Tourists Visiting
the United States of America

Swarup Sahoo
Chih-Hsiang Hsieh
Xintong Zhou

Term-Project for

Time-Series Analysis, Modeling and Control

- Spring '17

05/09/2017

1. Introduction

Tourism in the United States is a large industry that serves millions of international and domestic tourists. We collected the monthly international visitor data from the National Travel and Tourism Office (NTTO) database and analyzed the total inflow and country wise inflow. The 9/11 attack affected the numbers a by a big order of magnitude and it took some time before the numbers grew up to normal again. So our analysis starts from January of 2002.

Our project mainly focuses on four sections. The first section is the analysis of the data structure and transformation to facilitate modeling. The second section is stationary modeling of the total number people from all other countries to the US. The third section is modeling using the non-stationary series. In the last section, we combined the three typical countries along with the total countries' data to set up the Vectorial ARMA model. The predictions are plotted and compared to show the performance of the models with data fitting and forecasting. The entire analysis has been performed in R programming language. Certain packaged statistical tools have been used for simplicity of analysis.

2. Data Processing

We plotted the monthly data and observed that the data is heteroscedastic [1],as shown below. The variances were blowing up, which is not a suitable condition to fit an ARMA model.

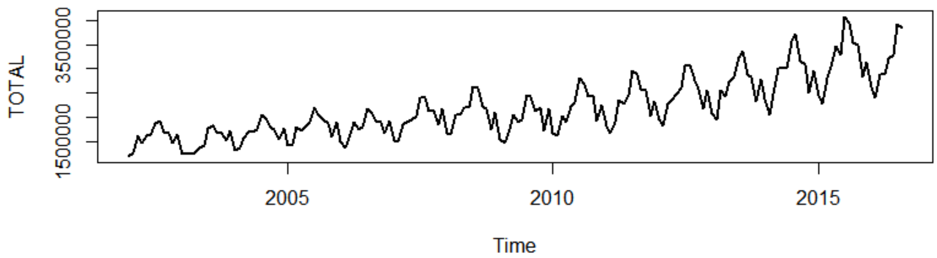


Fig.2.1.Total data plot(Time Span: Jan-2002 to Aug-2016)

A *log* transformation [1] gives more controlled variances, suitable for modeling and analysis.

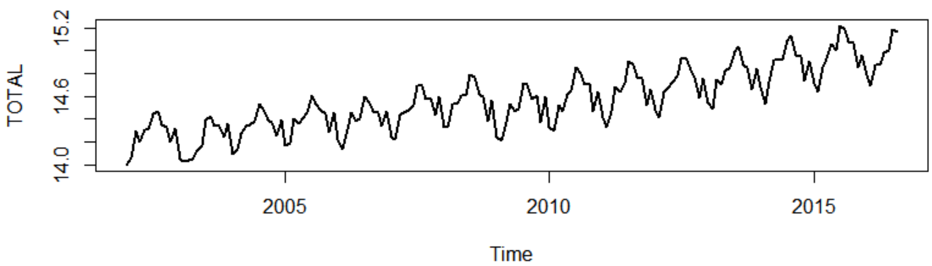


Fig.2.2.Data plot after log transformation(Time Span: Jan-2002 to Aug-2016)

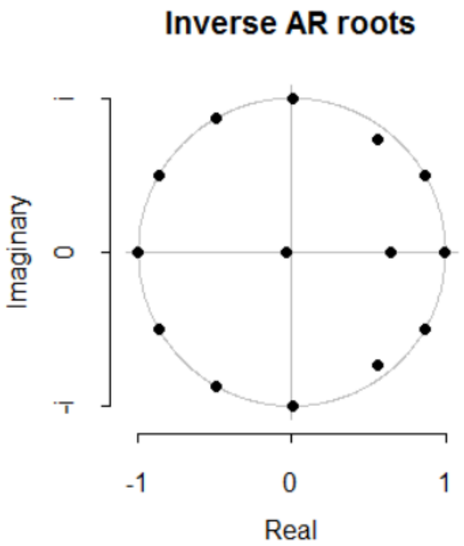
We will be using the log transformed series for modeling. For forecasts we will reverse transform the data, since we need to test it against the test set, which is untransformed. For all modeling steps we have considered the data from January 2002 to October 2015 (166 data points) as our training set and the data from November 2015 to August 2016 (10 data points) as the testing set.

3. Stationary Times Series Modeling

First, we assumed that the data is stationary. By using R, we wrote the function to do the F-test. In the function we used an ARMA(2n,2n-1) incremental method. The function *Arima* was used from the *fpp* package, with parameter *order=c(2n,0,2n-1)*. The F value for the (2n,2n-1) vs (2(n+1),2(n+1)-1) was compared against the F(4,166) at 95% confidence interval (*qf(.95, df1=4, df2=N)*) and found the best-fit model. In the end, we got an ARMA(14,13), which is the adequate model. (Appendix 1)

The summary of the ARMA model would then show the coefficients. RSS = 0.28314064.

