Problem Set 2 suggested answers

Ch 6. Problems and Complements:

2. (p. 108)

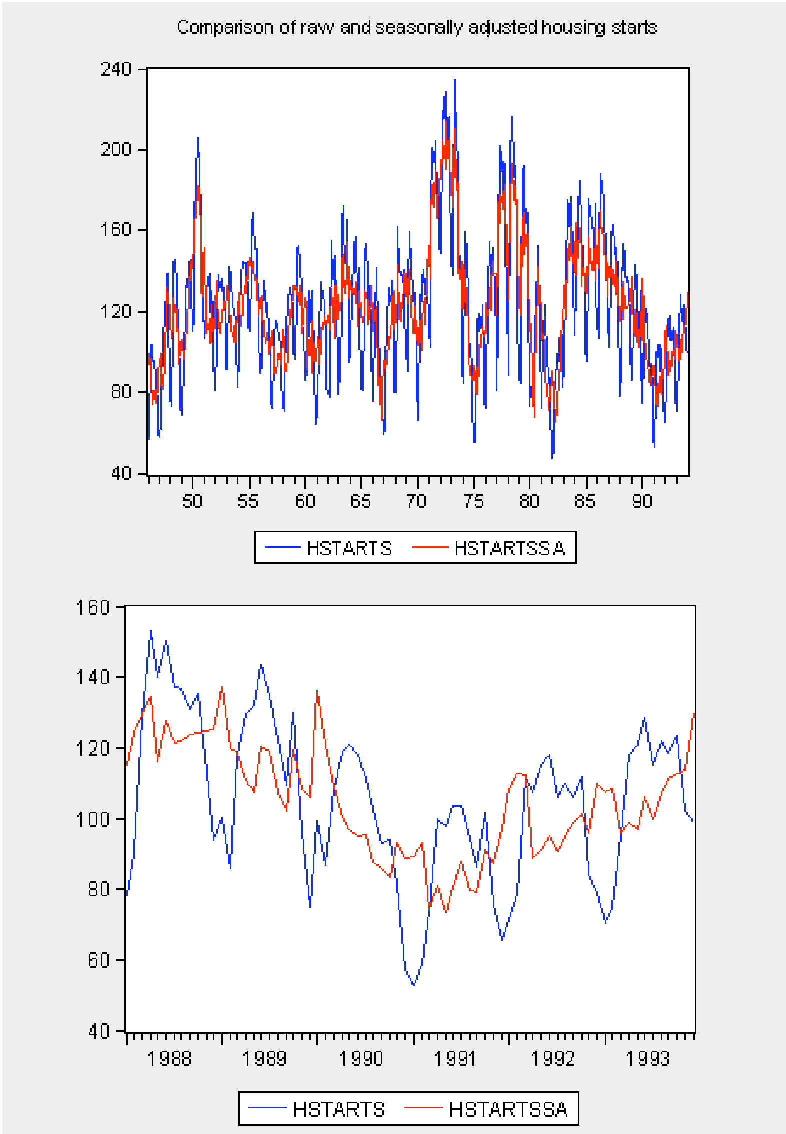
   a. To use a dummy variables regression as the basis for a seasonally-adjusted series, first estimate the model with terms as needed, and seasonal dummies. The residuals from this regression will have been "purged" of seasonal factors, but also of the constant and trend terms. You would then add back in the series mean and trend terms. That should give you a series with the same mean, trend, and cyclical behavior as the original series, but without the seasonality.
   b. Here we replicate the housing starts model with 12 monthly seasonal dummies from the textbook (there is no significant trend).

Dependent Variable: HSTARTS
Method: Least Squares
Date: 02/15/10 Time: 17:20
Sample: 1946M01 1993M12
Included observations: 576

	Coefficient	Std. Error	t-Statistic	Prob.
DM1	86.50417	4.029055	21.47009	0.0000
DM2	89.50417	4.029055	22.21468	0.0000
DM3	122.8833	4.029055	30.49929	0.0000
DM4	142.1687	4.029055	35.28588	0.0000
DM5	147.5000	4.029055	36.60908	0.0000
DM6	145.9979	4.029055	36.23627	0.0000
DM7	139.1125	4.029055	34.52733	0.0000
DM8	138.4167	4.029055	34.35462	0.0000
DM9	130.5625	4.029055	32.40524	0.0000
DM10	134.0917	4.029055	33.28117	0.0000
DM11	111.8333	4.029055	27.75671	0.0000
DM12	92.15833	4.029055	22.87344	0.0000

R-squared	0.383780	Mean dependent var	123.3944
Adjusted R-squared	0.371762	S.D. dependent var	35.21775
S.E. of regression	27.91411	Akaike info criterion	9.516755
Sum squared resid	439467.5	Schwarz criterion	9.607507
Log likelihood	-2728.825	Hannan-Quinn criter.	9.552147
Durbin-Watson stat	0.154140		

In this case, the only thing that needs to be added back to the residuals is the mean of the dependent variable. So the seasonally-adjusted series can be obtained from: "series hstartssa = resid + 123.3944".



