EC3304 Econometrics II

AY 2018/2019 Semester 2

Forecasting Project

Chua Ming Yuan Mervin

Tutorial Group: E5

1. Summary

This report uses Stata to perform statistical analysis to forecast the Consumer Price Index (CPI) of the monthly All Items in Singapore for the month February 2019.

The method of analysis is as follows. Firstly, a brief explanation of how the data is being formatted into the appropriate format, so as to ensure that we are able to carry out the statistical analysis in Stata. Secondly, we assessed whether the time series that we are working on is stationary or not stationary by plotting the variable All Items against time. Thirdly, we used the Akaike Information Criterion (AIC) to determine the number of lags required in our forecasting model. After acquiring the model, we next test for whether all the lags are significant in our model. Lastly, we then use the model we acquired to forecast the monthly (All Items) CPI for February 2019 and obtained a 95% forecast interval for it.

2. Our Model

In summary, we obtain our forecasting model to be:

$$\mathbf{Dlog(X)_t = \beta_0 + \beta_1 Dlog(X)_{t-1} + \beta_2 Dlog(X)_{t-2} + \beta_3 Dlog(X)_{t-3} + \ldots}$$

$$\mathbf{+\ \beta_{11} Dlog(X)_{t-11} + \beta_{12} Dlog(X)_{t-12},}$$

where X = All item and Dlog is the first log different between the time period t and t-1, which is the log different between this month and last month.

Detailed analysis and explanation will be provided at the descriptive of methods section to explain on how we derived at this forecasting model.

3.The sample period we used for our model

To obtain our forecasting model, the sampling period that we employed is from Jan 1961 to Jan 2019. The reason that we had chosen this period of data to forecast our model is because the first log different of the time series during this period is stationary, which mean that the time series is stable and that its distribution does not change over time. Hence, it is appropriate for us to use this sampling period to obtain our forecast model.

4. A description of any other data we had used

For our forecasting model, we only included the dependent variable All Items in it. As, since most of the other variables value are not available in the earlier part of the time series, if we include these variables, we will have to omit many of the missing values, causing our forecasting model to be inaccurate. Thus, to ensure that we had a more accurate model, only the dependent variable All Items is included. Yes, there is a tradeoff

5. A description of the methods we had used to evaluate our model

We first start off by transposing the given data in excel into the appropriate format before we import it into Stata. Since the date given was in a string form, we need to first convert it into a numeric form before we can apply the time series function to it. We created a new variable call Date by using the monthly function and format it accordingly as the picture below.

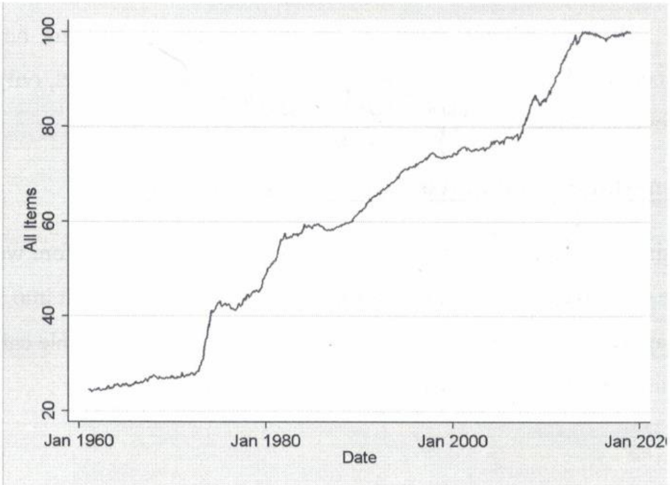
```
                                  College S
                                  800-STATA
                                  979-696-4
                                  979-696-4

100-user Stata network perpetual license:
       Serial number:  301406308777
         Licensed to:  NUS
                       NUS

Notes:
      1.  Unicode is supported; see help unicod
      2.  New update available; type -update al

. use "C:\Users\Mervin\Desktop\Proj.dta"

. import excel "C:\Users\Mervin\Desktop\Proj.xl

. generate Date = monthly(Variables,"YM")

. format %tmMon_CCYY Date
```