As for the roots, we used the *polyroots* function to calculate the roots. The following shows the roots on the unit circle. All the roots lie either inside or on the unit circle.



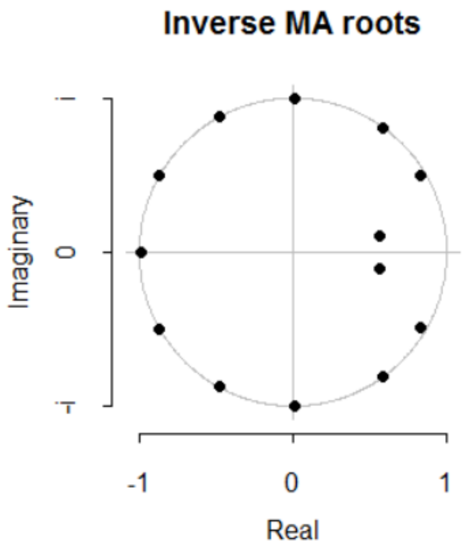


Fig 3.1. Inverse AR and MA roots for ARMA(14,13)

In the model, it was found that λ_1 to λ_8 are seasonal parts(Appendix 2), which show periodicity = 12 and their harmonics. λ_1 is the real root of -1, and it also shows a seasonal but not in sinusoidal form but rather in a zig-zag pattern.[6] The following is the model in Parsimonious form:

$$(1-B)(1+B)(1+1.729B+B^2)(1+0.986B+B^2)(1-1.729B+B^2)(1-0.009B+B^2)$$

(1-B) is the trend part (because λ=1), and
(1+B) $(1+1.729B+B^2)$ $(1+0.986B+B^2)$ $(1-1.73B+B^2)$ $(1-0.009B+B^2)$ is the seasonal part (Appendix 3).
RSS of the model = 0.2495025.

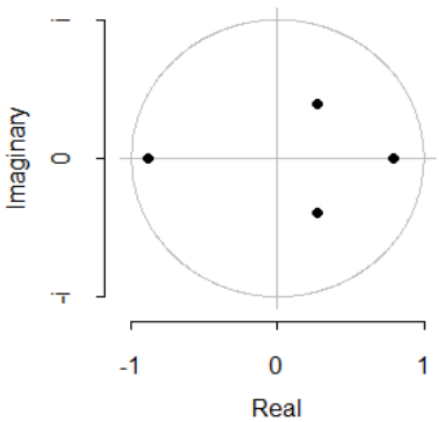


Fig 3.2.Plot of AR roots for parsimonious model

In the last part of this section, we did an ARMA forecast [4] from Nov 2015 to Oct 2016, and compared it with the actual data we kept aside for testing. For plotting the data in the proper scale, we transformed all the data points to an exponential order. The test set data was used to judge that if the model is suitable or not at a 95% confidence interval. The forecast shows that the model is a good fit. But two actual data points lie close to the 95% boundary, so maybe we can find a better model.

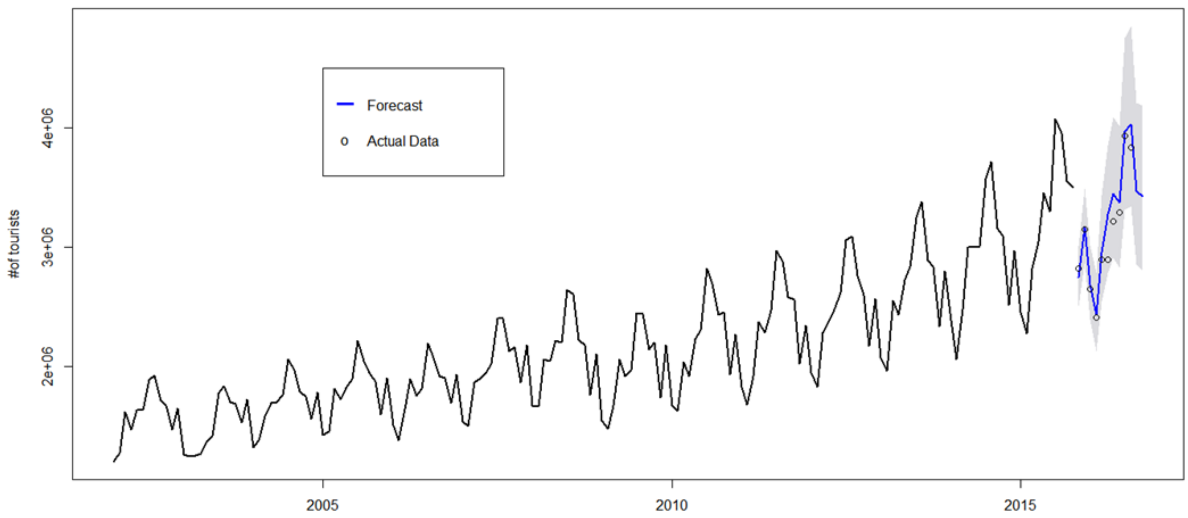


Fig 3.3. Forecast of ARMA(14,13) model

Both the parsimonious and the original models would give practically the same forecasts and probability limits.[5]

