Econ 427, Spring 2010

Problem Set 4—Suggested Answers

Ch 9 Exercises, Problems and Complements

2. (p. 184)

You did the first part of this back in Ch. 8, prob. 6. Recall that the ARMA selection process suggested that the best fitting model was an ARMA(1,1).

(a) First 175 business days takes you thru Sep 1, 2000 the way I set up my workfile. Here are the results for the shorter sample.

Akaike

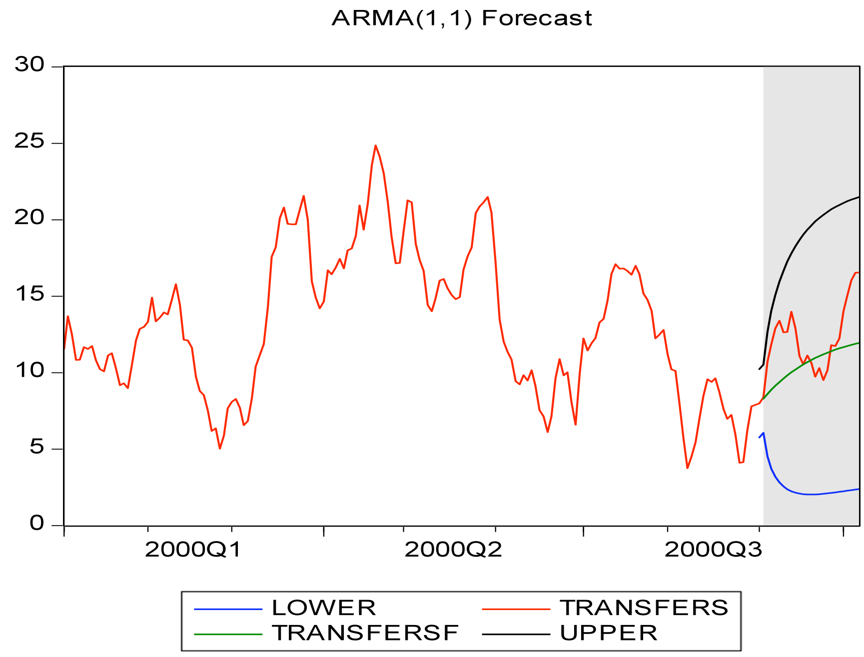
	MA Order					
	AR \ MA	0	1	2	3	4
AR order	0	6.010980	4.794815	4.060573	3.944296	3.444045
	1	3.363355	3.078906	3.090251	3.089203	3.097132
	2	3.129039	3.089926	3.095747	3.098996	3.097294
	3	3.120731	3.094749	3.103918	3.036291	3.108725
	4	3.109935	3.101321	3.111499	3.047191	3.053343

Schwartz

	MA Order					
AR order	AR \ MA	0	1	2	3	4
	0	6.029353	4.831560	4.115690	4.017785	3.535907
	1	3.400099	3.134023	3.163740	3.181065	3.207366
	2	3.184156	3.163415	3.187609	3.209230	3.225900
	3	3.194220	3.186611	3.214152	3.164898	3.255703
	4	3.201796	3.211555	3.240105	3.194170	3.218694

Well, a lot of work to get basically the same result. It's either an ARMA (3,3) or ARMA(1,1). Again, because of parsimony and the possible "common factors" I would go with the ARMA(1,1) Of course, in general we would not necessary expect to select the same model when using different samples of data.

(b) Here's the point and interval forecasts for days 176-200 using the ARMA(1,1) estimated over the first 175 days of data.



(c) The forecast captures the general upward drift in the series but clearly not all of the short-run swings. There is a lot of inherent volatility in the series, and we see that in the very wide error bands around our forecast. While the actuals do not cross outside these wide bands, the forecast is far from perfect.

(d) See the Eviews help files under "Forecast basics" and "forecasting with ARMA errors."