	Variables	AllItems	Date
1	1961 Jan	24.665	Jan 1961
2	1961 Feb	24.689	Feb 1961
3	1961 Mar	24.709	Mar 1961
4	1961 Apr	24.309	Apr 1961
5	1961 May	24.174	May 1961
6	1961 Jun	24.345	Jun 1961
7	1961 Jul	24.398	Jul 1961
8	1961 Aug	24.64	Aug 1961
9	1961 Sep	24.516	Sep 1961
10	1961 Oct	24.586	Oct 1961
11	1961 Nov	24.645	Nov 1961
12	1961 Dec	24.61	Dec 1961
13	1962 Jan	24.766	Jan 1962
14	1962 Feb	24.904	Feb 1962
15	1962 Mar	24.548	Mar 1962
16	1962 Apr	24.425	Apr 1962
17	1962 May	24.449	May 1962

After formatting the date, we can now apply the series function to the data. We then plot the All Items against Date to see if the time series is stationary or not.

```
tsset Date, monthly
      time variable:  Date, Jan 1961 to Jan 2019
              delta:  1 month
```

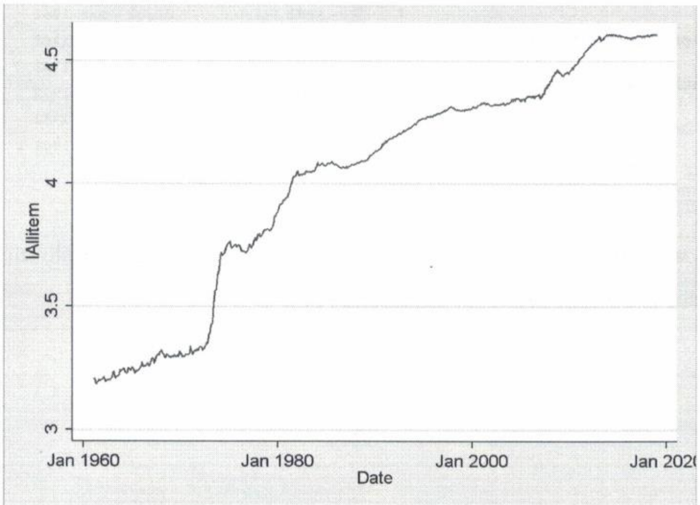
```
tsline AllItems
```



From the diagram above, we can see that there is an increasing trend in the consumer price index of All Items over the years and so the data is not stationary. Since the data is increasing and always positive, we can take the natural log of it to make the data stationary. Applying logarithm to All Items and plot it against date, we get the following graph.

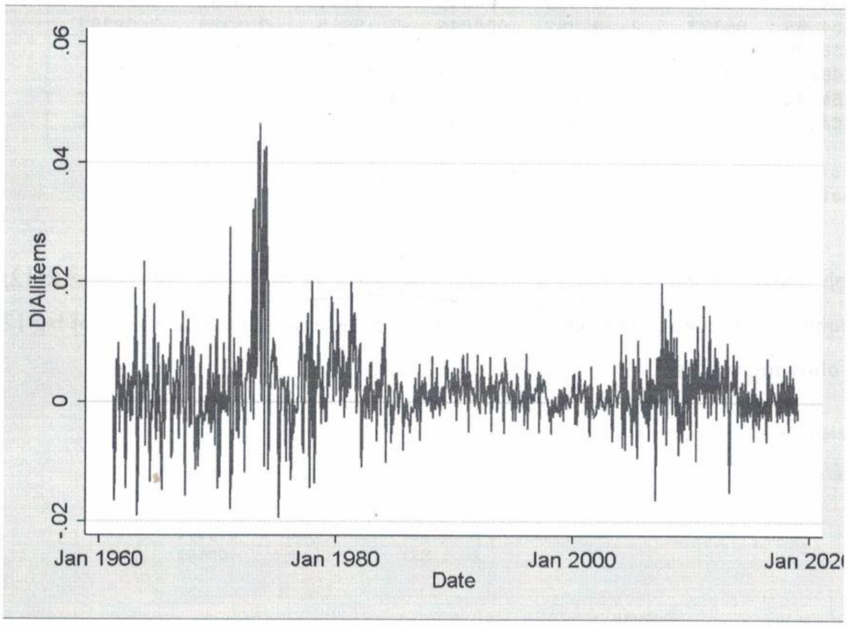
```
gen lAllItems= log(AllItems)
```

```
tsline lAllItems
```