We performed the Ljung-Box test (*tsdiag(model)*) to test the residuals of ARMA(14,13) model, to check whether it is the white noise. The null hypothesis is that it is white noise. Since the p-values for ten lags are comfortably greater than 5%, hence we can conclude that the residuals are uncorrelated. The same is shown by the ACF plots. [2]

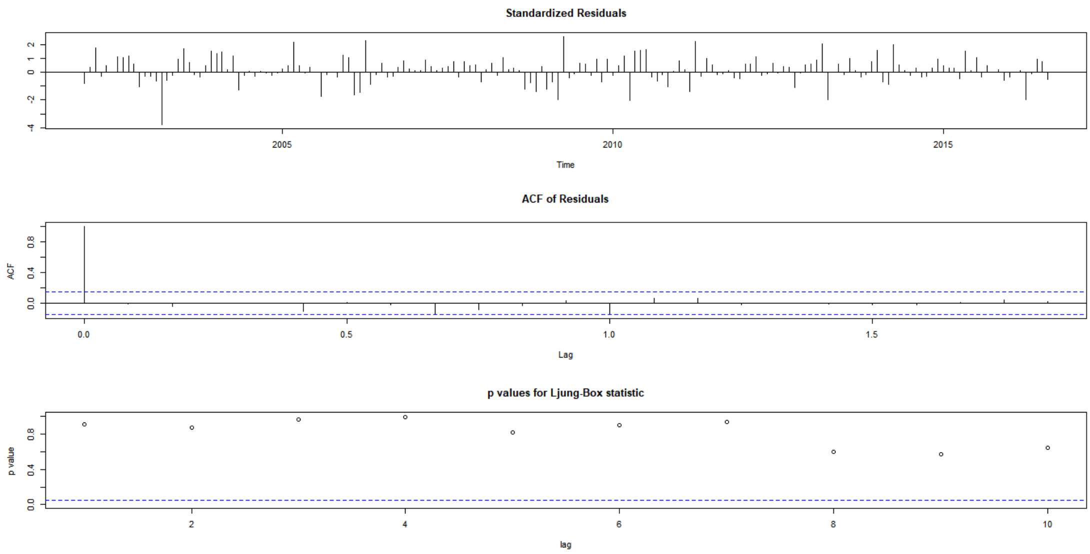


Fig.3.4. Check for whiteness of residuals for Stationary ARMA(14,13)

Explanation on Ljung-Box Test is at the end of section 4.

4. Non-stationary Time Series Model

Now, we try an alternative method of modeling obtained by decomposing a model into two parts, deterministic and stochastic. From the data, we observed the periodic trend easily. By using:

$$Y_t = kt + a \sin \omega t + b \cos \omega t \qquad \omega = 2\pi/12$$
$$k = 0.00438, \qquad a = -0.1248, \qquad b = -0.1137$$

We fitted the periodic trend. The trend was removed from the data and we followed the same F-test procedure as before. The best fit non-stationary model was found to be an ARMA(14,13). The roots of the model are shown as Fig 4.1, and the RSS of the model = 0.233169 (Appendix 4).

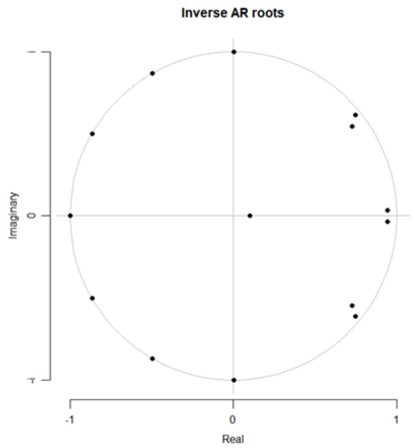


Fig 4.1. AR roots for non-stationary ARMA(14,13) model

For ARMA forecasting [4] we added the previously removed trend curve, back to the model. Further we extrapolated the trend for 12 future data points and added the forecasted value on top of that. The uncertainty intervals were also added to the forecast. Finally all the data points were transformed to exponentials to match with the scale of the actual test set data.

In Fig 4.2, black dots show the actual data from Nov 2015 to Aug 2016, and blue line is our prediction. It is shown that the prediction result is close to the actual data. The RSS and AIC[3] results also show that the non-stationary is the best model in our project.

