Texas Tech University

Department of Economics

Spring 2018

Eco 4306: Economic and Business Forecasting

Midterm 2

Name:

ID:

Short questions (45 points)

Q1. 7.5 points
Q2. 7.5 points
Q3. 7.5 points
Q4. 7.5 points
Q5. 7.5 points
Q6. 7.5 points

Applied problems (60 points)

Q7. 10 points
Q8. 10 points
Q9. 10 points
Q10. 10 points
Q11. 10 points
Q12. 10 points

Good luck!

Question 1 (7.5 points)

Explain the difference between in-sample evaluation and out-of-sample evaluation.

See slide 2 in lec13slides.pdf and slide 10 in lec16slides.pdf.

Question 2 (7.5 points)

Explain how Mean Squared Error and Mean Loss are used in the assessment of forecasts.

See slides 9 and 14 in lec13slides.pdf.

Question 3 (7.5 points)

Give an example of a deterministic trend $g(t)$ other than a linear trend and plot its graph. Write the equation of a model with this trend.

See slides 10 to 12 in lec16slides.pdf.

Question 4 (7.5 points)

Explain the difference between a trend stationary time series and a difference stationary time series.

See slide 15 in lec18slides.pdf.

Question 5 (7.5 points)

Explain what it means for a time series process to be $I(1)$, and what it means for a process to be $I(0)$.

See slide 14 in lec18slides.pdf.

Question 6 (7.5 points)

Write down the equation for a pure seasonal S-AR(1) model. Describe how its AC and PAC functions look like.

See slides 9 and 10 in lec11slides.pdf.

Question 7 (10 points)

Consider two candidate models for change in monthly private residential construction spending, AR(1) and AR(2)+SAR(1), the results for which are below. Evaluate the adequacy of these models based on the correlograms of residuals, AIC and BIC, and statistical significance of coefficients.

Dependent Variable: DCONST
Method: ARMA Maximum Likelihood (BFGS)
Date: 04/11/18 Time: 16:28
Sample: 1993M02 2013M12
Included observations: 251
Convergence achieved after 3 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	34.02301	284.5681	0.119560	0.9049
AR(1)	0.503787	0.082472	6.108569	0.0000
SIGMASQ	4263658.	311155.6	13.70266	0.0000

R-squared	0.254386	Mean dependent var	49.44223
Adjusted R-squared	0.248373	S.D. dependent var	2396.078
S.E. of regression	2077.314	Akaike info criterion	18.12859
Sum squared resid	1.07E+09	Schwarz criterion	18.17072
Log likelihood	-2272.138	Hannan-Quinn criter.	18.14554
F-statistic	42.30579	Durbin-Watson stat	2.030264
Prob(F-statistic)	0.000000		

Date: 04/11/18 Time: 16:30
Sample: 1993M01 2013M12
Included observations: 251
Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.016	-0.016	0.0654	
		2	0.196	0.196	9.8361	0.002
		3	-0.156	-0.156	16.045	0.000
		4	-0.152	-0.202	21.954	0.000
		5	-0.073	-0.014	23.343	0.000
		6	-0.437	-0.430	72.819	0.000
		7	-0.077	-0.174	74.368	0.000
		8	-0.167	-0.097	81.652	0.000
		9	-0.141	-0.424	86.868	0.000
		10	0.174	-0.055	94.847	0.000
		11	-0.004	-0.168	94.852	0.000
		12	0.928	0.873	323.74	0.000
		13	-0.022	0.004	323.87	0.000
		14	0.191	-0.179	333.60	0.000
		15	-0.170	-0.052	341.37	0.000
		16	-0.138	0.000	346.48	0.000
		17	-0.073	-0.027	347.93	0.000
		18	-0.426	0.036	397.29	0.000
		19	-0.074	0.040	398.77	0.000
		20	-0.181	-0.046	407.75	0.000
		21	-0.107	0.110	410.89	0.000
		22	0.145	-0.070	416.75	0.000
		23	0.014	-0.017	416.81	0.000
		24	0.859	-0.003	623.42	0.000
		25	-0.023	0.005	623.56	0.000
		26	0.187	-0.009	633.38	0.000
		27	-0.183	-0.013	642.88	0.000
		28	-0.128	-0.051	647.55	0.000
		29	-0.074	-0.018	649.09	0.000
		30	-0.408	0.052	696.98	0.000
		31	-0.070	-0.020	698.38	0.000
		32	-0.203	-0.115	710.36	0.000
		33	-0.069	0.028	711.77	0.000
		34	0.118	-0.001	715.82	0.000
		35	0.035	-0.014	716.18	0.000
		36	0.783	-0.102	897.23	0.000