Since the data still shows an increasing trend, we next take the first difference of Log(All Items) and plot it against Date and get the following graph.

```
. gen DlAllItems = D.lAllItems
(1 missing value generated)

. tsline DlAllItems
```



The time series is much more stationary since the values of DlAllitems only differ by a very small margin across the years.

Next, we used the Akaike Information Criterion (AIC) to determine the number of lags required in our forecasting model.

```
varsoc DlAllitems, maxlag(13)
```

Selection-order criteria
Sample: **Mar 1962 - Jan 2019** Number of obs = **683**

lag	LL	LR	df	p	FPE	AIC	HQIC	SBIC
0	2408.87				.000051	-7.05086	-7.0483	-7.04423
1	2417.29	16.84	1	0.000	.00005	-7.07259	-7.06746	-7.05933
2	2419.74	4.9089	1	0.027	.000049	-7.07685	-7.06915	-7.05697
3	2431.69	23.899	1	0.000	.000048	-7.10891	-7.09865	-7.0824
4	2434.55	5.7174	1	0.017	.000048	-7.11435	-7.10153	-7.08122
5	2442.09	15.086	1	0.000	.000047	-7.13351	-7.11812	-7.09375
6	2453.75	23.308	1	0.000	.000045	-7.16471	-7.14676*	-7.11832*
7	2454.49	1.4877	1	0.223	.000045	-7.16396	-7.14344	-7.11094
8	2454.81	.64272	1	0.423	.000045	-7.16197	-7.13889	-7.10233
9	2454.85	.06973	1	0.792	.000046	-7.15915	-7.1335	-7.09287
10	2458.98	8.2583	1	0.004	.000045	-7.16831	-7.1401	-7.09541
11	2461.4	4.8495	1	0.028	.000045	-7.17248	-7.1417	-7.09295
12	2464.43	6.0616*	1	0.014	.000045*	-7.17843*	-7.14509	-7.09227
13	2464.43	.00014	1	0.990	.000045	-7.1755	-7.13959	-7.08272

Endogenous: DlAllitems
Exogenous: _cons

From the table above, we can see that AIC is minimize at lag 12 and hence we chose AR(12) as the most appropriate model to forecast the data. We run the auto regression model of lag 12 and get the following results.

```
. reg DlAllitems L(1/12).DlAllitems ,r
```

Linear regression

Number of obs	=	684
F(12, 671)	=	3.32
Prob > F	=	0.0001
R-squared	=	0.1504
Root MSE	=	.00662

DlAllitems	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
DlAllitems						
L1.	.0604164	.0666652	0.91	0.365	-.0704811	.1913139
L2.	.0174347	.0683412	0.26	0.799	-.1167537	.151623
L3.	.1332162	.0654844	2.03	0.042	.0046372	.2617953
L4.	.0758289	.0615502	1.23	0.218	-.0450253	.1966832
L5.	.1191923	.0723346	1.65	0.100	-.022837	.2612216
L6.	.1597339	.0649221	2.46	0.014	.032259	.2872089
L7.	.0398477	.0595327	0.67	0.504	-.077045	.1567404
L8.	.0139495	.0522883	0.27	0.790	-.0887189	.1166179
L9.	-.0163699	.0541485	-0.30	0.763	-.1226907	.089951
L10.	-.115329	.0541234	-2.13	0.033	-.2216006	-.0090574
L11.	.0769739	.053781	1.43	0.153	-.0286255	.1825732
L12.	.0932869	.0506305	1.84	0.066	-.0061265	.1927003
_cons	.0006867	.0003376	2.03	0.042	.0000238	.0013496