The seasonally adjusted series has the same mean and cyclical path but does not have the sharp annual swing. This is clearer in the second figure that shows only the last several years of the sample.

c. A brief overview of the Census X-11 procedure is at http://www.statsoftinc.com/textbook/sttimser.html#x. There is also a fairly detailed description in the full Eviews manual that is in the "Docs" directory on your CD. (But the procedure itself is not included in the student version of the software.) The X-11 procedure allows for more than just the simple linear additive trend that you have tried. It also tests for outliers, etc. There is a newer Census X-12 ARIMA model that is now used more widely.

6. (p. 109)

a. Here are linear, quadratic and exponential models. I have estimated the exponential model rather than the similar log-linear model to facilitate comparisons.

Dependent Variable: HITS
Method: Least Squares
Date: 02/23/03 Time: 20:31
Sample: 1/01/1998 9/28/1998
Included observations: 271

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13702.08	524.2944	26.13432	0.0000
TIME	38.61834	3.360241	11.49273	0.0000

R-squared	0.329316	Mean dependent var	18915.55
Adjusted R-squared	0.326822	S.D. dependent var	5274.313
S.E. of regression	4327.437	Akaike info criterion	19.59069
Sum squared resid	5.04E+09	Schwarz criterion	19.61727
Log likelihood	-2652.539	F-statistic	132.0829
Durbin-Watson stat	1.260251	Prob(F-statistic)	0.000000

Dependent Variable: HITS
Method: Least Squares
Date: 02/23/03 Time: 20:31
Sample: 1/01/1998 9/28/1998
Included observations: 271

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	12904.17	781.5426	16.51115	0.0000
TIME	56.41562	13.37270	4.218715	0.0000
TIME^2	-0.065916	0.047945	-1.374829	0.1703

R-squared	0.334013	Mean dependent var	18915.55
Adjusted R-squared	0.329043	S.D. dependent var	5274.313
S.E. of regression	4320.294	Akaike info criterion	19.59104
Sum squared resid	5.00E+09	Schwarz criterion	19.63092
Log likelihood	-2651.586	F-statistic	67.20504
Durbin-Watson stat	1.269116	Prob(F-statistic)	0.000000

Dependent Variable: HITS
Method: Least Squares
Date: 02/23/03 Time: 20:32
Sample: 1/01/1998 9/28/1998
Included observations: 271
Convergence achieved after 45 iterations
HITS=C(1)*EXP(C(2)*TIME)

	Coefficient	Std. Error	t-Statistic	Prob.
C(1)	14271.74	456.5735	31.25836	0.0000
C(2)	0.001999	0.000182	11.01016	0.0000

R-squared	0.322391	Mean dependent var	18915.55
Adjusted R-squared	0.319872	S.D. dependent var	5274.313
S.E. of regression	4349.719	Akaike info criterion	19.60096
Sum squared resid	5.09E+09	Schwarz criterion	19.62755
Log likelihood	-2653.930	Durbin-Watson stat	1.247390

Notice that there are not huge differences in adjusted R^2, AIC or SIC criteria, although AIC and SIC pick the linear model. The quadratic model has a slightly higher adjusted R^2, but the AIC and SIC criteria are telling us it is not worth the added parameter. And when they are so close in in-sample performance, we would probably opt for the "smaller" model anyway.

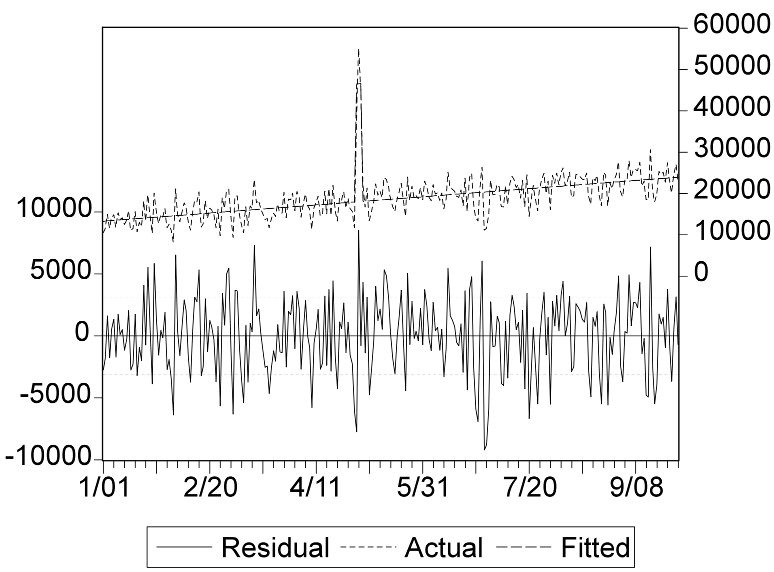
b. Here is the linear model with a dummy variable to "take out" the three days with extraordinarily high numbers of hits. (The dummy variable has "1" in 4/30/98-5/2/98, and zeros for all other periods.) Notice that while it improves the fit substantially it does not markedly change the other coefficients. That makes sense since (visually) the model seemed to be capturing the trend fairly well in the original linear model.

