STUDENT PROJECT: TIME SERIES

NAME: SAOWALUK KONGSUKTHAI

REGISTRATION ID: 62280478

COMPANY: THAI LIFE INSURANCE PCL

COURSE: TIME SERIES

SESSION: WINTER 2014

Consider the following: Weekly market share data of Crest toothpaste from January 1958 to April 1963

Week	Data	Week	Data	Week	Data	Week	Data
0001 W01	0.108	0001 W39	0.153	0002 W25	0.198	0003 W11	0.138
0001 W02	0.166	0001 W40	0.078	0002 W26	0.197	0003 W12	0.216
0001 W03	0.126	0001 W41	0.114	0002 W27	0.251	0003 W13	0.132
0001 W04	0.115	0001 W42	0.088	0002 W28	0.146	0003 W14	0.120
0001 W05	0.119	0001 W43	0.165	0002 W29	0.133	0003 W15	0.083
0001 W06	0.176	0001 W44	0.160	0002 W30	0.243	0003 W16	0.118
0001 W07	0.155	0001 W45	0.075	0002 W31	0.192	0003 W17	0.125
0001 W08	0.118	0001 W46	0.118	0002 W32	0.150	0003 W18	0.109
0001 W09	0.136	0001 W47	0.100	0002 W33	0.221	0003 W19	0.119
0001 W10	0.137	0001 W48	0.102	0002 W34	0.183	0003 W20	0.154
0001 W11	0.124	0001 W49	0.131	0002 W35	0.136	0003 W21	0.122
0001 W12	0.131	0001 W50	0.148	0002 W36	0.206	0003 W22	0.126
0001 W13	0.120	0001 W51	0.137	0002 W37	0.127	0003 W23	0.126
0001 W14	0.133	0001 W52	0.090	0002 W38	0.139	0003 W24	0.130
0001 W15	0.067	0002 W01	0.088	0002 W39	0.189	0003 W25	0.158
0001 W16	0.086	0002 W02	0.172	0002 W40	0.194	0003 W26	0.141
0001 W17	0.140	0002 W03	0.111	0002 W41	0.114	0003 W27	0.145
0001 W18	0.122	0002 W04	0.097	0002 W42	0.229	0003 W28	0.127
0001 W19	0.105	0002 W05	0.098	0002 W43	0.148	0003 W29	0.171
0001 W20	0.079	0002 W06	0.090	0002 W44	0.155	0003 W30	0.152
0001 W21	0.130	0002 W07	0.127	0002 W45	0.106	0003 W31	0.211
0001 W22	0.142	0002 W08	0.116	0002 W46	0.156	0003 W32	0.309
0001 W23	0.120	0002 W09	0.137	0002 W47	0.053	0003 W33	0.242
0001 W24	0.115	0002 W10	0.111	0002 W48	0.112	0003 W34	0.380
0001 W25	0.103	0002 W11	0.107	0002 W49	0.084	0003 W35	0.362
0001 W26	0.078	0002 W12	0.097	0002 W50	0.191	0003 W36	0.328
0001 W27	0.093	0002 W13	0.134	0002 W51	0.149	0003 W37	0.359
0001 W28	0.086	0002 W14	0.160	0002 W52	0.143	0003 W38	0.352
0001 W29	0.099	0002 W15	0.147	0003 W01	0.094	0003 W39	0.322
0001 W30	0.078	0002 W16	0.104	0003 W02	0.184	0003 W40	0.333
0001 W31	0.095	0002 W17	0.128	0003 W03	0.205	0003 W41	0.365
0001 W32	0.094	0002 W18	0.128	0003 W04	0.206	0003 W42	0.367
0001 W33	0.056	0002 W19	0.165	0003 W05	0.191	0003 W43	0.305
0001 W34	0.050	0002 W20	0.184	0003 W06	0.195	0003 W44	0.298
0001 W35	0.065	0002 W21	0.172	0003 W07	0.179	0003 W45	0.307
0001 W36	0.091	0002 W22	0.207	0003 W08	0.272	0003 W46	0.318
0001 W37	0.094	0002 W23	0.221	0003 W09	0.203	0003 W47	0.280
0001 W38	0.124	0002 W24	0.159	0003 W10	0.165	0003 W48	0.298

Week	Data	Week	Data	Week	Data	Week	Data
0003 W49	0.336	0004 W35	0.41	0005 W21	0.388	0006 W07	0.388
0003 W50	0.339	0004 W36	0.425	0005 W22	0.373	0006 W08	0.377
0003 W51	0.344	0004 W37	0.358	0005 W23	0.385	0006 W09	0.466
0003 W52	0.31	0004 W38	0.393	0005 W24	0.314	0006 W10	0.478
0004 W01	0.317	0004 W39	0.375	0005 W25	0.347	0006 W11	0.365
0004 W02	0.369	0004 W40	0.273	0005 W26	0.408	0006 W12	0.472
0004 W03	0.32	0004 W41	0.237	0005 W27	0.341	0006 W13	0.399
0004 W04	0.29	0004 W42	0.331	0005 W28	0.361	0006 W14	0.391
0004 W05	0.361	0004 W43	0.335	0005 W29	0.414	0006 W15	0.473
0004 W06	0.235	0004 W44	0.395	0005 W30	0.38	0006 W16	0.384
0004 W07	0.32	0004 W45	0.357	0005 W31	0.274		
0004 W08	0.337	0004 W46	0.296	0005 W32	0.352		
0004 W09	0.289	0004 W47	0.307	0005 W33	0.439		
0004 W10	0.339	0004 W48	0.39	0005 W34	0.355		
0004 W11	0.187	0004 W49	0.298	0005 W35	0.435		
0004 W12	0.414	0004 W50	0.381	0005 W36	0.408		
0004 W13	0.373	0004 W51	0.354	0005 W37	0.383		
0004 W14	0.265	0004 W52	0.436	0005 W38	0.357		
0004 W15	0.316	0005 W01	0.357	0005 W39	0.374		
0004 W16	0.245	0005 W02	0.427	0005 W40	0.366		
0004 W17	0.328	0005 W03	0.432	0005 W41	0.346		
0004 W18	0.368	0005 W04	0.45	0005 W42	0.381		
0004 W19	0.287	0005 W05	0.53	0005 W43	0.329		
0004 W20	0.369	0005 W06	0.431	0005 W44	0.474		
0004 W21	0.406	0005 W07	0.42	0005 W45	0.397		
0004 W22	0.316	0005 W08	0.411	0005 W46	0.436		
0004 W23	0.362	0005 W09	0.423	0005 W47	0.417		
0004 W24	0.308	0005 W10	0.433	0005 W48	0.43		
0004 W25	0.286	0005 W11	0.393	0005 W49	0.388		
0004 W26	0.42	0005 W12	0.389	0005 W50	0.453		
0004 W27	0.299	0005 W13	0.387	0005 W51	0.316		
0004 W28	0.383	0005 W14	0.439	0005 W52	0.414		
0004 W29	0.354	0005 W15	0.421	0006 W01	0.396		
0004 W30	0.418	0005 W16	0.363	0006 W02	0.42		
0004 W31	0.425	0005 W17	0.401	0006 W03	0.432		
0004 W32	0.445	0005 W18	0.394	0006 W04	0.453		
0004 W33	0.408	0005 W19	0.459	0006 W05	0.43		
0004 W34	0.282	0005 W20	0.441	0006 W06	0.327		

a.) Build an ARIMA model for the series

Model Identification

From the data series in the previous page, the plot of this data is given in Figure1.

