0.136697	0.243914	-0.560432	0.5754
THU	0.708333	0.245041	2.890677	0.0040
FRI	1.202778	0.245041	4.908484	0.0000
SAT	0.974074	0.245041	3.975154	0.0001
SUN	0.244954	0.243914	1.004264	0.3156

R-squared	0.065461	Mean dependent var	0.260658
Adjusted R-squared	0.058014	S.D. dependent var	2.623782
S.E. of regression	2.546537	Akaike info criterion	4.716513
Sum squared resid	4883.091	Schwarz criterion	4.759189
Log likelihood	-1785.275	Hannan-Quinn criter.	4.732947
Durbin-Watson stat	1.218830		

d) Based on your results, on which day of the week people are the happiest? In which days are the unhappiest?

Looking at the coefficients on the estimated day dummies, we can see that the happiest day is Friday, followed by Saturday. The days where the happy index is the lowest is Monday and Tuesday.

e) Now, we want our model to capture the effects of particular dates on happiness. Estimate the following model:

$$Happy_t = \beta_1 mon_t + \beta_2 tue_t + \beta_3 wed_t + \beta_4 thu_t + \beta_5 fri_t + \beta_6 sat_t + \beta_7 sun_t + \beta_7 chris_t + \beta_7 easter_t + \beta_7 newyear_t + \beta_7 thanks_t + \beta_7 vale_t + \varepsilon_t$$

Dependent Variable: HAPPY
Method: Least Squares
Sample (adjusted): 9/09/2007 10/07/2009
Included observations: 760 after adjustments

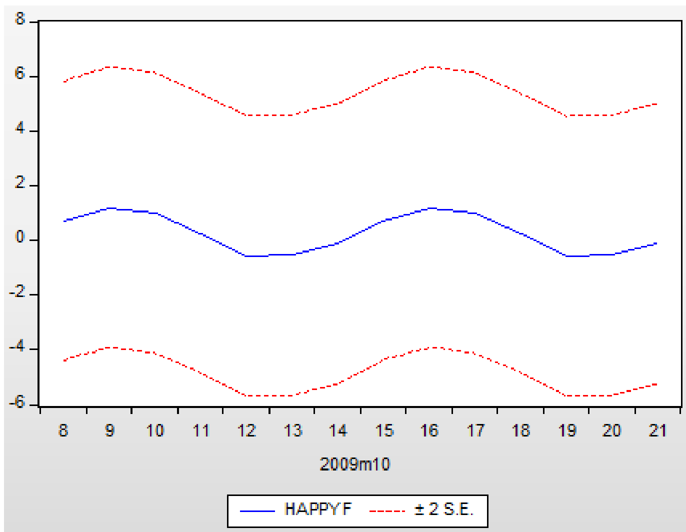
Variable	Coefficient	Std. Error	t-Statistic	Prob.
MON	-0.834221	0.152251	-5.479250	0.0000
TUE	-0.776423	0.152251	-5.099625	0.0000
WED	-0.714034	0.152971	-4.667774	0.0000
THU	0.064434	0.154051	0.418266	0.6759
FRI	1.202778	0.152598	7.882000	0.0000
SAT	0.912858	0.152954	5.968178	0.0000
SUN	0.091589	0.153309	0.597411	0.5504
CHRIS	17.29006	0.796668	21.70297	0.0000
EASTER	8.358411	1.131794	7.385102	0.0000
NEWYEAR	7.840061	0.796668	9.841063	0.0000
THANKS	18.89980	0.800422	23.61229	0.0000
VALE	6.611354	1.126628	5.868269	0.0000

R-squared	0.639981	Mean dependent var	0.260658
Adjusted R-squared	0.634686	S.D. dependent var	2.623782
S.E. of regression	1.585845	Akaike info criterion	3.775776
Sum squared resid	1881.149	Schwarz criterion	3.848934
Log likelihood	-1422.795	Hannan-Quinn criter.	3.803948
Durbin-Watson stat	1.888660		

f) Which one is the date that brings the most happiness? Is this consistent with the graph presented at the beginning of the assignment?

The date that brings more happiness appears to be Thanksgiving Holiday, followed by Christmas. This is consistent with the time series graph presented in the assignment as these dates appear with peaks.

g) Using the model estimated in part (c), forecast the level of happiness for the following dates: 10/08/2009 10/21/2009. Can you recognize the effect of the day of the week in the forecasted values?