Dependent Variable: HITS
Method: Least Squares
Date: 02/23/03 Time: 20:37
Sample: 1/01/1998 9/28/1998
Included observations: 271

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13282.00	380.2302	34.93146	0.0000
TIME	39.39197	2.431370	16.20155	0.0000
DUMNEW	28512.97	1817.892	15.68463	0.0000

R-squared	0.650310	Mean dependent var	18915.55
Adjusted R-squared	0.647700	S.D. dependent var	5274.313
S.E. of regression	3130.560	Akaike info criterion	18.94682
Sum squared resid	2.63E+09	Schwarz criterion	18.98670
Log likelihood	-2564.294	F-statistic	249.1965
Durbin-Watson stat	1.943985	Prob(F-statistic)	0.000000

Model with Dummy for 3 Huge days



c. Including the day-of-the-week effects in the linear model gives:

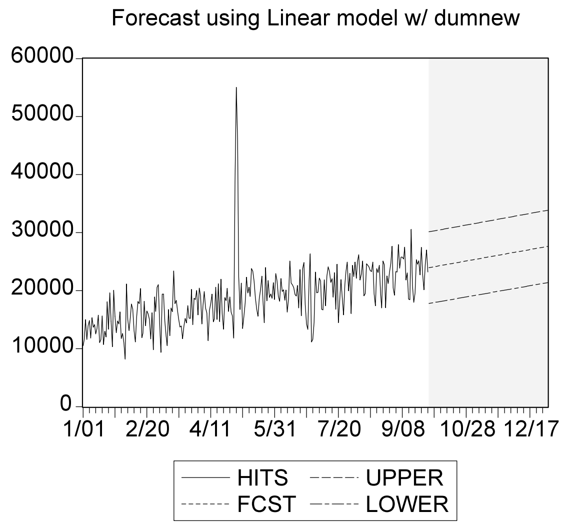
Dependent Variable: HITS
Method: Least Squares
Date: 02/23/03 Time: 20:50
Sample: 1/01/1998 9/28/1998
Included observations: 271

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	13342.38	599.9094	22.24066	0.0000
TIME	39.38135	2.425037	16.23948	0.0000
DUMNEW	28457.40	1826.439	15.58081	0.0000
WD1	976.3725	708.5659	1.377956	0.1694
WD2	-473.1114	708.5753	-0.667694	0.5049
WD3	-418.5919	713.1935	-0.586926	0.5578
WD4	-503.0259	713.1904	-0.705318	0.4812
WD5	-387.8783	707.0321	-0.548601	0.5837
WD6	377.2531	707.0196	0.533582	0.5941

R-squared	0.659985	Mean dependent var	18915.55
Adjusted R-squared	0.649603	S.D. dependent var	5274.313
S.E. of regression	3122.096	Akaike info criterion	18.96304
Sum squared resid	2.55E+09	Schwarz criterion	19.08267
Log likelihood	-2560.492	F-statistic	63.56926
Durbin-Watson stat	1.940700	Prob(F-statistic)	0.000000

Interestingly, none of the day dummies are significantly different from zero, so we conclude that there are no important day-of-the-week effects in this data series. (You could test this formally as you did in the first question, above.) So we don't need to include these terms in our model.

d. On the basis of the preceding analysis I would chose the simple linear trend model with DUMNEW and no seasonal factors.
e. To forecast through the end of the year, you need the dummies defined out that far. If you did not originally set up your workfile through year end, you can do so using the proc—structure/resize page menu item. You can then regenerate the "time" series and the dumnew (if you are including that). You would also need to extend the day-of-the-week dummies if you wanted to include them.



<u>Extra Problem</u>:

Go online and download a monthly or quarterly time series for the U.S. economy. Using all but he most recent year of data for your time series, select a forecasting model. You should allow for both trend and seasonal factors, and you should consider alternative types of trend modeling. Comment on your model's residual correlation, normality, equation fit, etc.

Use your model to forecast over the most recent year. Create a graph of forecast and actuals, with error bands, similar to Fig. 5.9 on page 115.

Some specific things that I was looking for here include: (a) that you document the source of your data series, (b) that you reported several alternative models and explained how you chose among them, (c) that you made appropriate decisions about the inclusion or exclusion of variables, such as seasonal dummies, and (d) that you carefully graphed and evaluated your selected model’s out-of-sample performance.

Note: If you want to be able to perform true out-of-sample forecast evaluation, you must exclude that period from your estimation sample period. Otherwise, you will be biasing your results in favor of your model, since least squares tried to fit the data of the sample period.