4. (p. 185) $y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2}$

(a) optimal 1-step ahead forecast. Write out the process at time T+1:

$$y_{T+1} = \phi_1 y_T + \phi_2 y_{T-1} + \varepsilon_{T+1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}$$

Project on the time T information set:

$$y_{T+1,T} = \phi_1 y_T + \phi_2 y_{T-1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}$$

(b) optimal 2-step ahead forecast. Write out the process at time T+2:

$$y_{T+2} = \phi_1 y_{T+1} + \phi_2 y_T + \varepsilon_{T+2} + \theta_1 \varepsilon_{T+1} + \theta_2 \varepsilon_T$$

Project on the time T information set:

$$y_{T+2,T} = \phi_1 y_{T+1,T} + \phi_2 y_T + \theta_2 \varepsilon_T$$

But we already have an optimal forecast of y_{T+1}. Substituting that,
$$y_{T+1,T} = \phi_1 y_T + \phi_2 y_{T-1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}$$
Substituting,

$$y_{T+2,T} = \phi_1 [\phi_1 y_T + \phi_2 y_{T-1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}] + \phi_2 y_T + \theta_2 \varepsilon_T$$
$$= \phi_1^2 y_T + \phi_1 \phi_2 y_{T-1} + \phi_1 \theta_1 \varepsilon_T + \phi_1 \theta_2 \varepsilon_{T-1} + \phi_2 y_T + \theta_2 \varepsilon_T$$
$$= \left(\phi_1^2 + \phi_2\right) y_T + \phi_1 \phi_2 y_{T-1} + \left(\phi_1 \theta_1 + \theta_2\right) \varepsilon_T + \phi_1 \theta_2 \varepsilon_{T-1}$$

(c) optimal 3-step ahead forecast. Write out the process at time T+3:

$$y_{T+3} = \phi_1 y_{T+2} + \phi_2 y_{T+1} + \varepsilon_{T+3} + \theta_1 \varepsilon_{T+2} + \theta_2 \varepsilon_{T+1}$$

Project on the time T information set:

$$y_{T+3,T} = \phi_1 y_{T+2,T} + \phi_2 y_{T+1,T}$$

Substituting optimal forecasts for y_{T+1} and y_{T+2} gives:

$$y_{T+3,T} = \phi_1 \left(\left(\phi_1^2 + \phi_2\right) y_T + \phi_1 \phi_2 y_{T-1} + \left(\phi_1 \theta_1 + \theta_2\right) \varepsilon_T + \phi_1 \theta_2 \varepsilon_{T-1}\right) + \phi_2 \left(\phi_1 y_T + \phi_2 y_{T-1} + \theta_1 \varepsilon_T + \theta_2 \varepsilon_{T-1}\right)$$

(You can do the simplifying!)

(d) optimal h-step ahead forecast. Write out the process at time T+h:

$$y_{T+h} = \phi_1 y_{T+h-1} + \phi_2 y_{T+h-2} + \varepsilon_{T+h} + \theta_1 \varepsilon_{T+h-1} + \theta_2 \varepsilon_{T+h-2}$$

Project on the time T information set:

$$y_{T+h,T} = \phi_1 y_{T+h-1,T} + \phi_2 y_{T+h-2,T}$$

for h>2.

Ch 10. Exercises, Problems and Complements:

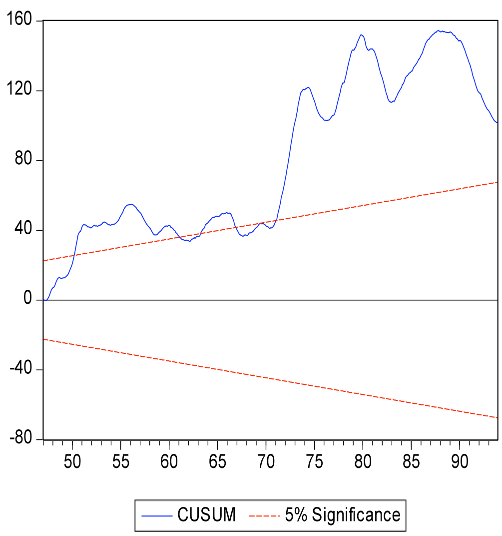
4. (p. 215) (a) Here's the model from Chapter 6 (reestimated in the preferred form with a constant plus 11 of the monthly dummies):

