Kingdom of Saudi Arabia
KING SAUD UNIVERSITY
College of Adminis-
trative Sciences
Research Center

FORECASTING THE NUMBER OF EXTERNAL PILGRIMS: A BOX-JENKINS APPROACH

By:

Mohamed Fathi Abou-El-Fetouh
Assistant Professor
Dept. of Quantitative Methods
King Saud University
Riyadh, Saudi Arabia.

1407 H / 1987

بسم الله الرحمن الرحيم

King Saud University
College of Administrative Sciences
Research Center

FORECASTING THE NUMBER OF EXTERNAL PILGRIMS: A BOX-JENKINS APPROACH

By:

Mohamed Fathi Abou-El-Fetouh
Assistant Professor
Dept. of Quantitative Methods
King Saud University
Riyadh, Saudi Arabia.

1407 H / 1987

King Saud University Press – 1407 A.H.

SUMMARY

A time series model has been developed for forecasting the number of pilgrims using Box-Jenkins approach. The identified model represents a random walk process. This leads to the finding that the best forecast for next year's Hajj is just simply the present year's Hajj.

I- INTRODUCTION

For the purpose of effective planning of Hajj services it is necessary to forecast the future numbers of pilgrims and in particular external Hajjis. In the present research we have tried to forecast the number of external pilgrims by making use of Box-Jenkins approach.

The forecasting of the numbers of external pilgrims through the application of other methods has also been attempted. H. Pasha and H. Jamal (1983) estimated and predicted the external demand for pilgrimage using an econometric model. They assumed that a measure for external Hajj is influenced through a partial adjustment hypothesis by the lagged effect of the measure, the income level, distance from Saudi Arabia, system's capacity expansion, congestion, and random effects.

The present research uses time series models, namely Box-Jenkins models in an attempt to obtain high forecasting accuracy of the generated predictions.

This research is divided into nine sections: The first contains a broad outline of the Box-Jenkins Approach. The second deals with a preliminary examination of data and the third with the identification of the original time series. The fourth section is concerned with estimating the identified models. In the fifth section the original time series is differenced and models are identified from this differenced series. Estimation of these models is performed in the sixth section. This is followed by diagnostic checking and forecasting in the seventh and eighth sections respectively. The last section contains conclusions drawn from the empirical results.

I- BROAD OUTLINE OF THE BOX-JENKINS PROCEDURE

A stationary process can often be parsimoniously represented by a mixture of autoregressive and moving average models. A homogeneous non-stationary process needs to be differenced in order to fit a stationary model. Conversly a stationary process may be summed or integrated to give a non-stationary process and an integrated model for a non-stationary process is one which yields a stationay model when differenced in the appropriate way. The Box-Jenkins procedure is concerned with fitting a mixed autoregressive integrated moving average ARIMA model to a given set of data.

The general algebraic expression of an ARIMA model for seasonal or non-seasonal realizations and the important special cases that occur more frequently in practice are widely discussed in the forecasting literature, and hence repetition is unnecessary.

We concentrate on describing briefly the main stages in setting up a Box-Jenkins forecasting model. These are as follows:

1- Model Identification:

This involves examination of data to see which model in the class of ARIMA processes appears to be the most appropriate.

2- Estimation:

This means estimating the parameters of the selected model by least squares or other methods.[1]

3- Diagnostic Checking:

This is done by examining the estimated residuals from the fitted model to see if it is adequate.

4- Consideration of Alternative Models if Necessary:

If the first model appears to be inadequate for some reason, then other ARIMA models may be studied by repeating the above procedure until a satisfactory model is found.

5- Forecasting:

This involves the use of the adequate estimated model(s) for forecasting for one or more periods into the future.

---

(1) For example Maximum Likelihood and Bayesian methods.

2- THE DATA

The data are given in table 1 and plotted in figure 1. The series rises through time, so its mean may not be fixed. But deciding if the mean is fixed with visual inspection can be misleading. We must rely on the appearance of the autocorrelation and autoregressive coefficients at the estimation stage to decide if the external Hajj data has fixed mean.

Inspection of figure 1 suggests that variance is stationary since it does not change over time, and hence no transformation is required.