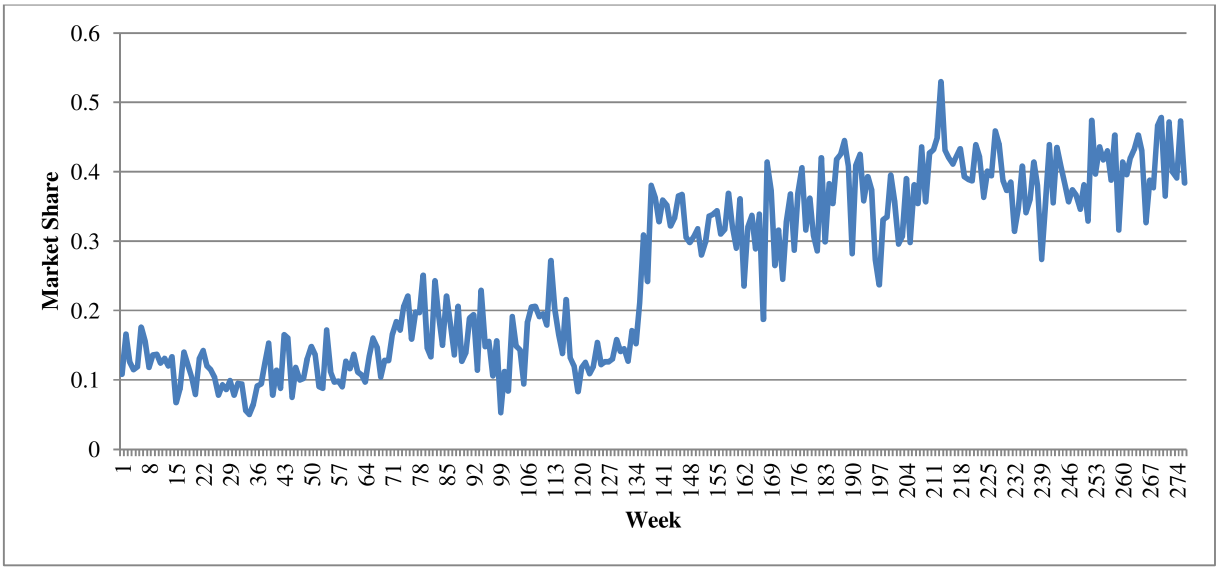


Figure 1 Deaths from suicides in Australia between 1915 and 2004

To investigate whether the series is stationary in the variance, we calculated the residual sum of squares as follow

$$S(\lambda) = \sum_{t=1}^{n} (Z_t(\lambda) - \hat{\mu})^2$$

where $\hat{\mu}$ is the corresponding simple mean of the transformed series and the results of the power transformation is given in Table 1.

Table 1 Results of the power transformation

λ	Residual sum of squares
1.0	**4.432059638**
0.5	3097.029644
0.0	352.6978389
-0.5	91.09809945
-1.0	4.760653534

From Table 1, the power transformation indicates that the power transformation is not needed.

The sample autocorrelation function (ACF) and the sample partial autocorrelation function (PACF) of the data are calculated by using the following formula and the results are given in Table 2.

$$\hat{\rho}_k = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2} \quad , k = 0, 1, 2, \dots$$

Where $\bar{Z} = \frac{1}{n}\sum_{t=1}^{n} Z_t$ and

$$\hat{\phi}_{k+1,k+1} = \frac{\hat{\rho}_{k+1} - \sum_{j=1}^{k} \hat{\phi}_{kj}, \hat{\phi}_{k+1-j}}{1 - \sum_{j=1}^{k} \hat{\phi}_{kj}, \hat{\rho}_j}$$

And $\hat{\phi}_{k+1,j} = \hat{\phi}_{kj} - \hat{\phi}_{k+1,k+1}, \hat{\phi}_{k,k+1-j} \quad , j = 1, 2, \dots, k$

Where $\hat{\phi}_{1,1} = \hat{\rho}_1$

The standard error of $\hat{\rho}_k$ is approximately by

$$S_{\hat{\rho}_k} \simeq \sqrt{\frac{1}{n}\left(1 + 2{\hat{\rho}_1}^2 + \dots + 2{\hat{\rho}_{k-1}}^2\right)} \ , k = 1, 2, \dots$$

And standard error of $\hat{\phi}_{kk}$ is approximately by

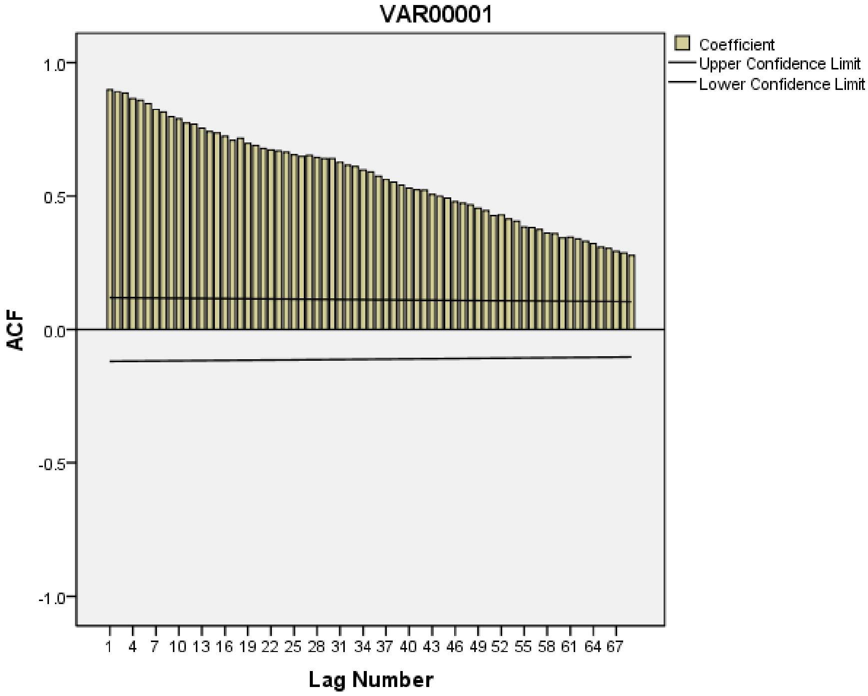
$$S_{\hat{\phi}_{kk}} \simeq \sqrt{\frac{1}{n}} \quad , k = 1, 2, \dots$$

Table 2 The sample ACF and PACF of the data (Z_t)