Dependent Variable: DCONST
Method: ARMA Maximum Likelihood (BFGS)
Date: 04/11/18 Time: 16:26
Sample: 1993M02 2013M12
Included observations: 251
Convergence achieved after 7 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	214.4219	1374.109	0.156044	0.8761
AR(1)	0.497140	0.054595	9.105947	0.0000
AR(2)	0.116143	0.052147	2.227211	0.0268
SAR(12)	0.944592	0.013249	71.29646	0.0000
SIGMASQ	373960.6	26485.07	14.11968	0.0000

R-squared	0.934603	Mean dependent var	49.44223
Adjusted R-squared	0.933540	S.D. dependent var	2396.078
S.E. of regression	617.7066	Akaike info criterion	15.81783
Sum squared resid	93864109	Schwarz criterion	15.88805
Log likelihood	-1980.137	Hannan-Quinn criter.	15.84609
F-statistic	878.9103	Durbin-Watson stat	1.975678
Prob(F-statistic)	0.000000		

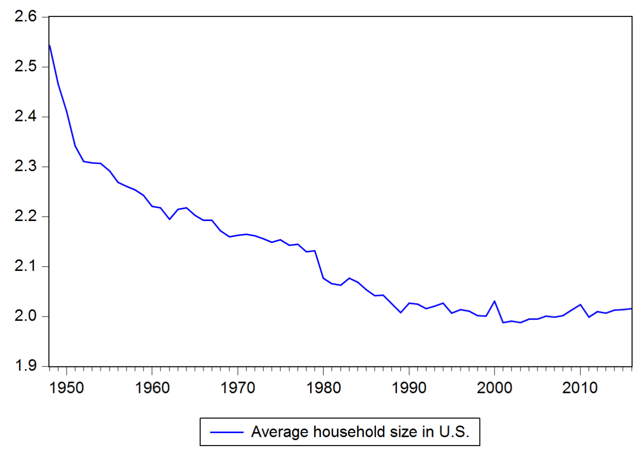
Date: 04/11/18 Time: 16:26
Sample: 1993M01 2013M12
Included observations: 251
Q-statistic probabilities adjusted for 3 ARMA terms

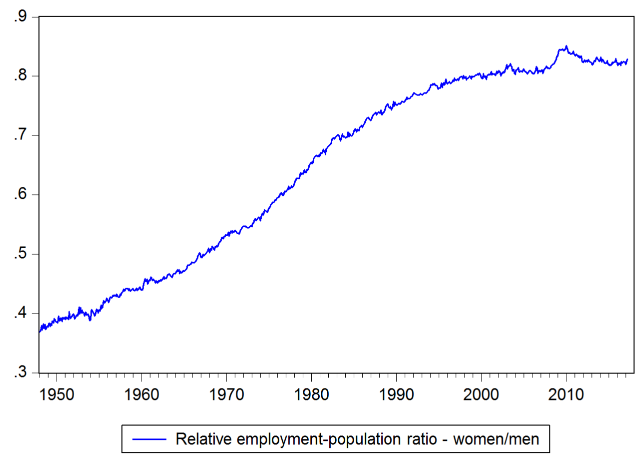
Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.011	0.011	0.0322	
		2	0.054	0.054	0.7743	
		3	-0.054	-0.056	1.5282	
		4	-0.032	-0.034	1.7972	0.180
		5	0.042	0.049	2.2539	0.324
		6	-0.091	-0.092	4.3953	0.222
		7	-0.081	-0.089	6.1091	0.191
		8	0.112	0.132	9.3965	0.094
		9	-0.073	-0.078	10.799	0.095
		10	-0.031	-0.065	11.048	0.137
		11	-0.041	-0.009	11.501	0.175
		12	-0.005	-0.001	11.507	0.243
		13	-0.013	-0.053	11.553	0.316
		14	-0.007	0.014	11.565	0.397
		15	0.006	0.021	11.575	0.480
		16	0.019	-0.024	11.670	0.555
		17	0.040	0.045	12.109	0.598
		18	-0.095	-0.096	14.580	0.482
		19	0.042	0.037	15.068	0.520
		20	0.070	0.082	16.416	0.495
		21	-0.041	-0.061	16.875	0.532
		22	0.010	-0.006	16.904	0.596
		23	-0.005	0.036	16.912	0.659
		24	0.158	0.146	23.933	0.296
		25	0.008	-0.038	23.952	0.350
		26	-0.075	-0.038	25.527	0.324
		27	-0.033	-0.021	25.837	0.362
		28	0.024	0.018	25.999	0.408
		29	-0.044	-0.047	26.564	0.432
		30	-0.043	-0.024	27.097	0.459
		31	-0.037	-0.000	27.498	0.491
		32	0.064	0.024	28.673	0.482
		33	0.022	0.028	28.816	0.527
		34	0.044	0.058	29.388	0.549
		35	-0.016	-0.004	29.459	0.596
		36	-0.071	-0.119	30.955	0.569