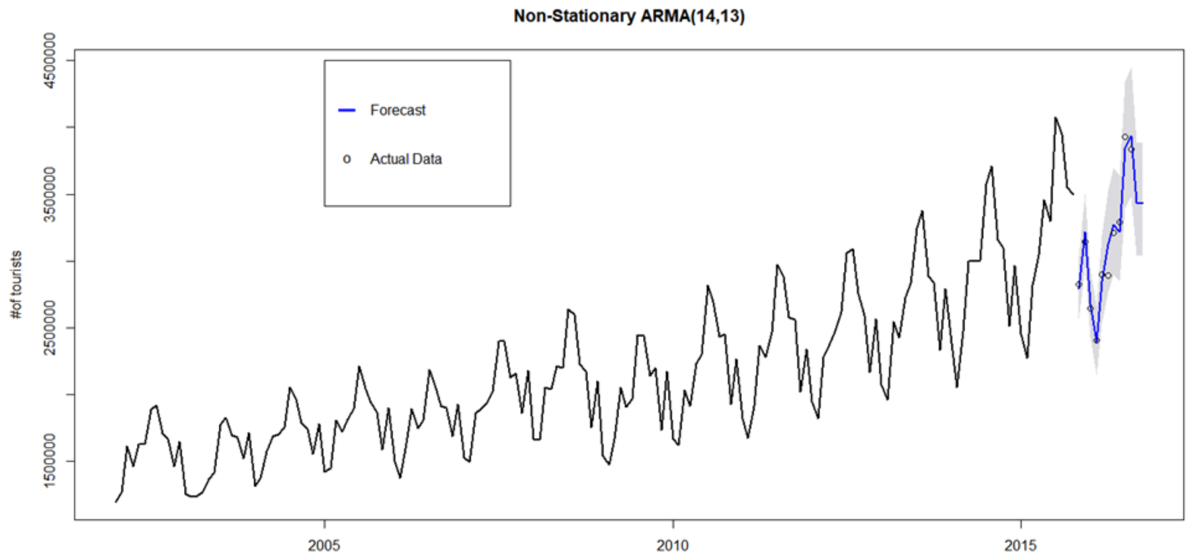


Figure 4.2. Forecast of non-stationary ARMA(14,13) model

The Ljung-Box test on the residuals of non-stationary model shows that the residuals are just white noise. [2] Hence, the modeling is complete for this section.

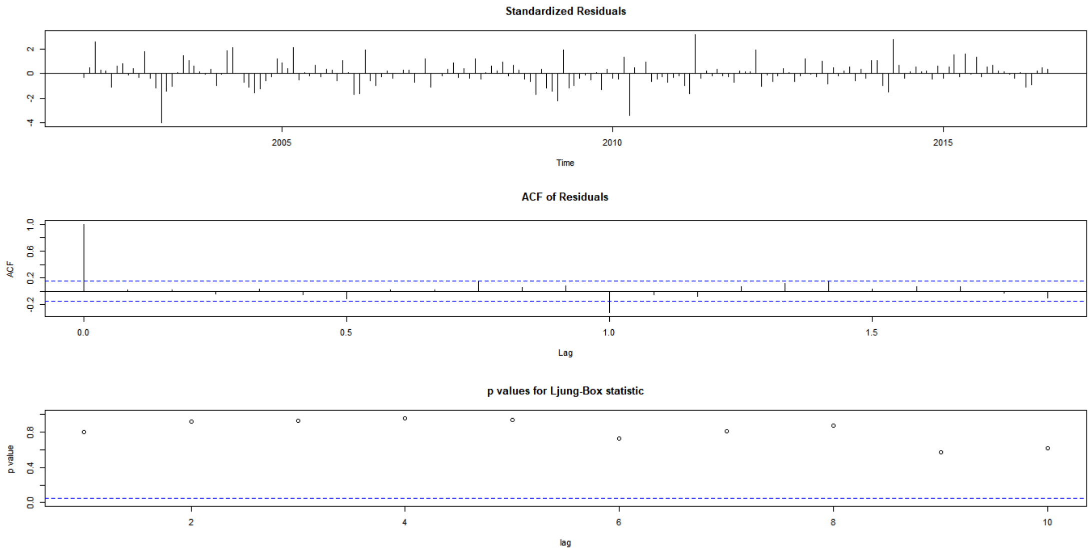


Fig 4.3. Check for whiteness of residuals (non-stationary model) ARMA(14,13)

Ljung-Box test used for white noise check of residuals. The test statistic is: $Q = n(n+2)\sum_{k=1}^{h}\frac{\hat{\rho}_k^2}{n-k}$ where n is the sample size, $\hat{\rho}_k^2$ is the sample autocorrelation at lag k, and h is the number of lags being tested. Under H_0 the statistic Q follows a $\chi^2_{(h)}$.

For significance level α, the critical region for rejection of the hypothesis of randomness is $Q > \chi^2_{1-\alpha,h}$

The blue dotted lines in ACF plot signify the 95% confidence interval of the distribution of white noise, i.e. $\pm 1.96\sigma/\sqrt{n}$.