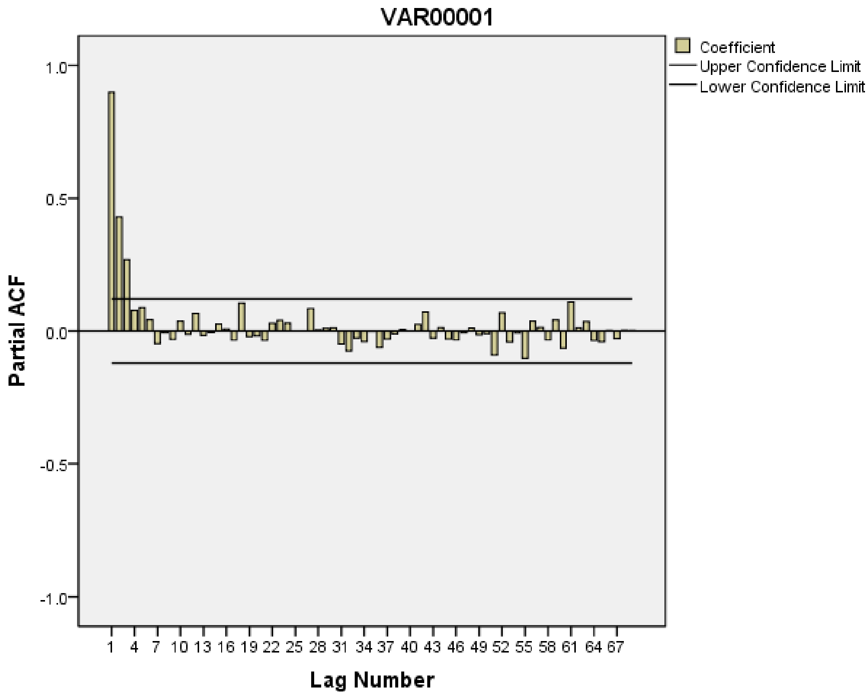
Lag	Autocorrelation	Std. Error ACF	Partial Autocorrelation	Std. Error PACF
1	.900	.060	.900	.060
2	.891	.060	.429	.060
3	.886	.060	.269	.060
4	.865	.060	.077	.060
5	.859	.059	.089	.060
6	.847	.059	.043	.060
7	.825	.059	-.048	.060
8	.815	.059	-.006	.060
9	.797	.059	-.031	.060
10	.790	.059	.037	.060
11	.774	.059	-.012	.060
12	.770	.059	.065	.060
13	.754	.059	-.016	.060
14	.742	.058	-.006	.060
15	.736	.058	.027	.060
16	.724	.058	.007	.060
17	.709	.058	-.034	.060
18	.717	.058	.105	.060
19	.698	.058	-.020	.060
20	.689	.058	-.018	.060
21	.678	.058	-.035	.060
22	.673	.058	.030	.060
23	.669	.057	.041	.060
24	.665	.057	.031	.060
25	.655	.057	.000	.060
26	.649	.057	.000	.060
27	.653	.057	.085	.060
28	.646	.057	.004	.060
29	.641	.057	.009	.060
30	.641	.057	.011	.060
31	.627	.057	-.049	.060
32	.615	.056	-.075	.060
33	.610	.056	-.026	.060
34	.598	.056	-.040	.060
35	.590	.056	-.001	.060
36	.575	.056	-.061	.060

37	.562	.056	-.030	.060
38	.552	.056	-.010	.060
39	.541	.056	.005	.060
40	.530	.055	.000	.060
41	.524	.055	.025	.060
42	.523	.055	.071	.060
43	.507	.055	-.026	.060
44	.500	.055	.012	.060
45	.492	.055	-.030	.060
46	.479	.055	-.033	.060
47	.474	.055	-.006	.060
48	.467	.055	.011	.060
49	.455	.054	-.014	.060
50	.446	.054	-.009	.060
51	.426	.054	-.090	.060
52	.430	.054	.068	.060
53	.414	.054	-.041	.060
54	.406	.054	-.007	.060
55	.384	.054	-.103	.060
56	.382	.054	.038	.060
57	.376	.053	.013	.060
58	.361	.053	-.033	.060
59	.359	.053	.043	.060
60	.343	.053	-.064	.060
61	.345	.053	.109	.060
62	.340	.053	.011	.060
63	.331	.053	.036	.060
64	.322	.053	-.036	.060
65	.309	.052	-.041	.060
66	.304	.052	.002	.060
67	.292	.052	-.028	.060
68	.286	.052	.002	.060
69	.277	.052	.001	.060

(a)



(b)



<u>Figure 2</u> (a) The sample ACF of the data
(b) The sample PACF of the data

Next, return to identify our model. From the patterns of the sample ACF and PACF of the series of the data in Figure 2, the sample ACF tails off and the sample PACF cuts off after lag 3 then the AR(3) model is suggested for the data. Hence, the following AR(3) model is entertained for the data,

$$(1-\phi_1 B-\phi_2 B^2-\phi_3 B^3)Z_t=a_t.$$

Moreover if we try to gets a better shape of sample ACF damped sine wave by use differencing and calculate sample ACF and sample PACF. Hence, the differenced data $\omega_t=(1-B)Z_t$ are plotted in Figure 3 and the sample ACF and PACF for the differenced data are calculated in Table 3 with their plots in Figure 4.