While coefficients are statistically significant in both AR(1) and AR(2)+S-AR(1) model, both the AIC and the BIC favor AR(2)+S-AR(1) model, and the correlograms of residuals shows that there is a serious problem with seasonality that is not accounted for in the AR(1). Overall, AR(2)+SAR(1) is a much better model for change in monthly private residential construction spending.

Question 8 (10 points)

Which deterministic trends would you use when developing models for the two series plotted below?



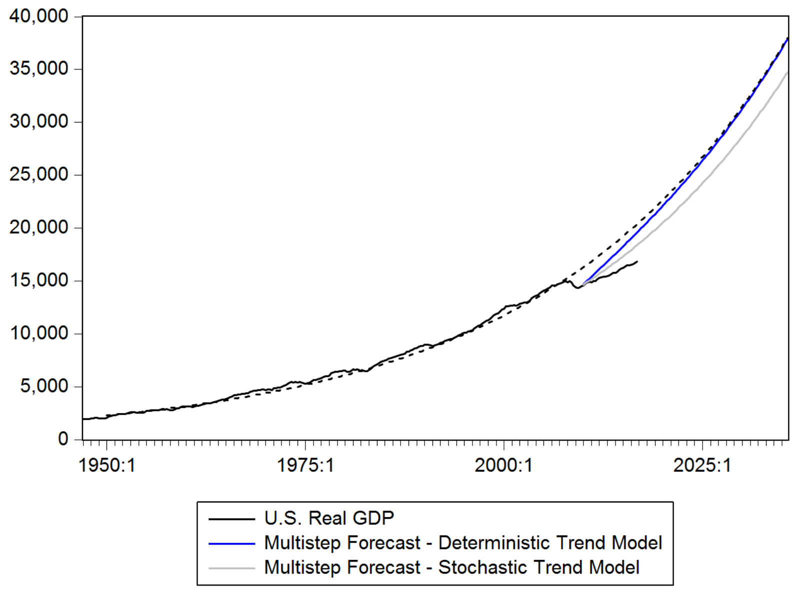


Exponential or logistic trend look like reasonable candidates to capture the time path of the average household size.

Logistic trend would be a good candidate to model relative employment-population ratio of women vs men.

Question 9 (10 points)

The following figure shows the multistep forecasts for the U.S. real GDP, from the deterministic model and from the stochastic trend model, both for the period 2010Q1-2035Q4. Discuss the main difference in the behavior of the two forecasts and explain the reason for this difference.



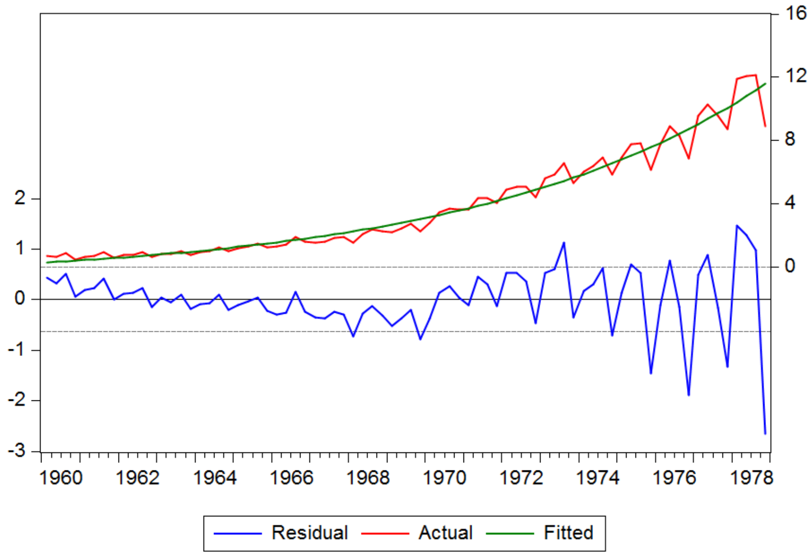
The effect of a one time shock is only temporary in the deterministic trend model, multistep forecast thus converges back to the original trend.
The effect of a one time shock is permanent in the stochastic trend model, multistep forecast thus does not converge back to the original trend, but rather stay permanently below the original trend.