5. Vectorial ARMA modeling

We picked different countries from different continents. Information of tourists' data from Germany, Australia and South Korea were found to be most correlated with the total tourist inflow. We used these data sets to prepare a Vectorial ARMA model.[6] For this purpose we used the MTS package in R.[7]

A Vectorial ARMA(2n,2n-1) incremental modeling with F-test method was used to find the best fit. The function *VARMA* of the *MTS* package was used for this loop. An ARMAV(4,3) was found to be the best fit. Appendix 5 and 6 show the coefficient matrix of ϕ and θ.

The covariance matrix is as shown below. It shows that the correlation between different countries and total data. The variables tend to show similar behavior, since the covariances are positive.

	TOTAL	GERMANY	SOUTH KOREA	AUSTRALIA
TOTAL	0.004477493	0.006089834	0.001548936	0.002940034
GERMANY	0.006089834	0.012151337	0.002053682	0.003839761
SOUTH KOREA	0.001548936	0.002053682	0.007620268	0.000681619
AUSTRALIA	0.002940034	0.003839761	0.000681619	0.006104997

Table 5.1 Covariance matrix of ARMAV model

The predictions of the Total, Germany, South Korea and Australia are shown as Fig 5.1~5.4., from the predictions of ARMAV(4,3) model.

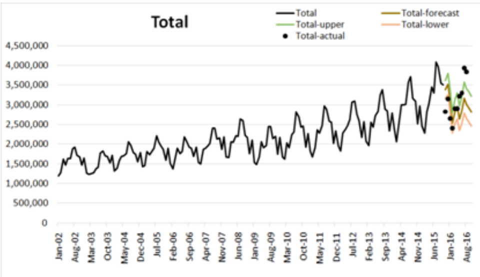


Fig.5.1. Total Arrival forecast(ARMAV)

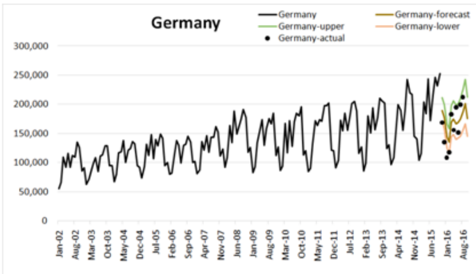


Fig.5.2. German Arrival forecast(ARMAV)

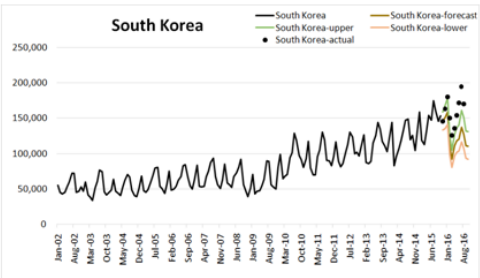


Fig.5.3. South Korean Arrival forecast(ARMAV)

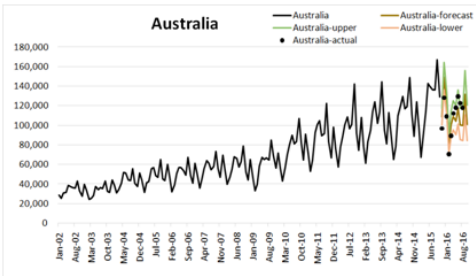


Fig.5.4. Australian Arrival forecast(ARMAV)

From the forecasts we can see the the seasonal effects are damping out very fast [11] and the actual data is far outside the confidence interval. The forecast is converging towards the mean. Hence, the data sets in consideration are not good explanatory variables for the forecast of tourist inflow.

6. Conclusion

To sum up, the non-stationary time series model gives us the best prediction for the tourists traveling to US, since the actual data points lie close to the mean forecast and further the model has the lowest RSS and AIC.

Model		RSS	AIC [3]
Stationary	ARMA(14,13)	0.2831	-484.28
Parsimonious	ARMA(4,13)	0.2495	-498.95
Nonstationary	ARMA(14,13)	0.2332	-520.19
Vectorial	ARMAV(4,3)	0.7235	-20

Table 6.1 Comparison of all the models

Further, we noticed that all the models accurately accounted for the long range period drop in tourist inflow. The numbers tend to drop every 7.5 years because of some major event. The terrorist attack in mid 2001, economic depression in early 2009 and probably the US elections in late 2016 accounted for a bad international perception of United States and hence the drops in tourist inflow.

It is quite interesting to analyze the tourism in US that indicates not only the development of tourism but also the impact of economic and political events on international tourists' perception. Since, these factors affect tourism, the analysis of the economic and political data of USA and other major countries can provide a more robust vectorial ARMA model for forecasting.