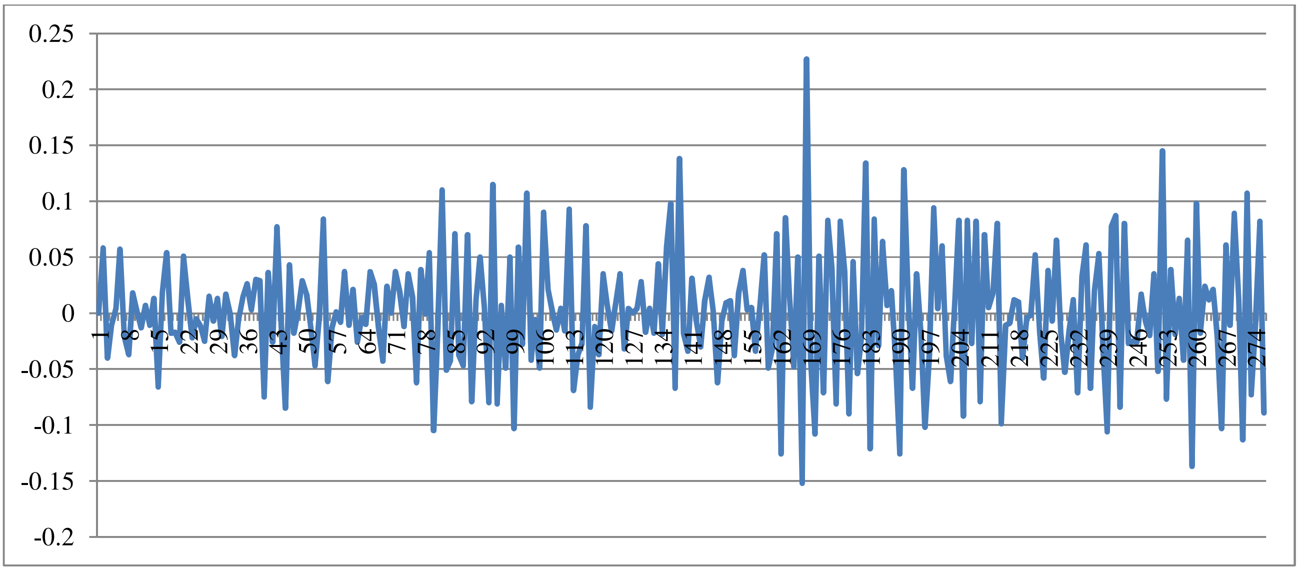


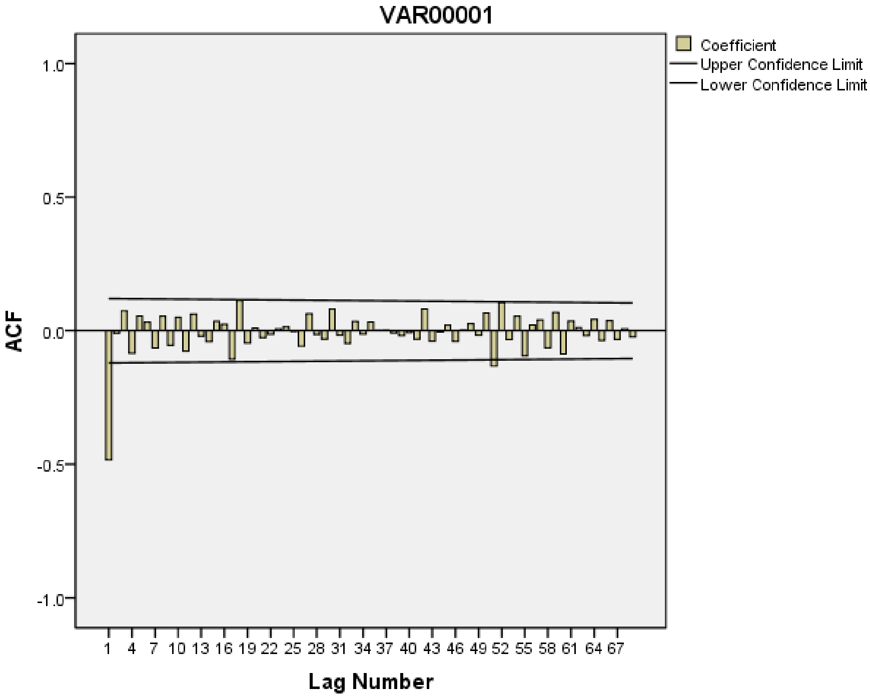
Figure 3 The differenced series of the weekly market share data of Crest toothpaste from January 1958 to April 1963, $\omega_t=(1-B)Z_t$

Table 3 The sample ACF and PACF for the differenced series of the data, $\omega_t = (1 - B)Z_t$ ($\bar{\omega} = 0.001003636,\ S_\omega = 0.055721839,\ n = 275$)

Lag	Autocorrelation	Std. Error ACF	Partial Autocorrelation	Std. Error PACF
1	-.483	.060	-.483	.060
2	-.010	.060	-.316	.060
3	.074	.060	-.122	.060
4	-.085	.060	-.144	.060
5	.055	.060	-.063	.060
6	.032	.059	.025	.060
7	-.066	.059	-.021	.060
8	.054	.059	.022	.060
9	-.054	.059	-.037	.060
10	.050	.059	.022	.060
11	-.077	.059	-.085	.060
12	.062	.059	-.019	.060
13	-.021	.059	-.027	.060
14	-.040	.059	-.071	.060
15	.035	.058	-.049	.060
16	.025	.058	.018	.060
17	-.106	.058	-.102	.060
18	.114	.058	-.007	.060
19	-.046	.058	-.001	.060
20	.011	.058	.017	.060
21	-.026	.058	-.043	.060
22	-.014	.058	-.068	.060
23	.008	.058	-.065	.060
24	.015	.057	-.051	.060
25	-.003	.057	-.027	.060
26	-.058	.057	-.116	.060
27	.062	.057	-.032	.060
28	-.014	.057	-.041	.060
29	-.032	.057	-.049	.060
30	.081	.057	.032	.060
31	-.017	.057	.069	.060
32	-.047	.056	.003	.060
33	.035	.056	-.005	.060
34	-.013	.056	-.020	.060

35	.033	.056	.026	.060
36	.001	.056	.026	.060
37	.002	.056	.043	.060
38	-.009	.056	.030	.060
39	-.019	.056	-.013	.060
40	-.007	.056	-.048	.060
41	-.032	.055	-.100	.060
42	.082	.055	.009	.060
43	-.039	.055	-.018	.060
44	-.005	.055	.020	.060
45	.021	.055	.024	.060
46	-.040	.055	-.014	.060
47	.003	.055	-.040	.060
48	.027	.055	-.020	.060
49	-.017	.054	-.020	.060
50	.066	.054	.081	.060
51	-.133	.054	-.077	.060
52	.106	.054	.021	.060
53	-.033	.054	.007	.060
54	.054	.054	.100	.060
55	-.094	.054	-.047	.060
56	.022	.054	-.030	.060
57	.040	.053	.020	.060
58	-.065	.053	-.073	.060
59	.068	.053	.039	.060
60	-.087	.053	-.124	.060
61	.036	.053	-.053	.060
62	.012	.053	-.086	.060
63	-.018	.053	-.030	.060
64	.042	.053	.006	.060
65	-.036	.053	-.017	.060
66	.037	.052	.027	.060
67	-.032	.052	-.008	.060
68	.008	.052	-.013	.060
69	-.023	.052	-.070	.060

(a)



(b)

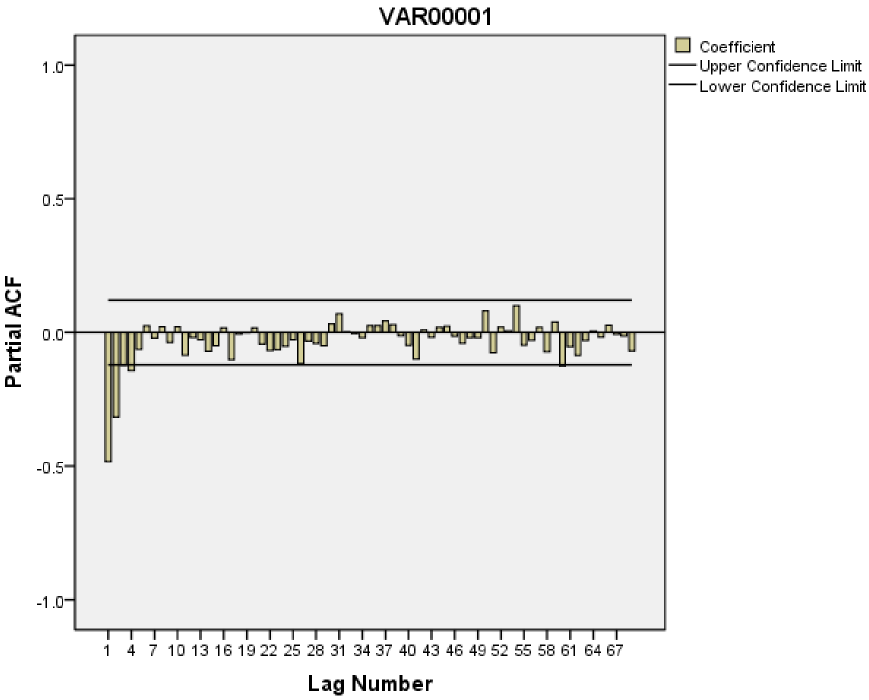


Figure 4 (a) The sample ACF of the differenced series of the data
(b) The sample PACF of the differenced series of the data, $\omega_t = (1-B)Z_t$

Next, return to identify our model. From the patterns of the sample ACF and PACF of the differenced series in Figure 4, the sample ACF cuts off after lag 1 and the sample PACF tails off. Hence, the following ARIMA(0,1,1) model is entertained for the data,

$$(1-B)Z_t = \theta_0 + (1-\theta_1 B)a_t.$$

To determine whether a deterministic trend term θ_0 is needed, we calculate the t- ratio with $\bar{\omega} = 0.001003636,\ S_\omega = 0.055721839\ \ (n = 275)$

$$t = \frac{\bar{w}}{s_{\bar{w}}} = \frac{\bar{w}}{s_w/\sqrt{n-d}} = \frac{0.001003636}{0.055721839/\sqrt{275}} \simeq 0.2987$$

Table 4 The value of $t_{\alpha/2}$ at the significant level $\boldsymbol{\alpha} = 0.05$

No. of observations	**Model**	**Degrees of freedom**	$t_{\alpha/2}$
275	$(1-B)Z_t = \theta_0 + (1-\theta_1 B)a_t$	273	1.9687

For α =0.05, from Table 4 for ARIMA(0,1,1) model for the differenced series $\omega_t = (1-B)Z_t$, since t-ratio t = 1.9687 is not in the rejection area the θ_0 is not significant.