Dependent Variable: HSTARTS
Method: Least Squares
Date: 02/15/10 Time: 17:21
Sample: 1946M01 1993M12
Included observations: 576

	Coefficient	Std. Error	t-Statistic	Prob.
C	92.15833	4.029055	22.87344	0.0000
DM1	-5.654167	5.697944	-0.992317	0.3215
DM2	-2.654166	5.697944	-0.465811	0.6415
DM3	30.72500	5.697944	5.392295	0.0000
DM4	50.01042	5.697944	8.776923	0.0000
DM5	55.34167	5.697944	9.712567	0.0000
DM6	53.83958	5.697944	9.448949	0.0000
DM7	46.95417	5.697944	8.240545	0.0000
DM8	46.25833	5.697944	8.118425	0.0000
DM9	38.40417	5.697944	6.740004	0.0000
DM10	41.93333	5.697944	7.359379	0.0000
DM11	19.67500	5.697944	3.453000	0.0006

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
F-statistic	31.93250	Durbin-Watson stat	0.154140
Prob(F-statistic)	0.000000		

(a) CUSUM plot of the model with constant and seasonal dummies is:



Clearly there is evidence of parameter instability in this model.

(b) And (c) Unfortunately, Eviews does not permit us to do CUSUM analysis on models with ARMA terms. That's why I suggested that you use one or more lagged dependent variables to capture cyclicality.

Lagged Dep. Vars:	1	AIC: 7.6115	SIC: 7.7099
Lagged Dep. Vars:	2	AIC: 7.5921	SIC: 7.6983
Lagged Dep. Vars:	3	AIC: 7.5961	SIC: 7.7101
Lagged Dep. Vars:	4	AIC: 7.5957	SIC: 7.7174

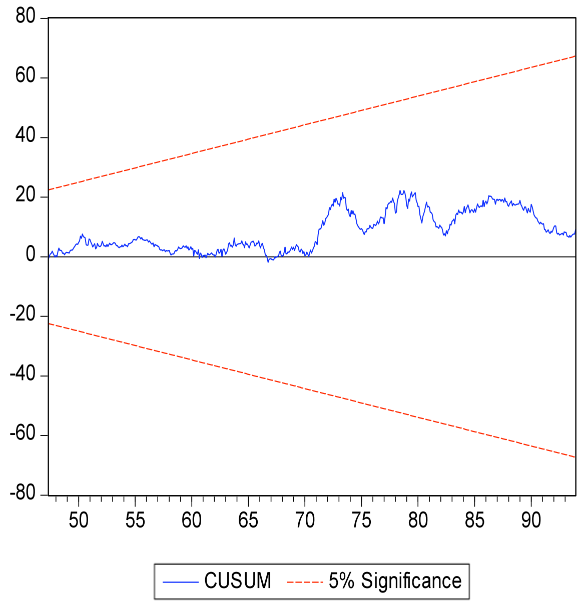
The model with 2 lags of the dependent variable fits best and it yields serially uncorrelated residuals (there is some non-normality but not much we can do about that). Here is the model

Dependent Variable: HSTARTS
Method: Least Squares
Date: 03/25/10 Time: 09:58
Sample (adjusted): 1946M03 1993M12
Included observations: 574 after adjustments