References

1. Mann, HB; Wald, A. “On the Statistical Treatment of Linear Stochastic Difference Equations”. Econometrica 1943. 11: 173–200
2. Ljung, GM; Box, GEP. “On a measure of lack of fit in time series models”. Biometrika 1978. 65 (2): 297-303.
3. Bozdogan, H. “Model Selection And Akaike’s Information Criterion (AIC): The General Theory And Its Analytical Extensions”. Psychometrika 1987. 52 (3): 345-37.
4. Torresa, JL; Garcíaa, AM. "Forecast of hourly average wind speed with ARMA models in Navarre". Solar Energy 2005. 79(1), 65–77.
5. Anastasakis, LN. "Prediction of the GBPUSD Exchange Rate using Statistical and Neural Network Models". Proceeding (403) Artificial Intelligence and Applications 2003. 777-780.
6. Patrik, G; Jonas, H. "The Impact of Seasonal Unit Roots and Vector ARMA Modelling on Forecasting Monthly Tourism Flows". Sage Journals, 2001, 117–133.
7. Ruey, S; Tsay, GC. "Consistent Estimates of Autoregressive Parameters and Extended Sample Autocorrelation Function for Stationary and Nonstationary ARMA Models" Journal of the American Statistical Association 2012. 84-96.
8. Pandit, SM; Wu, SM. "Time Series And System Analysis With Applications", 1983
9. Taylor, AMR. "On the practical problems of computing seasonal unit root tests", International Journal of Forecasting, Vol 13, 1997, 307–318.
10. Taylor, AMR. "Testing for unit roots in monthly time series", Journal of Time Series Analysis, Vol 19, 1998, 349–368.
11. Oke, T; Öller, LE. "Testing for Short Memory in a VARMA Process, National Institute of Economic Research", Stockholm, Working Paper No 56, 1997.
12. Tiao, G; Box, GEP. "Modeling multiple time series with applications", Journal of the American Statistical Association, Vol 76, 1981, 802–816.
13. Lütkephol, H. "Introduction to Multiple Time Series Analysis", Springer-Verlag, Berlin 1991. 232–233.

Appendix 1: Stationary Analysis of data. Incremental model statistics

Parameters	ARMA order								Seasonal & trend parts
	(2,1)	(4,3)	(6,5)	(8,7)	(10,9)	(12,11)	(14,13)	(16,15)	(4,13)
φ_1	0.730 ± 0.558	0.912 ± 0.247	-0.226 ± 0.257	-0.019 ± 0.103	-0.394 ± 0.549	-0.448 ± 0.401	0.744 ± 0.731	0.717 ± 1.278	0.464 ± 0.524
φ_2	0.175 ± 0.496	0.191 ± 0.398	0.427 ± 0.264	1.153 ± 0.097	-0.008 ± 0.038	0.172 ± 0.367	0.186 ± 0.402	-0.11 ± 2.320	0.521 ± 0.468
φ_3		-0.346 ± 0.314	0.394 ± 0.227	0.038 ± 0.205	0.981 ± 0.027	0.635 ± 0.330	-0.005 ± 0.222	0.018 ± 1.490	-0.408 ± 0.512
φ_4		0.042 ± 0.175	-0.002 ± 0.279	-0.331 ± 0.190	0.394 ± 0.528	0.461 ± 0.389	-0.201 ± 0.171	-0.17 ± 0.923	0.161 ± 0.293
φ_5			0.207 ± 0.249	-0.038 ± 0.205	0.010 ± 0.011	-0.177 ± 0.362	-0.204 ± 0.115	-0.094	
φ_6			-0.379 ± 0.205	-0.660 ± 0.190	-0.985 ± 0.014	-0.641 ± 0.328	0.006 ± 0.233	0.013 ± 0.132	
φ_7				0.016 ± 0.104	-0.393 ± 0.546	-0.457 ± 0.390	0.211 ± 0.172	0.108 ± 0.118	
φ_8				0.827 ± 0.095	-0.010 ± 0.036	0.172 ± 0.362	0.192 ± 0.118	0.082	
φ_9					0.986 ± 0.027	0.643 ± 0.327	0.001 ± 0.226	-0.00 ± 0.132	
φ_{10}					0.414 ± 0.531	0.482 ± 0.392	-0.194 ± 0.181	-0.08 ± 0.098	
φ_{11}						-0.188 ± 0.369	-0.238 ± 0.111	-0.123	
φ_{12}						0.341 ± 0.335	1.016 ± 0.248	1.027 ± 0.143	
φ_{13}							-0.506 ± 0.601	-0.59 ± 1.390	
φ_{14}							-0.018 ± 0.355	0.173 ± 2.312	
φ_{15}								-0.02 ± 1.414	
φ_{16}								0.074 ± 0.908	
θ_1	0.001 ± 0.555	-0.399 ± 0.207	1.070 ± 0.271	0.336 ± 0.157	0.758 ± 0.516	0.924 ± 0.332	-0.297 ± 0.703	-0.26 ± 1.208	0.921 ± 0.500
θ_2		0.084 ± 0.238	0.610 ± 0.360	-0.973 ± 0.158	0.790 ± 0.262	0.732 ± 0.427	0.149 ± 0.476	0.449 ± 1.709	-0.232 ± 0.603
θ_3		0.777 ± 0.214	0.673 ± 0.331	0.351 ± 0.217	-0.089 ± 0.508	0.429 ± 0.462	0.200 ± 0.343	0.393 ± 1.068	-0.529 ± 0.431
θ_4			1.045 ± 0.216	0.978 ± 0.167	0.055 ± 0.105	0.351 ± 0.464	0.267 ± 0.393	0.393 ± 0.729	-0.388 ± 0.593
θ_5			0.283 ± 0.309	-0.393 ± 0.223	-0.028 ± 0.122	0.557 ± 0.321	0.276 ± 0.477	0.398 ± 0.601	0.142 ± 0.521
θ_6				-0.044 ± 0.156	0.860 ± 0.146	0.810 ± 0.191	0.144 ± 0.574	0.301 ± 0.706	0.471 ± 0.365
θ_7				0.627 ± 0.134	0.638 ± 0.456	0.830 ± 0.351	-0.011 ± 0.430	0.176 ± 0.606	0.219 ± 0.532
θ_8					0.692 ± 0.253	0.625 ± 0.407	0.078 ± 0.460	0.263 ± 0.567	-0.251 ± 0.426
θ_9					-0.197 ± 0.433	0.317 ± 0.410	0.098 ± 0.267	0.256 ± 0.427	-0.481 ± 0.339
θ_{10}						0.273 ± 0.413	0.273 ± 0.284	0.299 ± 0.519	-0.273 ± 0.556
θ_{11}						0.616 ± 0.309	0.340 ± 0.381	0.405 ± 0.551	0.332 ± 0.502
θ_{12}							-0.809 ± 0.537	-0.67 ± 0.430	-0.319 ± 0.338
θ_{13}							0.298 ± 0.401	0.493 ± 1.589	-0.598 ± 0.347
θ_{14}								-0.083 ± 1.407	
θ_{15}								-0.119 ± 1.374	
RSS	2.83718	1.82057	1.35687	0.56014	0.34617	0.3037	0.283068	0.279557	0.249502
F		23.17	14.18	59.03	25.65	5.80	3.02	0.52	
AIC	-192.79	-250.19	-290.14	-415.66	-475.82	-483.99	-484.28	-477.08	-498.95