Therefore, the following models are entertained,

$$(1-\phi_1 B - \phi_2 B^2 - \phi_3 B^3)Z_t = a_t$$
and $(1-B)Z_t = \theta_0 + (1-\theta_1 B)a_t$.

Model Estimation

By using SPSS to estimate parameters in our model the results are given in Table 5.

Table 5 Summary of models fitted to the data

No. of observations	Fitted models	$\hat{\sigma}_a{}^2$
276	AR(3): $(1-0.364B-0.318B^2-0.312B^3)Z_t = a_t$	0.00220851
276	ARIMA(0,1,1): $(1-B)Z_t = (1-0.664B)a_t$	0.00221647

Model Diagnostic Checking

From Table 5, we will see that all parameters in both models are significant. To check model adequacy Table 6 and Table 7 give the residual ACF and PACF of AR(3) and ARIMA(0,1,1) model respectively and their plots are also given in Figure 5 and Figure 6 respectively.

We test whether model is adequate based on the entire residual sample ACF with the test statistic.

$$Q = n(n+2)\sum_{k=1}^{K}(n-k)^{-1}\hat{\rho}_k^{\,2}$$

Under the null hypothesis of model adequacy where Q statistic approximately follows the $\chi^2(K-m)$ distribution where m is the number of parameter estimated in the model and the results of the residual analyses is given in Table 6.

Table 6 The residual ACF and PACF of AR(3) model of the data

Lag	Autocorrelation	Std. Error ACF	Partial Autocorrelation	Std. Error PACF
1	-0.03407	0.060193	-0.03407	0.060193
2	-0.09706	0.060263	-0.09834	0.060193
3	-0.14707	0.060827	-0.15574	0.060193
4	-0.02528	0.062102	-0.0502	0.060193
5	0.06776	0.062139	0.034335	0.060193
6	0.041429	0.062406	0.017205	0.060193
7	-0.04504	0.062506	-0.04393	0.060193
8	0.015523	0.062623	0.03193	0.060193
9	-0.05177	0.062637	-0.047	0.060193
10	0.000705	0.062792	-0.0137	0.060193
11	-0.07278	0.062792	-0.08602	0.060193
12	0.013236	0.063097	-0.00632	0.060193
13	-0.03473	0.063107	-0.05908	0.060193
14	-0.03933	0.063176	-0.06934	0.060193
15	-0.00052	0.063265	-0.01703	0.060193
16	0.011072	0.063265	-0.01158	0.060193
17	-0.06215	0.063272	-0.08282	0.060193
18	0.071411	0.063493	0.054117	0.060193
19	-0.02054	0.063783	-0.02053	0.060193
20	-0.01291	0.063807	-0.03516	0.060193
21	-0.07143	0.063816	-0.07597	0.060193
22	-0.0399	0.064105	-0.06292	0.060193
23	0.002835	0.064195	-0.04206	0.060193
24	0.001817	0.064196	-0.05994	0.060193

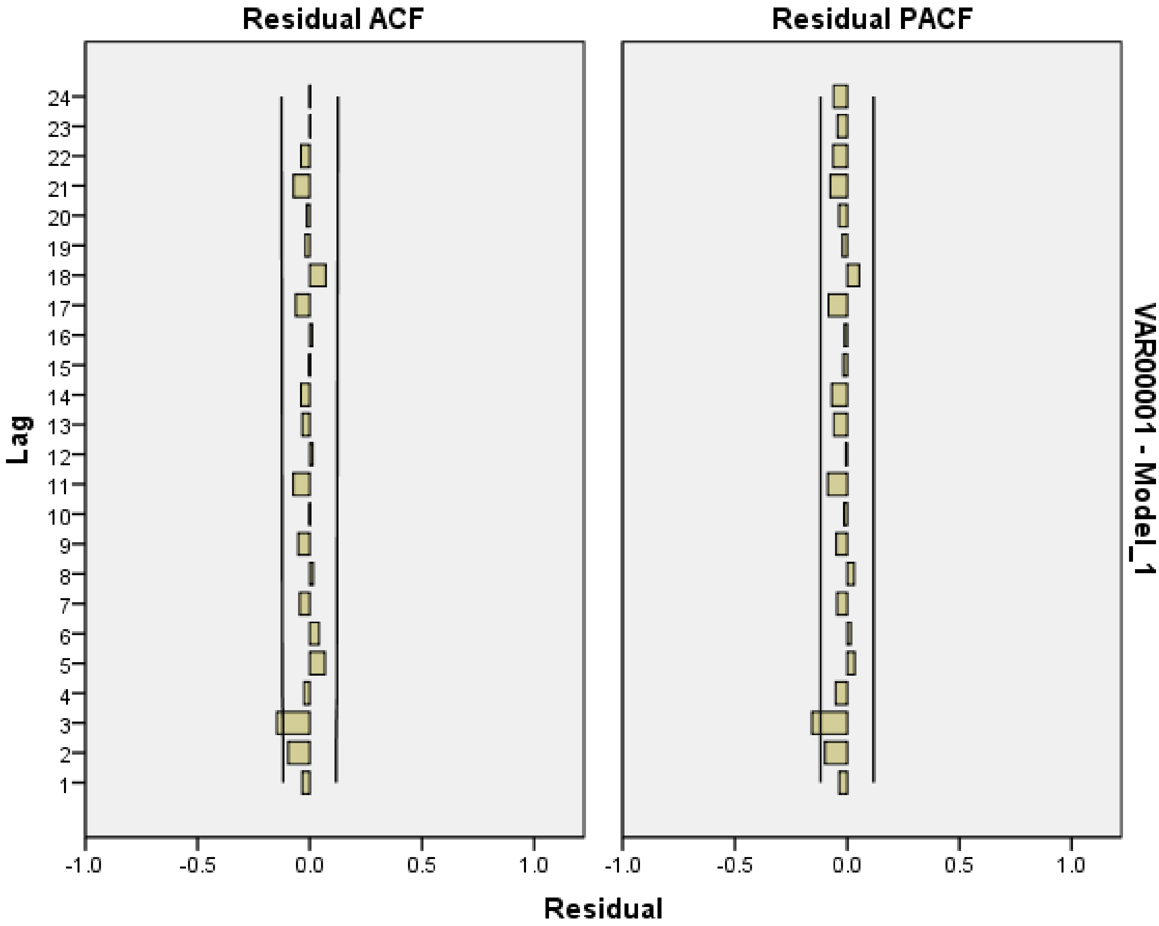


Figure 5 The residual ACF and PACF of AR(3) model

Table 7 The residual ACF and PACF of ARIMA(0,1,1) model of the data

Lag	Autocorrelation	Std. Error ACF	Partial Autocorrelation	Std. Error PACF
1	-0.04368	0.060302	-0.04368	0.060302
2	0.001221	0.060417	-0.00069	0.060302
3	0.060575	0.060417	0.060714	0.060302
4	-0.03457	0.060638	-0.02944	0.060302
5	0.062075	0.060709	0.05953	0.060302
6	0.035372	0.06094	0.037222	0.060302
7	-0.0464	0.061014	-0.04017	0.060302
8	0.011254	0.061143	-0.00054	0.060302
9	-0.05852	0.06115	-0.05921	0.060302
10	-0.01572	0.061353	-0.01746	0.060302
11	-0.09342	0.061368	-0.10382	0.060302