H_0: $\beta_i = 0$ for i in {1, 2, 3 .., 12}
H_1: $\beta_i \neq 0$ for some i in {1, 2 ,3 ..,12}

Since p-value=0.0001<0.01, we reject H_0 at the 1% level and hence the model is significant at the 1% level. We next test for whether all the lags are significant in the model. We run the test in Stata and get the following results.

```
. test L1.DlAllitems L2.DlAllitems L3.DlAllitems L4.DlAllitems L5.DlAllitems L6.DlAllitems L7.DlAll
> items L8.DlAllitems L9.DlAllitems L10.DlAllitems L11.DlAllitems L12.DlAllitems

 ( 1)  L.DlAllitems = 0
 ( 2)  L2.DlAllitems = 0
 ( 3)  L3.DlAllitems = 0
 ( 4)  L4.DlAllitems = 0
 ( 5)  L5.DlAllitems = 0
 ( 6)  L6.DlAllitems = 0
 ( 7)  L7.DlAllitems = 0
 ( 8)  L8.DlAllitems = 0
 ( 9)  L9.DlAllitems = 0
 (10)  L10.DlAllitems = 0
 (11)  L11.DlAllitems = 0
 (12)  L12.DlAllitems = 0

       F( 12,   671) =    3.32
            Prob > F =    0.0001
```

Since the p-value=0.0001<0.01, it means that all the 12 lags are significant in model at the 1% level. Therefore, our forecasting model will be:

$$\mathbf{Dlog(X)_t = \beta_0 + \beta_1 Dlog(X)_{t-1} + \beta_2 Dlog(X)_{t-2} + \beta_3 Dlog(X)_{t-3} + \ldots}$$

$$\mathbf{+ \beta_{11} Dlog(X)_{t-11} + \beta_{12} Dlog(X)_{t-12},}$$

where X = All item and Dlog is the first log different between the time period t and t-1, which is the log different between this month and last month.

6. Forecast for 2019M2 & 95% Forecast Interval

To forecast the consumer price index of All Items for February 2019, we append the time series by another month and predict the Dlog value of it.

```
tsappend ,add(1)

predict Dlhat
```

	Variables	AllItems	Date	lAllitem	DlAllitems	Dlhat
690	2018 Jun	100.006	Jun 2018	4.60523	.0005703	.0001766
691	2018 Jul	99.859	Jul 2018	4.603759	-.001471	.0004535
692	2018 Aug	100.28	Aug 2018	4.607966	.0042071	.0020296
693	2018 Sep	100.255	Sep 2018	4.607717	-.0002494	-.0001535
694	2018 Oct	99.947	Oct 2018	4.60464	-.003077	.0008864
695	2018 Nov	100.111	Nov 2018	4.606279	.0016394	.0023977
696	2018 Dec	100.187	Dec 2018	4.607038	.0007591	.0003687
697	2019 Jan	99.919	Jan 2019	4.60436	-.0026789	.0012018
698		.	Feb 2019	.	.	.0018364

From the table above, we can see that forecasted value at Feb 2019 is Dlhat = 0.0018364. Nonetheless, this forecasted value is the difference between Feb 2019 and Jan 2019 that is Dlhat = Feb 2019 – Jan 2019 = 0.0018364. So, to get Log (All items) in Feb 2019, we take 0.0018364 + Log(All items) in Jan 2019 and we will get 4.60436+0.0018364=4.6061964. Finally, to get the forecasted value of the consumer price index of All items in February 2019, we just need to exponential back the value.

Therefore, the Forecasted Value in Feb 2019 = Exp (4.6061964)

= 100.1026741

= 100.103

95 % Forecast Confidence Interval: Exp ([Dlhat +- 1.96*root MSE] + $\text{Log(Allitems)}_{\text{Jan2019}}$)

= Exp ([0.0018364 +- 1.96*0.00662] + 4.60436)

= (98.81221196, 101.4099893)

= (98.8, 101.4)