Forecasting Time Series Homework 5

Yoon Tae Park[yp2201], Yeong Koh[yk2678], Yoobin Cheong[yc5206]

A) For the log of the Unemployment series, use the ACF, PACF and AIC_C to identify an ARIMA model (perhaps including a constant term).

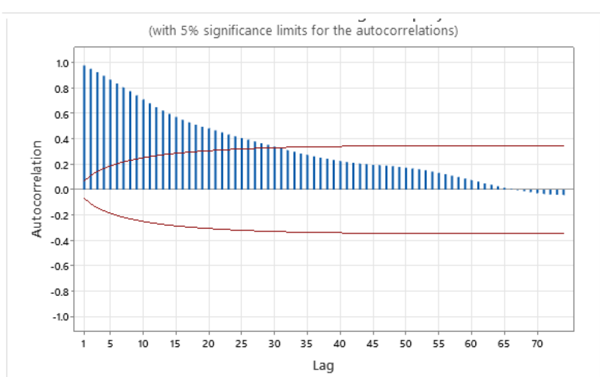
Ans) ARIMA (1,1,0) without constant

First, we checked the log of the Unemployment series, and didn't find any strong stationary or mean-reverting situation. Therefore, we took the 1st difference, and figured out that the difference of log Unemployment series is stationary and mean-reverting.

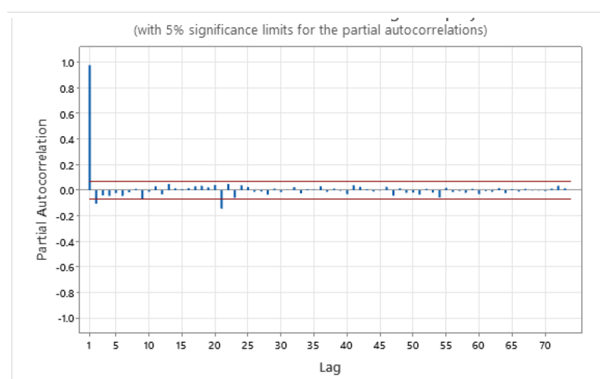
Moving on, we checked ACF and PACF of log Unemployment series to see any statistically significant lag point. Here, we found that there is statistically significant at lag 1 for both ACF and PACF. So, we created various combinations of ARIMA models and compared AICc. (p = 0, 1, 2/ d = 1 / q = 0,1, 2). We also compared the cases with a constant and without a constant.

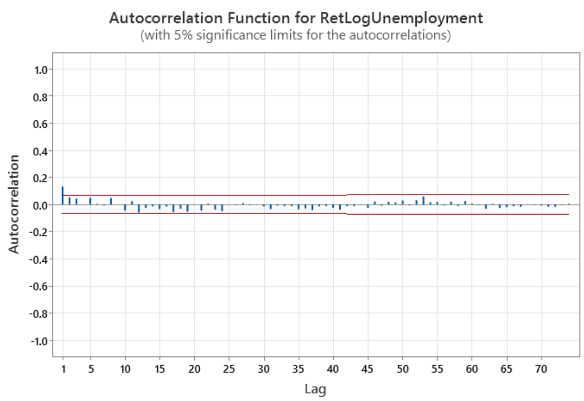
Based on the AICc table, we found that ARIMA (1,1,0) without constant shows the lowest AICc (-2563.84) among other ARIMA models. Therefore, we select ARIMA (1,1,0) without constant as the best ARIMA model.

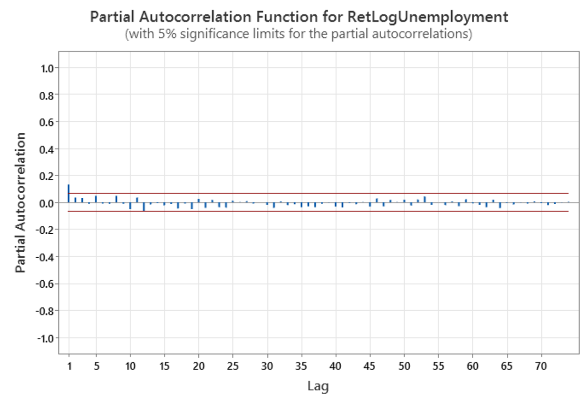
Autocorrelation Function: logUnemployment



Partial Autocorrelation Function: logUnemployment







[Minitab: Without constant]