	Coefficient	Std. Error	t-Statistic	Prob.
C	-15.76406	2.691372	-5.857257	0.0000
HSTARTS(-1)	0.778208	0.041817	18.60967	0.0000
HSTARTS(-2)	0.155810	0.041780	3.729330	0.0002
DM1	13.82960	2.213029	6.249172	0.0000
DM2	23.64535	2.361338	10.01354	0.0000
DM3	55.51633	2.473544	22.44404	0.0000
DM4	48.35837	3.161381	15.29660	0.0000
DM5	33.48076	2.765250	12.10768	0.0000
DM6	24.82500	2.475974	10.02636	0.0000
DM7	18.27786	2.366128	7.724797	0.0000
DM8	23.17435	2.285249	10.14085	0.0000
DM9	16.93450	2.360199	7.175033	0.0000
DM10	26.68427	2.259665	11.80895	0.0000
DM11	2.903259	2.421326	1.199037	0.2310

R-squared	0.910035	Mean dependent var	123.6118
Adjusted R-squared	0.907946	S.D. dependent var	35.08457
S.E. of regression	10.64480	Akaike info criterion	7.592107
Sum squared resid	63454.56	Schwarz criterion	7.698269
Log likelihood	-2164.935	Hannan-Quinn criter.	7.633516
F-statistic	435.7393	Durbin-Watson stat	2.005201
Prob(F-statistic)	0.000000		

Here is the CUSUM graph, which shows no breaks outside the 90% confidence interval. Modeling the cycle helps us obtain stable parameters for the other constant and seasonal factors.



6. (p. 215-216)

a. Airline passenger miles might be seasonal because of summer vacations and other holidays.

b. Here is a model with seasonal factors, and I also include a linear time trend since that seems to be significant:

Dependent Variable: AIRSPEED
Method: Least Squares
Date: 10/19/07 Time: 14:14
Sample: 1990Q1 2005Q4
Included observations: 64

	Coefficient	Std. Error	t-Statistic	Prob.
TIME	0.025296	0.006024	4.199129	0.0001
DM1	-1.143118	0.286386	-3.991524	0.0002
DM2	-0.422521	0.290226	-1.455835	0.1507
DM3	2.748285	0.294138	9.343518	0.0000
DM4	-0.712199	0.298121	-2.388960	0.0201

R-squared	0.782503	Mean dependent var	0.914446
Adjusted R-squared	0.767757	S.D. dependent var	1.843982
S.E. of regression	0.888644	Akaike info criterion	2.676664
Sum squared resid	46.59157	Schwarz criterion	2.845326
Log likelihood	-80.65324	Durbin-Watson stat	2.588298

The result of a Wald test that all four seasonal dummies have equal coefficients easily rejects that null hypothesis, so there is evidence of statistically-significant seasonality:

Wald Test:
Equation: EQSEAS

Test Statistic	Value	df	Probability
F-statistic	64.06964	(3, 59)	0.0000
Chi-square	192.2089	3	0.0000

Note that you cannot tell from the t-statistics on individual seasonal dummies whether they are significant in this model with s seasonal dummies and no constant. You may have done it with constant and 3 seasonal dummies, in which case the test for seasonality is a test of whether the coefficients on the 3 seasonal dummies are equal to zero.

Comparing the model to one with only a constant and trend, we can see that adding the seasonals reduces the AIC and SIC considerably, another indication that these seasonal dummies help to explain the evolution of our data over time:

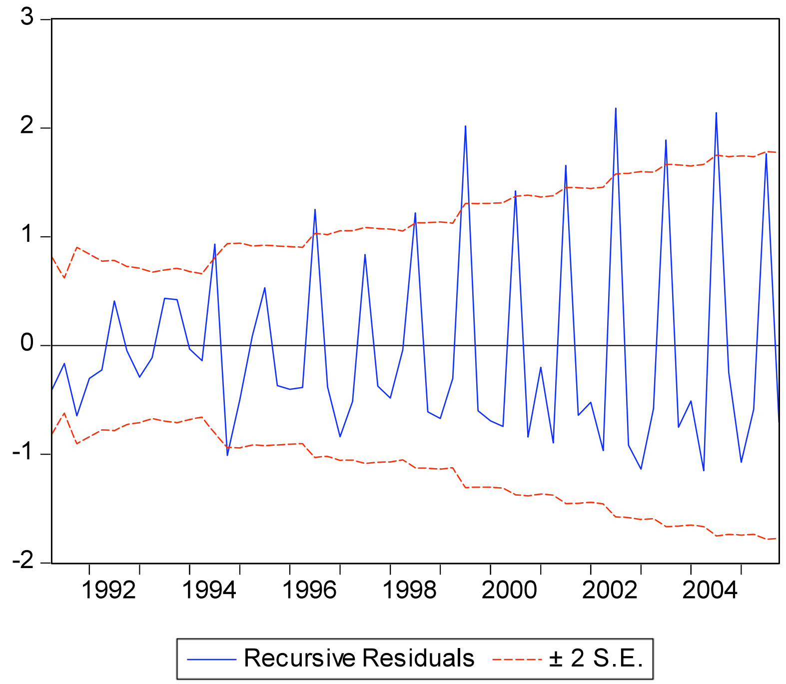
Dependent Variable: AIRSPEED
Method: Least Squares
Date: 10/31/07 Time: 13:57
Sample: 1990Q1 2005Q4
Included observations: 64

	Coefficient	Std. Error	t-Statistic	Prob.
C	0.066109	0.441997	0.149569	0.8816
TIME	0.026931	0.012104	2.225024	0.0297

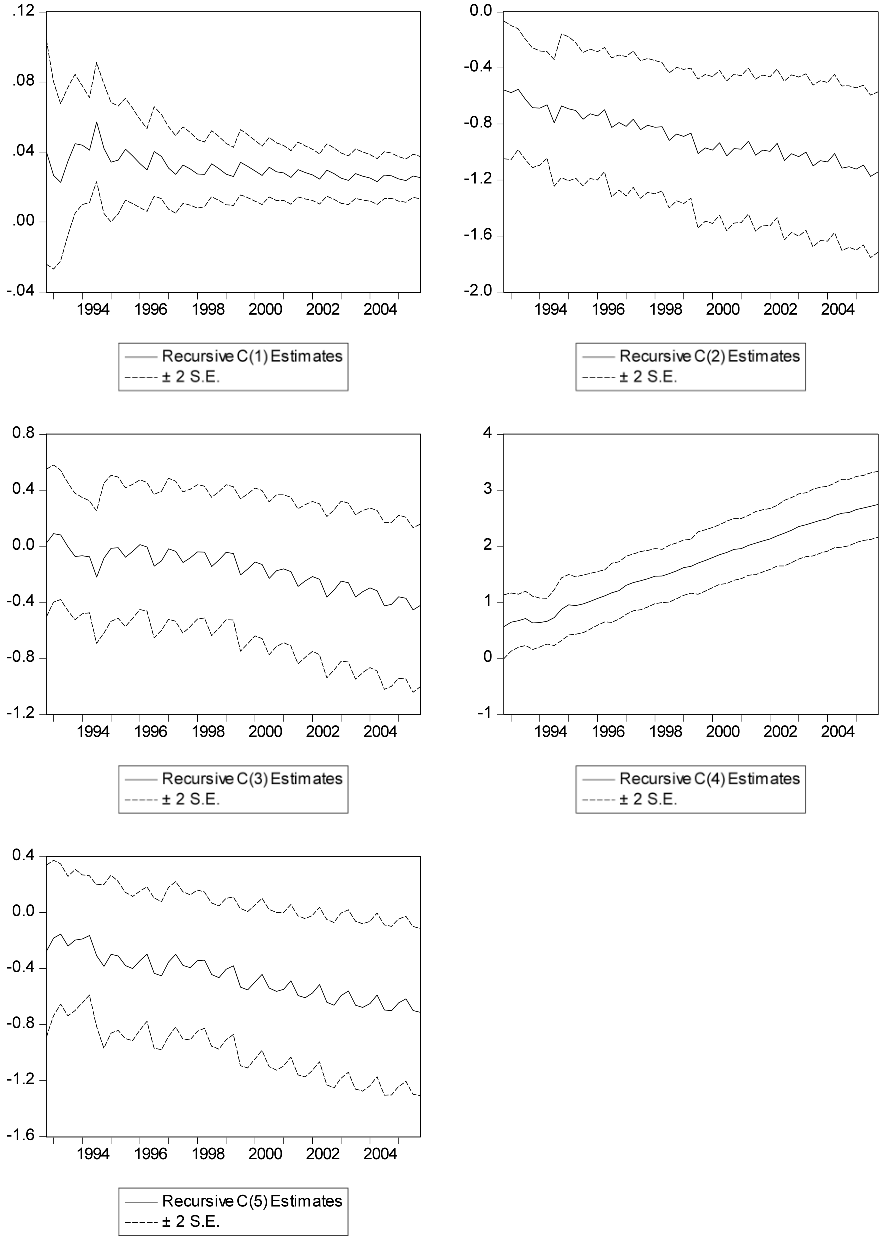
R-squared	0.073946	Mean dependent var	0.914446
Adjusted R-squared	0.059010	S.D. dependent var	1.843982
S.E. of regression	1.788749	Akaike info criterion	4.031661
Sum squared resid	198.3766	Schwarz criterion	4.099126
Log likelihood	-127.0132	Hannan-Quinn criter.	4.058239
F-statistic	4.950733	Durbin-Watson stat	2.444515
Prob(F-statistic)	0.029727		