Yes, the values where the forecast of happiness is higher corresponds to Fridays and Saturdays. Moreover, low values are associated with Mondays, Tuesdays, and Wednesdays.

h) Estimate the model using monthly seasonal dummies. You have to use the command: ls happy @seas(1) @seas(2) …. Recall that @seas(1) corresponds to January, @seas(2) to February and so on.[2] Is there any particular month of the year where people are happier?

[2] For some versions of EViews the following command can also work: ls happy @expand(@month)

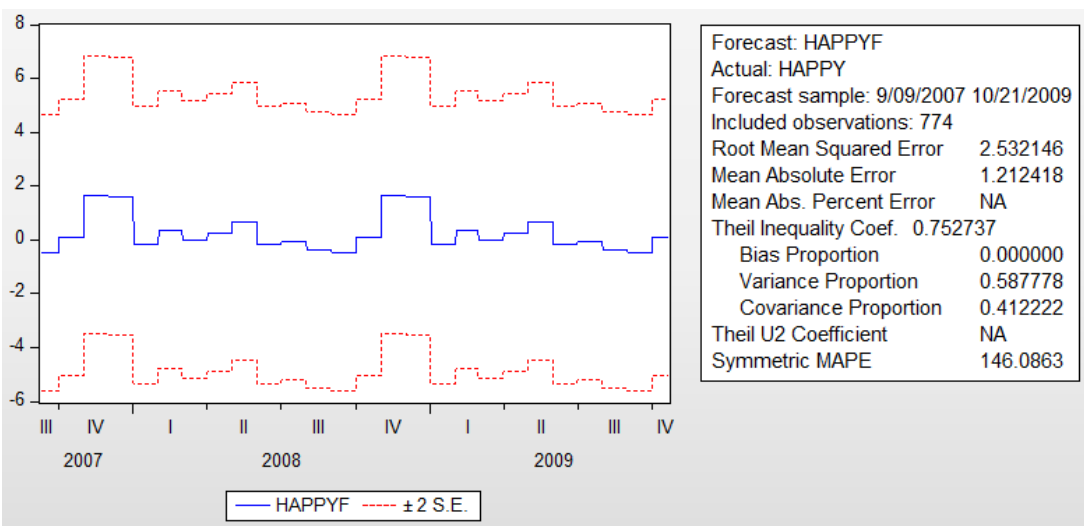
Dependent Variable: HAPPY
Method: Least Squares
Sample (adjusted): 9/09/2007 10/07/2009
Included observations: 760 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
@MONTH=1	-0.190323	0.324152	-0.587140	0.5573
@MONTH=2	0.364912	0.338071	1.079397	0.2808
@MONTH=3	0.003226	0.324152	0.009952	0.9921
@MONTH=4	0.263333	0.329510	0.799165	0.4244
@MONTH=5	0.687097	0.324152	2.119674	0.0344
@MONTH=6	-0.191667	0.329510	-0.581671	0.5610
@MONTH=7	-0.051613	0.324152	-0.159224	0.8735
@MONTH=8	-0.375806	0.324152	-1.159352	0.2467
@MONTH=9	-0.481707	0.281863	-1.709013	0.0879
@MONTH=10	0.108696	0.307270	0.353746	0.7236
@MONTH=11	1.670000	0.329510	5.068125	0.0000
@MONTH=12	1.617742	0.324152	4.990687	0.0000

R-squared	0.067404	Mean dependent var	0.260658
Adjusted R-squared	0.053689	S.D. dependent var	2.623782
S.E. of regression	2.552377	Akaike info criterion	4.727590
Sum squared resid	4872.941	Schwarz criterion	4.800748
Log likelihood	-1784.484	Hannan-Quinn criter.	4.755762
Durbin-Watson stat	1.273489		

Based on these regression results, the month with higher happiness is November followed by December. These appears to be the Thanksgiving and Christmas effects.

i) Obtain the in-sample forecast of the model estimated in part (h). Can you recognize the patterns of happiness obtained in part (h)?



The forecasting results are consistent with the ones reported in part (h). November and December are the months with higher happiness.