Question 10 (10 points)

Consider a model for quarterly earnings per share of the Johnson and Johnson company

$$JNJ_t = \beta_0 + \beta_1 e^{\beta_2 t} + \varepsilon_t$$

Given the plot with actual values, fitted values, and residuals below, explain how you would proceed with modifying/developing the model further.



There are two problems with the estimated model:

1. it can match the trend, but not the seasonal pattern
2. the variance of residuals does not appear to be constant, it is increasing over time

To deal with the issue of variance of residuals increasing over time we reestimate the model using log transformed data $\log JNJ_t$ as the dependent variable instead of JNJ_t.

To deal with he seasonal pattern in the residuals seasonal AR or MA model needs to be specified, based on the correlogram of residuals.

Question 11 (10 points)

Below are the results for the Augmented Dickey-Fuller unit root test for log transformed earnings per share $\log JNJ_t$, and for the first difference of the log transformed earnings per share $\Delta \log JNJ_t$.

Interpret the results, and determine whether $\log JNJ_t$ is $I(0)$ or $I(1)$.

Null Hypothesis: LJNJ has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 3 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.696535	0.7428
Test critical values:	1% level	-4.090602	
	5% level	-3.473447	
	10% level	-3.163967	

*MacKinnon (1996) one-sided p-values.

Null Hypothesis: D(LJNJ) has a unit root
Exogenous: Constant, Linear Trend
Lag Length: 2 (Automatic - based on SIC, maxlag=11)

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-19.93554	0.0001
Test critical values:	1% level	-4.090602	
	5% level	-3.473447	
	10% level	-3.163967	

*MacKinnon (1996) one-sided p-values.

The p-value in the unit root test for $\log JNJ_t$ is 0.7428 we thus can not reject the hypothesis that $\log JNJ_t$ has a unit root.
The p-value in the unit root test for first difference $\Delta \log JNJ_t$ is 0.0001 we thus strongly reject the hypothesis that $\Delta \log JNJ_t$ has a unit root.
These two results combined imply that $\log JNJ_t$ is an I(1) time series - it is non-stationary but taking first difference transforms it into a stationary time series.

Question 12 (10 points)

Consider two models for U.S. real GDP, used to construct forecast for the period 2010Q1-2016Q4:

- model A: deterministic trend model for which the sequence of 1-step ahead forecasts has RMSE=103.45 and the multistep forecast has RMSE=1649.06
- model B: stochastic trend model for which the sequence of 1-step ahead forecasts has RMSE=77.32 and the multistep forecast has RMSE=905.18.

The 1-step ahead forecasts are then used to perform the test of equal predictive ability be estimating

$$\Delta L_{t+j,1} = \beta_0 + u_{t+j} \quad \text{with } j = 0, 1, 2, \ldots, T - t - 1$$

where $\Delta L_{t+j,1} = (e^A_{t+j,1})^2 - (e^B_{t+j,1})^2$, and $e^k_{t+j,1}$ is the one step ahead forecast error for forecast from model k in period $t + j$. Explain the idea behind this test and interpret its results below. Discuss how we would use it together with above RMSE values in model selection process.

Dependent Variable: DL_TREND
Method: Least Squares
Date: 04/09/17 Time: 18:34
Sample (adjusted): 2010Q1 2016Q4
Included observations: 28 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	5454.311	1293.939	4.215275	0.0002

R-squared	0.000000	Mean dependent var	5454.311
Adjusted R-squared	0.000000	S.D. dependent var	6846.884
S.E. of regression	6846.884	Akaike info criterion	20.53604
Sum squared resid	1.27E+09	Schwarz criterion	20.58361
Log likelihood	-286.5045	Hannan-Quinn criter.	20.55058
Durbin-Watson stat	2.683486		

The main idea behind the test of equal predictive ability is to test whether the difference in MSE or RMSE between two two competing forecasts is statistically significant or not. The null hypothesis of equal predictive ability of forecasts from two models A and B is that the difference in the MSE and RMSE is not statistically significant, and thus the estimated coefficient β_0 in the test regression is not statistically significant. If β_0 is statistically significant we reject the hypothesis of equal predictive ability of forecasts A and B. This is exactly the case here when comparing model A: a deterministic trend model for real GDP growth vs model B: stochastic trend model since the p-value for β_0 is only 0.0002. Model B is thus a much better model for real GDP growth rate forecasting.