12	-0.0038	0.061883	-0.00121	0.060302
13	-0.05366	0.061884	-0.05453	0.060302
14	-0.06754	0.062053	-0.05855	0.060302
15	-0.00121	0.06232	-0.00583	0.060302
16	-0.017	0.06232	-0.00416	0.060302
17	-0.09025	0.062337	-0.08304	0.060302
18	0.050797	0.06281	0.035884	0.060302
19	-0.04153	0.062959	-0.02839	0.060302
20	-0.04335	0.063059	-0.05095	0.060302
21	-0.07362	0.063167	-0.09948	0.060302
22	-0.05847	0.063478	-0.06921	0.060302
23	-0.02125	0.063674	-0.04092	0.060302
24	-0.00916	0.063699	-0.0327	0.060302

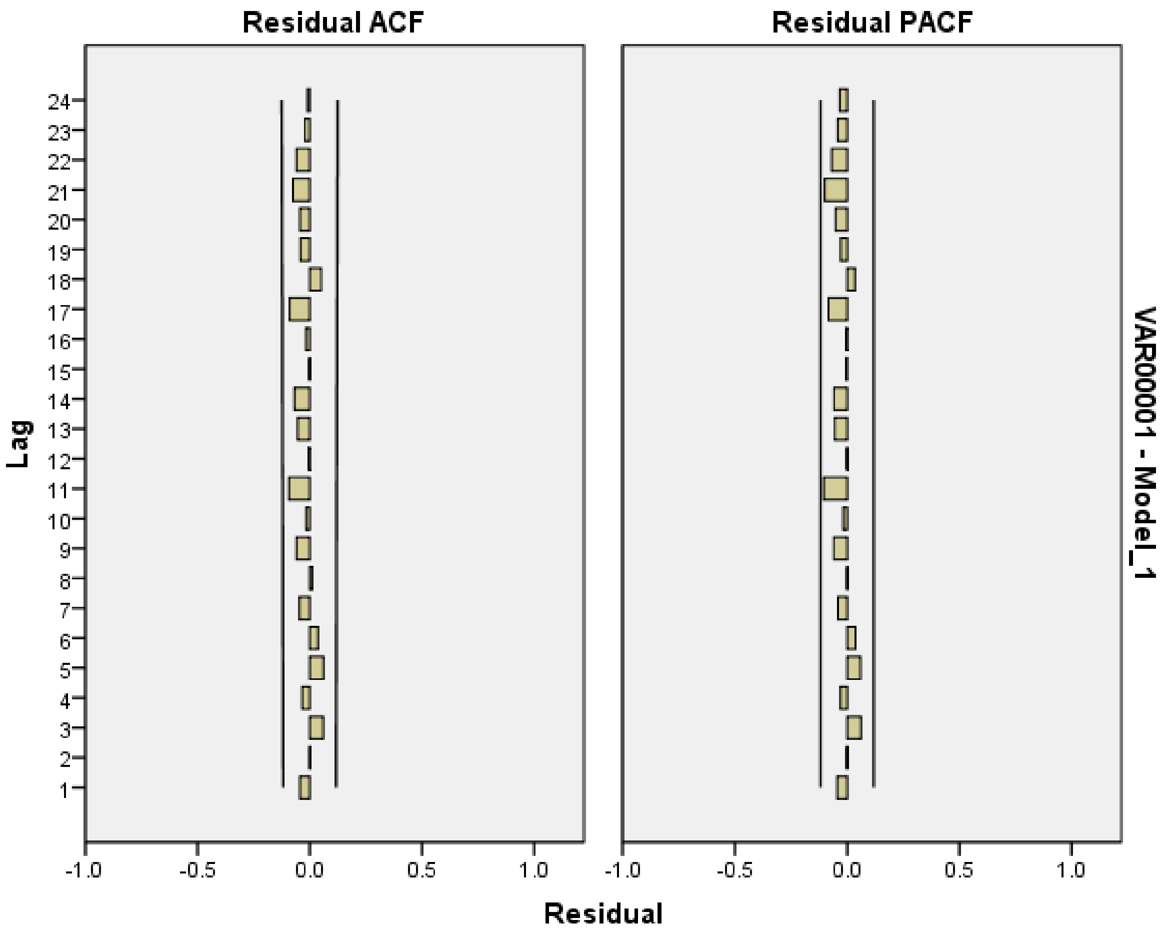


Figure 6 The residual ACF and PACF of ARIMA(0,1,1) model

Table 8 The results of residual analysis

Model	Degree of freedom $(K-m)$	Q- statistic	$\chi^2_{0.05}$
AR(3): $(1-0.364B-0.318B^2-0.312B^3)Z_t = a_t$	15	17.514	24.996
ARIMA(0,1,1): $(1-B)Z_t = (1-0.664B)a_t$	17	12.982	27.587

From the Table 6 and Table 7 with the plot in Figure 5 and Figure 6, we will see that the sample ACF and PACF of the residuals of AR(3) and ARIMA(0,1,1) are small and exhibit no pattern. Moreover, the results of residual analysis in Table 8 shows that AR(3) and ARIMA(0,1,1) model have Q less than $\chi^2_{0.05}$; therefore we can conclude that both models are adequate.

Model Selection

By using Akaike's BIC

$$BIC(M) = n\, ln\hat{\sigma}_a^2 - (n-M)ln\left(1-\frac{M}{n}\right) + M\, ln\, n + M\ln\left[\frac{\left(\frac{\hat{\sigma}_Z^2}{\hat{\sigma}_a^2}-1\right)}{M}\right]$$

where $\hat{\sigma}_a^2$ is the maximum likelihood estimate of $\hat{\sigma}_a^2$,

M is the number of parameters $(M = p+q)$, and $\hat{\sigma}_Z^2$ is the sample variance of the series. The optimal order of the model is chosen by the value of M so that $BIC(M)$ is minimum.

Table 9 BIC values $(\omega_t = (1-B)Z_t)$

No. of observations	Model	BIC
276	AR(3): $(1-0.364B-0.318B^2-0.312B^3)Z_t = a_t$	-6.060
275	ARIMA(0,1,1): $\omega_t = (1-0.664B)a_t$	-6.152

Therefore from Table 9 based on the BIC criterion an ARIMA(0,1,1) model,

$$\omega_t = (1 - 0.664B)a_t$$

should be selected to fit the data.

b.) Forecast the next 4 observations and calculates their associated 95% forecast limits.