Model Selection

Model (d = 1)	LogLikelihood	AICc	AIC	BIC
p = 1, q = 0*	1283.93	-2563.84	-2563.86	-2554.28
p = 1, q = 1	1284.78	-2563.54	-2563.57	-2549.21
p = 2, q = 0	1284.48	-2562.94	-2562.96	-2548.60
p = 0, q = 1	1283.36	-2562.71	-2562.72	-2553.15
p = 0, q = 2	1284.17	-2562.32	-2562.34	-2547.98
p = 1, q = 2	1284.96	-2561.87	-2561.91	-2542.77
p = 2, q = 1	1284.95	-2561.85	-2561.89	-2542.75
p = 2, q = 2	1285.41	-2560.75	-2560.82	-2536.88
p = 0, q = 0	1276.13	-2550.27	-2550.27	-2545.48

** Best model with minimum AICc. Output for the best model follows.*

[Minitab: With constant]

Model Selection

Model (d = 1)	LogLikelihood	AICc	AIC	BIC
p = 1, q = 0*	1283.93	-2561.84	-2561.87	-2547.51
p = 1, q = 1	1284.79	-2561.53	-2561.58	-2542.43
p = 2, q = 0	1284.49	-2560.93	-2560.97	-2541.83
p = 0, q = 1	1283.37	-2560.70	-2560.73	-2546.37
p = 0, q = 2	1284.18	-2560.31	-2560.35	-2541.21
p = 1, q = 2	1284.96	-2559.85	-2559.92	-2535.98
p = 2, q = 1	1284.95	-2559.83	-2559.90	-2535.97
p = 2, q = 2	1285.41	-2558.73	-2558.82	-2530.10
p = 0, q = 0	1276.14	-2548.27	-2548.28	-2538.71

** Best model with minimum AICc. Output for the best model follows.*

[Manual calculation using excel: shows the same result]

n		N	SS	p	d	q	AICc
887	wConstant	886		0	1	0	#NUM!
	wConstant	886	2.85901	1	1	0	-5076.2828
	wConstant	886	2.86272	0	1	1	-5075.1338
	wConstant	886	2.85343	1	1	1	-5075.9955
	wConstant	886	2.85536	2	1	0	-5075.3964
	wConstant	886	2.85739	0	1	2	-5074.7668
	wConstant	886	2.84935	2	1	2	-5073.2131
	woConstant	886	2.85904	1	1	0	-5078.2871
	woConstant	886	2.86276	0	1	1	-5077.1351
	woConstant	886	2.85346	1	1	1	-5078.0044
	woConstant	886	2.8554	2	1	0	-5077.4022
	woConstant	886	2.85743	0	1	2	-5076.7726
	woConstant	886	2.84937	2	1	2	-5075.2343

B) Estimate the parameters. Are they all statistically significant? Do you think that a constant should be included in the model?

Ans) AR 1 Coefficient shows statistically significant p-value. However, constant doesn't show statistically significant p-value(=0.913).

Regardless of the significance, we select the best model based on the AICc score. Therefore, based on the calculation on question A), the constant should not be included in the model as the best model is ARIMA (1,1,0) without the constant.

[Final model: ARIMA (1,1,0) w/o constant]

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.1326	0.0334	3.97	0.000

[For reference: ARIMA (1,1,0) w constant]

Final Estimates of Parameters

Type	Coef	SE Coef	T-Value	P-Value
AR 1	0.1326	0.0334	3.97	0.000
Constant	0.00021	0.00191	0.11	0.913

C) Write the complete form of the fitted model. (For example, $x_t = .3x_{t-1} + .2x_{t-2} + \varepsilon_t + .4\varepsilon_{t-1} - 2.351$).

Ans) $x_t = 0.1326x_{t-1} + \varepsilon_t$, $\quad$ *where* $x_t = diff(log\ unemployment)$

D) Examine the Ljung-Box statistics for lack of fit. Does the model seem to be adequate? The model is declared to be inadequate if "Chi-Square" in the Minitab output exceeds the 95'th percentile of a chi-square distribution with DF degrees of freedom, that is, if the corresponding p-value is less than .05.

Ans) The model seems to be adequate. We can see that all p-values are greater than 0.05 at all Chi-Square lag points.

Modified Box-Pierce (Ljung-Box) Chi-Square Statistic

Lag	12	24	36	48
Chi-Square	14.06	26.14	29.06	34.14
DF	11	23	35	47
P-Value	0.230	0.294	0.750	0.919