Looking at the coefficients in the model with trend and seasonal dummies, above, it is clear that the first quarter has lower-than-average travel and the third quarter has greater-than-average travel, consistent with out notion that summer vacations should contribute to seasonality.

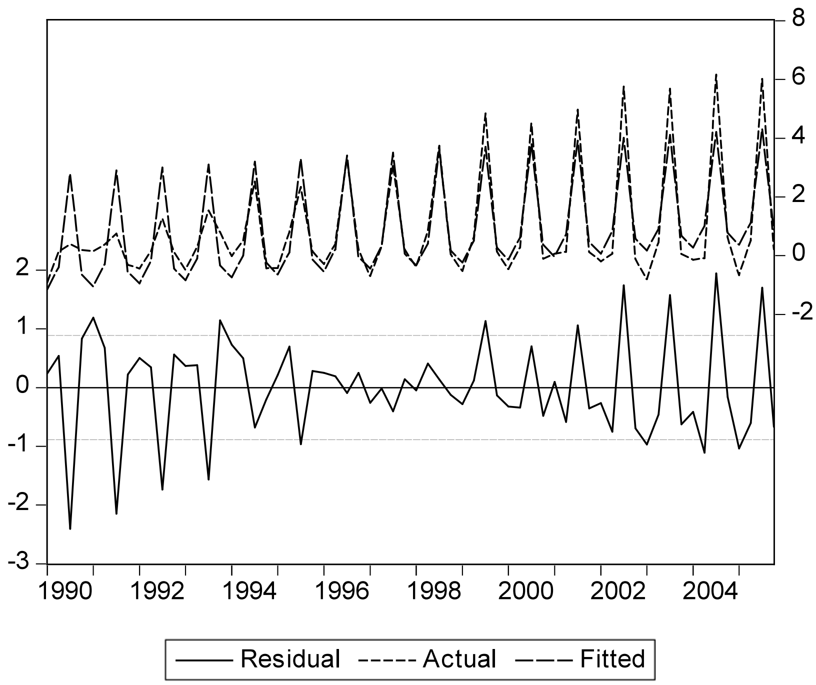
c. The CUSUM test does not show any obvious instability, although the CUSUM-of-square diagram does indicate some possible instability. (I am not showing these figures here.) The plot of recursive residuals shows a pattern of repeated breakouts, suggesting parameter instability:



A plot of recursive coefficients indicates that there is evidence of parameter instability, primarily in the seasonal dummy for the third quarter. It appears that more and more of the annual travel is occurring in the third quarter as time goes by:



This is also evidence (but maybe harder to see) from an inspection of the plot of actual, fitted and residuals:



The spikes in residuals for the 3rd quarter are mostly negative early in the sample and mostly positive later on as 3rd quarter travel becomes relatively more important.

d. One can allow for the coefficient on a seasonal dummy to change over time by "interacting" the dummy with time:

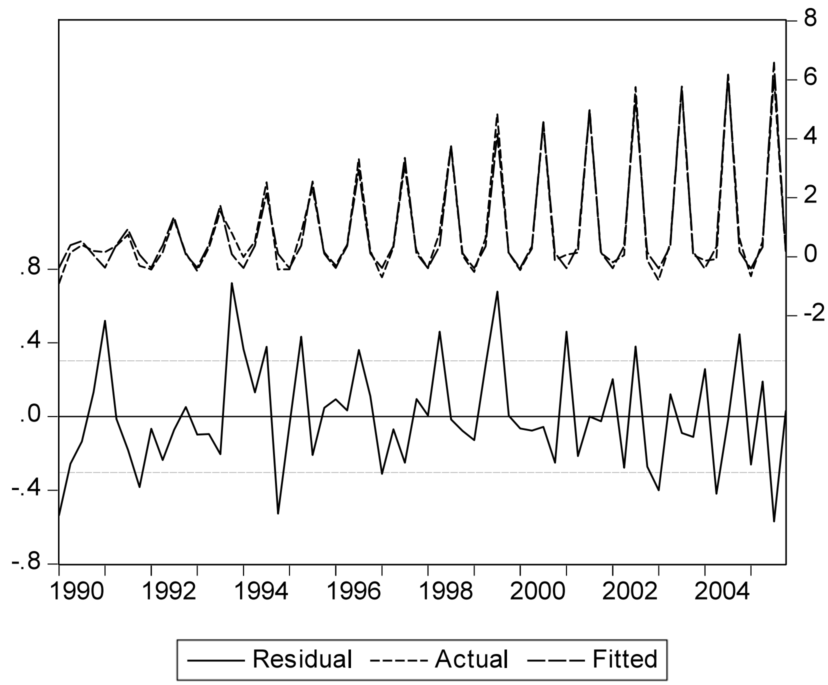
Dependent Variable: AIRSPEED
Method: Least Squares
Date: 10/19/07 Time: 14:17
Sample: 1990Q1 2005Q4
Included observations: 64

	Coefficient	Std. Error	t-Statistic	Prob.
DM1	-0.367811	0.144543	-2.544651	0.0137
DM2	0.396916	0.148055	2.680862	0.0096
DM3	0.329870	0.151598	2.175954	0.0338
DM4	0.056650	0.155168	0.365087	0.7164
DM1*TIME	-0.000547	0.004105	-0.133322	0.8944
DM2*TIME	-0.001137	0.004105	-0.277026	0.7828
DM3*TIME	0.100872	0.004105	24.57446	0.0000
DM4*TIME	0.001998	0.004105	0.486723	0.6284