Forecasting

From part a.) we selected the ARIMA(0,1,1) model

$$(1 - B)Z_t = (1 - 0.664B)a_t$$

with $\hat{\sigma}_a{}^2 = 0.00221647$

1.) Difference Equation Approach

$$(1 - B)Z_t = (1 - 0.664B)a_t$$

To derive the variance of the forecast for the general ARIMA model, we rewrite the model at time $t + l$ in an AR representation that exists because the model is invertible. Thus,

$$\pi(B)Z_{t+l} = a_{t+l},$$

where

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{(1-B)}{(1-0.664B)},$$

or

$$(1 - B) = 1 - (\pi_1 + 0.664)B - (\pi_2 - 0.664\pi_1)B^2 - \cdots$$

Equating the coefficients of B^j on both sides gives

$$\pi_j = (1 - 0.664)0.664^{j-1}, \quad j \geq 1$$

Now, applying the equation above, we obtain

$$\psi_1 = \pi_1 = 1 - 0.664 = 0.336$$

$$\psi_2 = \pi_2 + \pi_1\psi_1 = (1 - 0.664)(0.664) + (1 - 0.664)^2 = 1 - 0.664 = 0.336$$

$$\vdots$$

That is, we have

$$\psi_j = 0.336, \quad 1 \le j \le l-1$$

then

$$(1 + 0.336B + 0.336B^2 + 0.336B^3 + \cdots)(1-B)Z_t = a_t$$

$$(1 - 0.336B - 0.223B^2 - 0.148B^3 - \cdots)Z_t = a_t$$

$$Z_t = a_t + 0.336Z_{t-1} + 0.223Z_{t-2} + 0.148Z_{t-3} + \cdots$$

$$Z_t = a_t + \sum_{j=1}^{276}(1-0.664)0.664^{j-1}Z_{t-j}$$

and let $t = n + l$,

$$\hat{Z}_n(l) = \hat{a}_n(l) + \sum_{j=1}^{276}(1-0.664)0.664^{l-1}\hat{Z}_n(l-j)$$

Therefore,

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty}(1-0.664)0.664^{l-1}\hat{Z}_n(l-j)\,, l > 0$$

Note that $\hat{Z}_n(l) = Z_{n+l}$ for $l \le 0$.

<u>For $n = 276$</u>,

$$\hat{Z}_{276}(1) = \sum_{j=1}^{\infty} \pi_j{}^{(1)} Z_{n+1-j},$$

$$\hat{Z}_{276}(2) = \sum_{j=1}^{\infty} \pi_j{}^{(2)} Z_{n+1-j},$$

$$\hat{Z}_{276}(3) = \sum_{j=1}^{\infty} \pi_j{}^{(3)} Z_{n+1-j},$$

and $\hat{Z}_{276}(4) = \sum_{j=1}^{4} \pi_j{}^{(4)} Z_{n+1-j}$

where $\pi_j{}^{(l)} = \pi_{j+l-1} + \sum_{j=1}^{l-1} \pi_i \pi_j{}^{(l-1)}$, and $\pi_j{}^{(1)} = \pi_j$.

2.) <u>Eventual Forecast Function Approach</u>

Consider

$$(1-B)Z_t = (1-0.664B)a_t$$

with $\hat{\sigma}_a{}^2 = 0.00221647$

Let $\omega_t = (1-B)Z_t$

Refer to Box & Jenkins (1994, P.159), we can rewrite the model as follows:

$$(\omega_t - \hat{\mu}_\omega) = (1-0.664B)a_t$$

where $\hat{\mu}_\omega = 0$.

We can rewrite the model as follows:

$$\omega_t = a_t - 0.664a_{t-1}$$

Then,

$$\widehat{\omega}_n(l) = \hat{a}_n(l) - 0.664\hat{a}_n(l-1)$$

$$\widehat{\omega}_n(1) = \hat{a}_n(1) - 0.664a_n = -0.664a_n$$

$$\widehat{\omega}_n(2) = \hat{a}_n(2) - 0.664\hat{a}_n(1) = 0$$

$$\vdots$$

$$\widehat{\omega}_n(l) = \hat{a}_n(l) - 0.664\hat{a}_n(l-1) = 0\,, (l \neq 1)$$

$$\Rightarrow \hat{Z}_n(l) - \hat{Z}_n(l-1) = 0$$

So that,

$$\hat{Z}_n(l) = \begin{cases} Z_n - \theta a_n & \text{for } l = 1 \\ \hat{Z}_n(1) & \text{for } l > 1 \end{cases}$$

<u>For $n = 276$</u>,

$$\hat{Z}_{276}(1) = Z_{276} - 0.664a_{276} = 0.384 - 0.664(-0.0477081) = 0.41567819$$

$$\hat{Z}_{276}(2) = \hat{Z}_{276}(1) = 0.41567819$$

$$\hat{Z}_{276}(3) = \hat{Z}_{276}(1) = 0.41567819$$

$$\hat{Z}_{276}(4) = \hat{Z}_{276}(1) = 0.41567819$$

The result of forecasting for the next 4 observations of the data from difference equation approach and the eventual forecast function approach are given in Table 10.

Table 10 Forecasting the next 4 observations from the origin $n = 276$ for the data Z_t

l	$\hat{Z}_{276}(l)$	
	Difference Equation Approach	Eventual Forecast Function Approach
1	0.41567819	0.41567819
2	0.41567819	0.41567819
3	0.41567819	0.41567819
4	0.41567819	0.41567819

Next consider for the variance of forecast $\hat{Z}_n(l)$ for Z_{n+l}

First, consider the general ARIMA(p,d,q) model

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t + \theta_0$$

where $\phi(B)$ is a stationary AR operator and $\theta(B)$ is an invertible MA operator, respectively.

To derive the variance of the forecast for the general ARIMA model, we rewrite the model at time $t + l$ in an AR representation that exists because the model is invertible.

Thus,

$$\pi(B) = 1 - \sum_{j=1}^{\infty} \pi_j B^j = \frac{\phi(B)(1-B)^d}{\theta(B)}$$

or equivalently,

$$Z_{t+l} = \sum_{j=1}^{\infty} \pi_j Z_{t+l-j} + a_{t+l} + \frac{\theta_0}{\theta(B)} \quad \Rightarrow (1)$$

Then, apply operator $1 + \psi_1 B + \psi_2 B^2 + \cdots + \psi_{l-1} B^{l-1}$ to (1).

We obtain

$$\sum_{k=0}^{l-1} \psi_k B^k \left(\sum_{j=1}^{\infty} \pi_j B^j Z_{t+l} + a_{t+l} + \frac{\theta_0}{\theta(B)} \right) = 0$$

$$\Rightarrow \sum_{j=0}^{\infty}\sum_{k=0}^{l-1}\pi_j\psi_k Z_{t+l-j-k} + \sum_{k=0}^{l-1}\psi_k a_{t+l-k} + \frac{\theta_0}{\theta(B)}\sum_{k=0}^{l-1}\psi_k = 0$$

and $\sum_{j=0}^{\infty}\sum_{k=0}^{l-1}\pi_j\psi_k Z_{t+l-j-k}$ can be written as

$$\pi_0 Z_{t+l} + \sum_{m=1}^{l-1}\sum_{i=0}^{m}\pi_{m-i}\psi_i Z_{t+l-m} + \sum_{j=1}^{\infty}\sum_{i=0}^{l-1}\pi_{l-1+j-i}\psi_i Z_{t-j+1}$$

By choosing ψ weights so that

$$(*)\ \sum_{i=0}^{m}\pi_{m-i}\psi_i = 0\ , m = 1,2,\dots,l-1$$

We have

$$Z_{t+l} = \sum_{j=1}^{\infty}\pi_j{}^{(l)}Z_{t-j+1} + \sum_{i=0}^{l-1}\psi_i a_{t+l-i} + \frac{\theta_0}{\theta(B)}\sum_{k=0}^{l-1}\psi_k$$

where $\pi_j{}^{(l)} = \sum_{j=0}^{l-1}\pi_{l-1+j-i}\psi_i$.

Then, for given Z_t for $t \le n$, we have

$$\hat{Z}_n(l) = E(Z_{n+l}|Z_t, t \le n) = \sum_{j=1}^{\infty}\pi_j{}^{(l)}Z_{t-j+1} + \frac{\theta_0}{\theta(B)}\sum_{k=0}^{l-1}\psi_k$$

because $E(a_{n+j}|Z_t, t \le n) = 0$ for $j > 0$.