Appendix 2: The roots for the ARMA(14,13) model

Root	Real	Imag	Absolute Values	Periodicity
λ_1	0.865	0.501	1.00	12.0
λ_2	0.865	-0.501	1.00	12.0
λ_3	0.005	1.000	1.00	4.0
λ_4	0.005	1.000	1.00	4.0
λ_5	-0.494	0.869	1.00	3.0
λ_6	-0.494	-0.869	1.00	3.0
λ_7	-0.865	0.502	1.00	2.4
λ_8	-0.865	-0.502	1.00	2.4
λ_9	-0.035	0.000	0.03	-
λ_{10}	0.648	0.000	0.65	-
λ_{11}	-0.999	0.000	1.00	-
λ_{12}	0.555	0.729	0.92	-
λ_{13}	0.555	-0.729	0.92	-
λ_{14}	0.997	0.000	1.00	-

Appendix 3: The roots for parsimonious model

Root	Real	Imag	Absolute Values	Periodicity
λ_1	0.276	0.391	0.48	-
λ_2	0.276	-0.391	0.48	-
λ_3	-0.883	0.000	0.88	-
λ_4	0.796	0.000	0.80	-

Appendix 4: Non-stationary Analysis of data. Incremental model statistics

Parameters	ARMA order							
	(2,1)	(4,3)	(6,5)	(8,7)	(10,9)	(12,11)	(14,13)	(16,15)
φ1	0.724 ± 0.234	0.371 ± 0.189	0.092 ± 0.144	0.311 ± 0.327	-0.043 ± 0.168	-0.070 ± 0.404	1.231 ± 0.129	1.18 ± 0.736
φ2	0.158 ± 0.164	-0.537 ± 0.147	0.493 ± 0.136	-0.030 ± 0.156	-0.292 ± 0.125	0.550 ± 0.201	0.035 ± 0.248	-0.364 ± 1.471
φ3		0.354 ± 0.189	0.46 ± 0.158	-0.013 ± 0.137	0.866 ± 0.115	0.686 ± 0.362	-0.23 ± 0.093	0.132 ± 1.066
φ4		0.464 ± 0.146	0.264 ± 0.171	0.124 ± 0.092	0.093 ± 0.187	0.208 ± 0.635	-0.374 ± 0.095	-0.098 ± 0.317
φ5			0.330 ± 0.119	0.237 ± 0.077	0.497 ± 0.182	-0.421 ± 0.420	-0.130 ± 0.084	0.145 ± 0.067
φ6			-0.677 ± 0.116	-0.733 ± 0.119	-0.552 ± 0.181	-0.606 ± 0.276	0.284 ± 0.065	0.186 ± 0.095
φ7				0.532 ± 0.222	0.248 ± 0.182	-0.218 ± 0.624	0.430 ± 0.089	0.112 ± 0.134
φ8				0.121 ± 0.188	-0.231 ± 0.116	0.272 ± 0.509	0.150 ± 0.077	-0.056 ± 0.132
φ9					0.684 ± 0.125	0.344 ± 0.188	-0.257 ± 0.051	-0.181 ± 0.131
φ10					-0.273 ± 0.161	-0.104 ± 0.412	-0.426 ± 0.085	-0.220 ± 0.134
φ11						-0.660 ± 0.323	-0.238 ± 0.085	-0.191 ± 0.165
φ12						0.328 ± 0.196	1.245 ± 0.052	0.994 ± 0.162
φ13							-0.813 ± 0.239	-1.133 ± 0.839
φ14							0.070 ± 0.188	0.356 ± 1.435
φ15								-0.088 ± 1.018
φ16								0.173 ± 0.305
θ1	-0.716 ± 0.192	-0.449 ± 0.178	0.071 ± 0.212	-0.239 ± 0.308	0.344 ± 0.060	0.449 ± 0.660	-0.950 ± 0.158	-0.890 ± 0.743