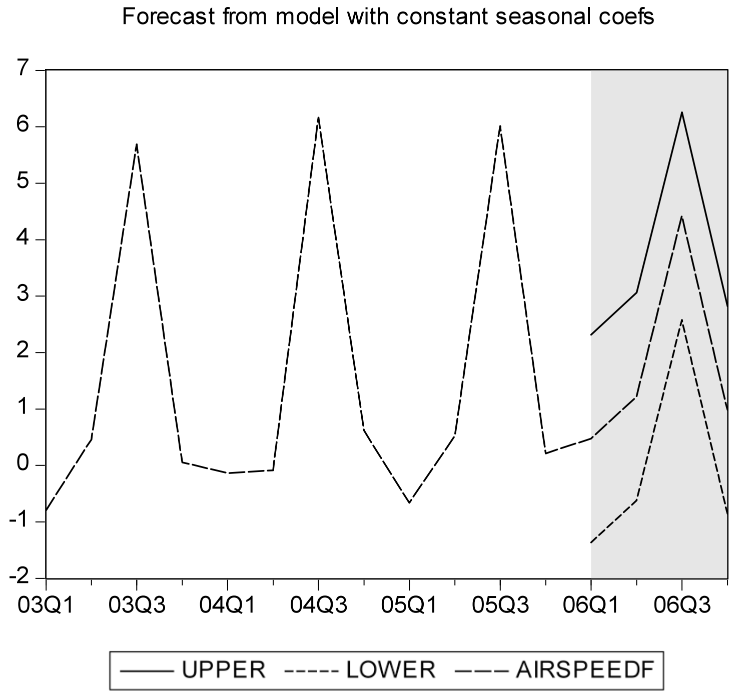
R-squared	0.976039	Mean dependent var	0.914446
Adjusted R-squared	0.973044	S.D. dependent var	1.843982
S.E. of regression	0.302751	Akaike info criterion	0.564655
Sum squared resid	5.132848	Schwarz criterion	0.834515
Log likelihood	-10.06895	Durbin-Watson stat	2.076128

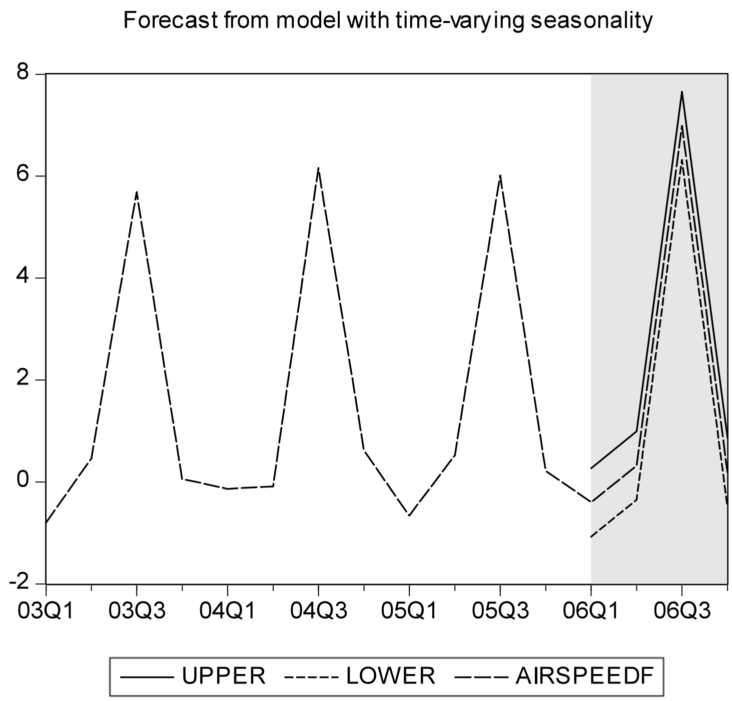
Notice that I also included the seasonal dummies by themselves. (I have to leave the trend out now or I will have perfect colinearity.) In this specification, the interacted terms tell us how the coefficient is changing over time. Notice that the only one interaction term that is significant is the one for the third quarter. You could consider dropping the other three interaction terms and see how that affects the model's fit measured by AIC and SIC.

This model clearly fits the data much better than the model with constant seasonal coefficients, as evidenced by the actual, fitted, residual graph:



e. The four quarter extrapolative forecasts for the two models are shown below:





As one might expect, the model with constant seasonal coefficients tends to under-predict the third-quarter spike compared with recent history; the model with time-varying seasonality produces a forecast that appears more "reasonable" compared with recent history. There also appears to be a tighter confidence interval, presumably because we are explaining more of the variation in the data series. (Note that I had to extend the workfile by one year and regenerate the trend and seasonal dummies before I could create these out-of-sample forecasts.)