The forecast error is

$$e_n(l) = Z_{n+l} - \hat{Z}_n(l) = \sum_{i=0}^{l-1}\psi_i a_{t+l-i}$$

when ψ weights are calculated by using $(*)$ as follow:

$$\psi_j = \sum_{i=0}^{j-1}\pi_{j-i}\psi_i\,, \quad j = 1, 2, \dots, l-1.$$

Therefore,

$$Var(e_n(l)) = \sigma_a^2\sum_{j=0}^{l-1}\psi_j^2$$

and the $(1-\alpha)100\%$ forecast limits are

$$\hat{Z}_n(l) \pm N_{\alpha/2}(\sigma_a)\sqrt{1+\sum_{i=1}^{l-1}\psi_{i-1}^2}.$$

For our ARIMA(0,1,1) model,

$$(1-B)Z_t = (1-0.664B)a_t$$

with $\hat{\sigma}_a{}^2 = 0.00221647$, we can rewrite this model in the form of

$$\pi(B)Z_t = a_t$$

where

$$\pi(B) = 1-\sum_{j=1}^{\infty}\pi_j B^j = \frac{(1-B)}{(1-0.664B)},$$

or

$$(1-B) = 1-(\pi_1+0.664)B-(\pi_2-0.664\pi_1)B^2-\cdots$$

Equating the coefficients of B^j on both sides gives

$$\pi_j = (1-0.664)0.664^{j-1}, \quad j \geq 1$$

That is, we have

$$\psi_j = 0.336, \quad 1 \leq j \leq l-1$$

Now, applying the equation above, we obtain

$$\psi_1 = \psi_2 = \psi_3 = 0.336$$

The 95% forecast limits for Z_{276+l} are

$$\hat{Z}_{276}(l) \pm N_{0.025}(\sigma_a)\sqrt{1+\sum_{j=1}^{l-1}\psi_j^2}$$

where $\hat{\sigma}_a^2 = 0.00221647$ and $N_{0.025} = 1.96$

Table 11 The 95% forecast limits for Z_{276+l}

l	The 95% forecast limits for Z_{276+l}	
	Difference Equation Approach	Eventual Forecast Function Approach
1	0.41567819 ± 0.08983	0.41567819 ± 0.08983
2	0.41567819 ± 0.09475	0.41567819 ± 0.09475
3	0.41567819 ± 0.09943	0.41567819 ± 0.09943
4	0.41567819 ± 0.10390	0.41567819 ± 0.10390

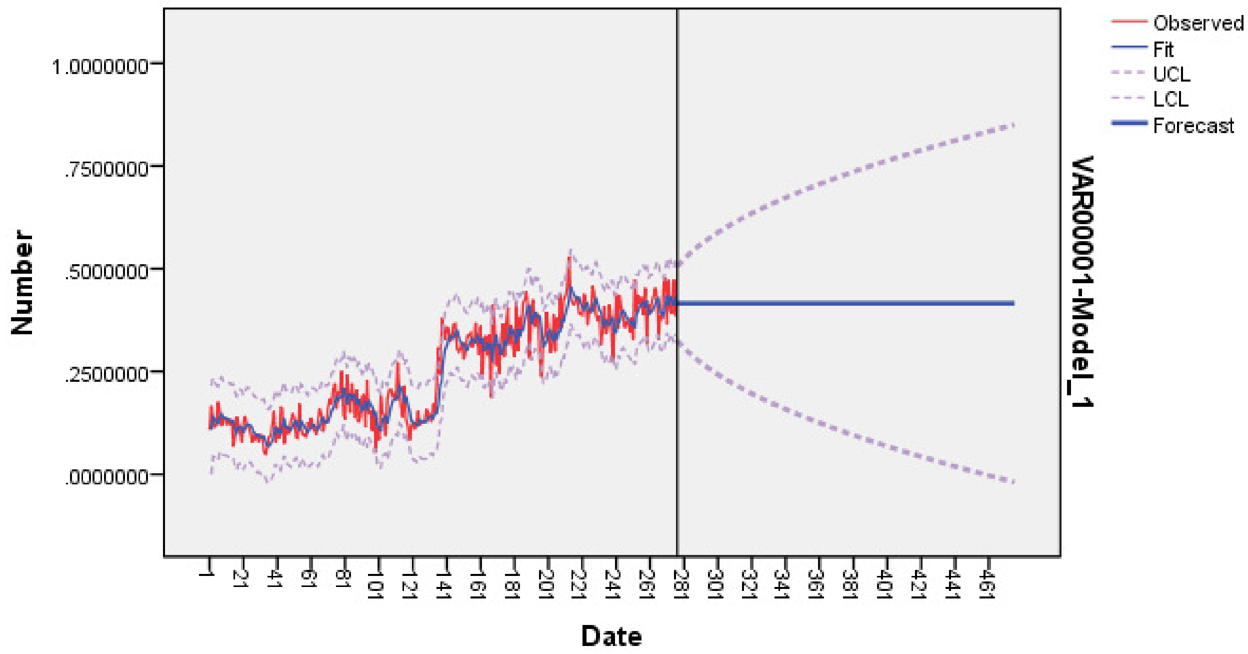


Figure 7 Forecasts with 95% forecast limits for the data

From table 10, we will see that the results of the forecasting by using difference equation approach and eventual forecast function approach are similarly therefore we can use both methods to forecast but for convenience when l is large, we can use the eventual function forecast function approach.

c.) Update your forecasts when one more observation became available and equaled 0.4224

Using the equation:

$$\hat{Z}_n(l) = \sum_{j=1}^{\infty}(1-0.664)0.664^{l-1}\hat{Z}_n(l-j)\,, l > 0 \text{ and } \hat{Z}_n(l) = \begin{cases} Z_n - \theta a_n & \text{for } l = 1 \\ \hat{Z}_n(1) & \text{for } l > 1 \end{cases}$$

Note that $\hat{Z}_n(l) = Z_{n+l}$ for $l \leq 0$.

And $Z_{277} = 110\%(Z_{276}) = 1.1(0.384) = 0.4224$

For $n = 277$,

$$\hat{Z}_{277}(1) = \sum_{j=1}^{\infty} \pi_j{}^{(1)} Z_{n+1-j}, \qquad \hat{Z}_{277}(1) = Z_{277} - 0.664a_{277}$$

$$\hat{Z}_{277}(2) = \sum_{j=1}^{\infty} \pi_j{}^{(2)} Z_{n+1-j}, \qquad \hat{Z}_{277}(2) = \hat{Z}_{277}(1)$$

$$\hat{Z}_{277}(3) = \sum_{j=1}^{\infty} \pi_j{}^{(3)} Z_{n+1-j}, \qquad \hat{Z}_{277}(3) = \hat{Z}_{277}(1)$$

and $\hat{Z}_{277}(4) = \sum_{j=1}^{4} \pi_j{}^{(4)} Z_{n+1-j}, \qquad \hat{Z}_{277}(4) = \hat{Z}_{277}(1)$

where $\pi_j{}^{(l)} = \pi_{j+l-1} + \sum_{j=1}^{l-1} \pi_i \pi_j{}^{(l-1)}$, and $\pi_j{}^{(1)} = \pi_j$.

Table 12 Forecasting the next 4 observations from the origin $n = 277$ for the data Z_t

l	$\hat{Z}_{277}(l)$	
	Difference Equation Approach	Eventual Forecast Function Approach
1	0.4179395	0.4179395
2	0.4179395	0.4179395
3	0.4179395	0.4179395
4	0.4179395	0.4179395