θ2		1.025 ± 0.053	-0.470 ± 0.148	0.434 ± 0.134	0.803 ± 0.115	-0.034 ± 0.426	0.117 ± 0.134	0.509 ± 1.274
θ3		-0.383 ± 0.177	-0.002 ± 0.202	0.467 ± 0.254	-0.169 ± 0.112	-0.270 ± 0.550	0.207 ± 0.112	0.034 ± 0.861
θ4			0.127 ± 0.182	0.107 ± 0.262	0.37	-0.129 ± 1.282	0.226	-0.005 ± 0.329
θ5			-0.725 ± 0.113	-0.042 ± 0.232	-0.371	0.120 ± 1.035	0.083	-0.193 ± 0.212
θ6				0.742 ± 0.152	0.167 ± 0.098	0.142 ± 0.187	-0.224 ± 0.089	-0.309 ± 0.202
θ7				-0.640 ± 0.312	-0.801 ± 0.014	-0.124 ± 1.179	-0.304 ± 0.095	-0.319 ± 0.342
θ8					-0.3475	-0.232 ± 1.183	-0.126 ± 0.105	-0.150 ± 0.294
θ9					-0.9965	-0.045 ± 0.304	0.062	-0.033 ± 0.345
θ10						0.462 ± 0.644	0.200	0.134 ± 0.206
θ11						0.966 ± 0.735	0.183 ± 0.047	0.331 ± 0.223
θ12							-1.152 ± 0.145	-0.788 ± 0.255
θ13							0.679 ± 0.181	1.092 ± 0.780
θ14								-0.149 ± 1.237
θ15								0.048 ± 0.917
RSS	1.22566	0.70158	0.55454	0.40194	0.28527	0.2542	0.23317	0.23108
F		31.00	11.00	15.75	16.97	5.07	3.74	0.38
AIC	-333.44	-408.84	-437.19	-471.11	-510.97	-519.76	-520.19	-513.64

Appendix 5: ϕ coefficient matrix table of ARMAV model

	ϕ1					ϕ2					ϕ3					ϕ4			
	Tot	Ger	SK	Aus		Tot	Ger	SK	Aus		Tot	Ger	SK	Aus		Tot	Ger	SK	Aus
Tot	1.06	-0.45	-0.2	0.42		-0.19	0.36	0.19	-0.19		-0.38	-0.23	0.07	0.62		0.63	-0.32	-0.20	-0.45
Ger	1.28	-0.01	-0.88	0.14		0.21	0.18	0.82	0.14		-0.71	-0.51	-0.31	0.47		0.36	-0.36	-0.28	-0.04
SK	1.13	-0.66	0.67	0.55		-1.27	0.54	-0.29	-0.29		0.32	-0.09	-0.17	0.62		-0.06	-0.32	0.22	-0.22
Aus	0.30	0.30	0.19	0.28		0.16	0.11	-0.27	-0.32		-0.13	-0.46	-0.53	1.35		-0.70	-0.07	0.52	-0.01

Appendix 6: θ coefficient matrix table of ARMAV model

	θ1					θ2					θ3			
	Tot	Ger	SK	Aus		Tot	Ger	SK	Aus		Tot	Ger	SK	Aus
Tot	0.86	-0.36	-0.45	0.09		-0.48	0.21	-0.03	0.08		-0.34	-0.21	0.04	0.38
Ger	1.32	-0.10	-1.05	-0.19		0.17	-0.35	0.41	0.20		-0.86	-0.23	0.14	0.56
SK	1.51	-0.76	0.00	0.21		-1.01	0.39	-0.19	0.14		-0.12	-0.12	-0.16	0.32
Aus	0.61	0.11	0.06	-0.37		-0.31	0.15	0.06	-0.05		0.00	-0.19	-0.64	0.70