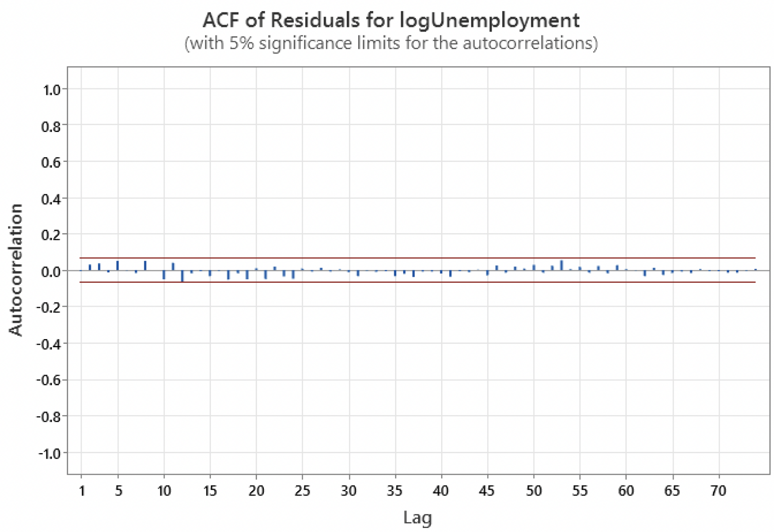
E) Plot the residuals from the fitted model, as well as the ACF and PACF of the residuals. Do these plots indicate any inadequacies in the model?

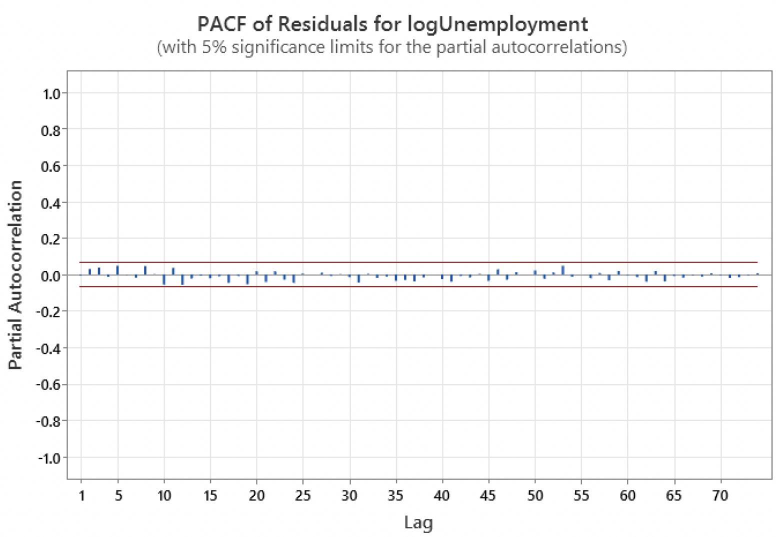
Ans) As shown in ACF and PACF of the residuals, we can see that there are no statistically significant lags. Therefore we can say that both ACF and PACF of the residuals don't have seasonality and therefore are considered as white noise.

Residual Sums of Squares

DF	SS	MS
885	2.85904	0.0032306

Back forecasts excluded





F) Obtain forecasts and 95% forecast intervals for lead times 1 to 30.

Ans)

Forecasts from period 887

		95% Limits		
Period	Forecast	Lower	Upper	Actual
888	1.42302	1.31160	1.53445	
889	1.42142	1.25307	1.58977	
890	1.42121	1.20963	1.63279	
891	1.42118	1.17368	1.66868	
892	1.42118	1.14233	1.70003	
893	1.42118	1.11416	1.72819	
894	1.42118	1.08837	1.75398	
895	1.42118	1.06444	1.77791	
896	1.42118	1.04202	1.80034	
897	1.42118	1.02085	1.82151	
898	1.42118	1.00074	1.84161	
899	1.42118	0.98156	1.86080	
900	1.42118	0.96318	1.87918	
901	1.42118	0.94550	1.89686	
902	1.42118	0.92846	1.91390	
903	1.42118	0.91199	1.93037	
904	1.42118	0.89604	1.94632	
905	1.42118	0.88056	1.96180	
906	1.42118	0.86550	1.97685	
907	1.42118	0.85085	1.99151	
908	1.42118	0.83656	2.00580	
909	1.42118	0.82261	2.01974	
910	1.42118	0.80899	2.03337	
911	1.42118	0.79565	2.04670	
912	1.42118	0.78260	2.05976	
913	1.42118	0.76981	2.07255	
914	1.42118	0.75726	2.08510	
915	1.42118	0.74495	2.09741	
916	1.42118	0.73286	2.10950	
917	1.42118	0.72097	2.12139	

G) Plot the data, forecasts and forecast intervals on a single plot.

Ans)

H) Do the forecasts seem reasonable? Do the forecast intervals seem excessively wide?

Ans)
The forecasts seem reasonable assuming the recent spike is an outlier/shock.
However, the 95% confidence forecast intervals seem excessively wide. Although the forecast intervals normally get wider as the lead time gets increased, this particular example shows excessively wide forecast intervals considering that we are only given relatively short lead time compared to the entire time the data is spread. For example, the forecast for lead time 30 includes the lower bound which is even lower than the all time low.