TABLE 1

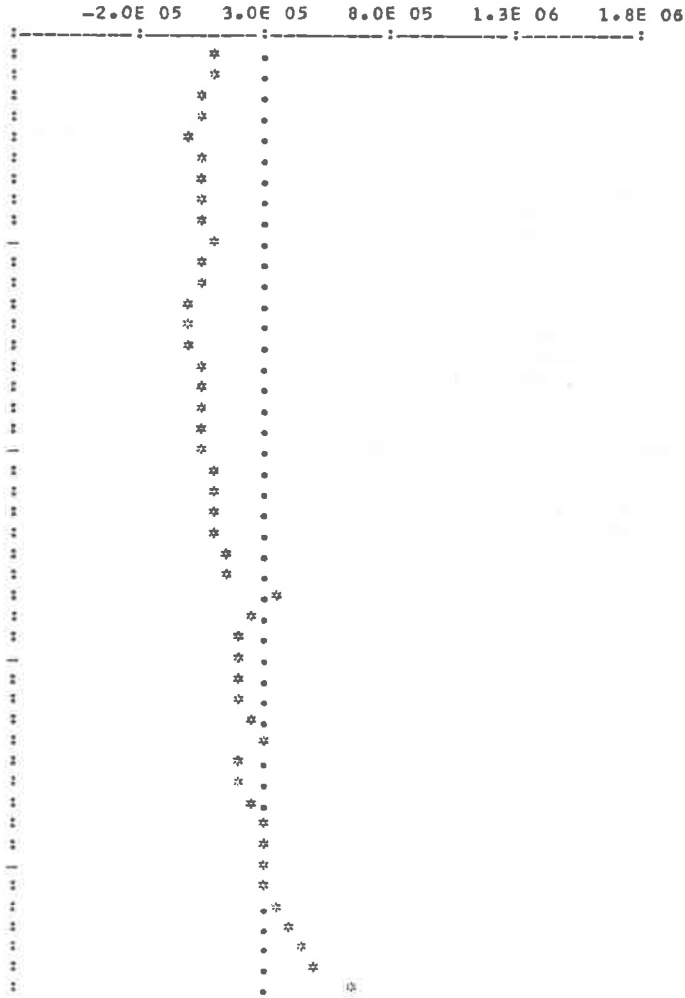
NUMBER OF PILGRIMS IN THE YEAR 1347 A.H. to 1404 A.H.

Year	Pilgrims	year	Pilgrims
1347	90764	1376	215575
1348	81666	1377	209197
1349	39045	1378	207171
1350	29065	1379	253369
1351	20181	1380	285948
1352	25291	1381	216455
1353	33898	1382	199038
1354	33830	1383	266555
1355	49517	1384	283319
1356	76224	1385	294118
1357	59577	1386	316226
1358	32152	1387	318507
1359	9024	1388	374784
1360	23863	1389	406295
1361	24743	1390	431270
1362	62590	1391	479339
1363	37857	1392	645182
1364	37630	1393	607755
1365	61286	1394	918777
1366	55244	1395	894573
1367	75614	1396	719040
1368	99069	1397	739319
1369	107652	1398	830236
1370	100578	1399	862520
1371	148515	1400	812892
1372	149841	1401	879368
1373	364072	1402	853555
1374	232971	1403	1005060
1375	220722	1404	919671

Source: Kingdom of Saudi Arabia, Ministry of Interior: Pilgrimage Statistics for 1404 A.H.(Dar AlAsfhani, Jeddah).

GRAPHIC DISPLAY OF SERIES FOR VARIABLE Y
DATA – *
MEAN – .

OBS.	DATA
1	90764.0
2	81666.0
3	39045.0
4	29065.0
5	20181.0
6	25291.0
7	33898.0
8	33830.0
9	49517.0
10	76224.0
11	59577.0
12	32152.0
13	9024.00
14	23863.0
15	24743.0
16	62590.0
17	37857.0
18	37630.0
19	61286.0
20	55244.0
21	75614.0
22	99069.0
23	107652.
24	100578.
25	148515.
26	149841.
27	364072.
28	232971.
29	220722.
30	215575.
31	209197.
32	207171.
33	253369.
34	285948.
35	216455.
36	199038.
37	266555.
38	283319.
39	294118.
40	316226.
41	318507.
42	374784.
43	406295.
44	431270.
45	479339.
46	645182.



```
OBS.     DATA                  -2.0E 05   3.0E 05   8.0E 05   1.3E 06   1.8E 06
                          :---------:---------:---------:---------:---------:
  47   607755.            :                   .     *
  48   918777.            :                   .           *
  49   894573.            :                   .           *
  50   719040.            -                   .       *
  51   739319.            :                   .        *
  52   830236.            :                   .          *
  53   862520.            :                   .          *
  54   812392.            :                   .         *
  55   879368.            :                   .           *
  56   853555.            :                   .          *
  57  0.100506E 07        :                   .             *
  58   919671.            :                   .           *
```

MEAN VALUE OF THE PROCESS
0.30737E 06

STANDARD DEVIATION OF THE PROCESS
0.30392E 06

3- IDENTIFICATION

We begin with the analysis of the autocorrelation function using the undifferenced data to decide on the stationarity of the time series. To decide this we look at the estimated ACF[1]. For nonstationary data the auto-correlations are typically significantly different from zero for the first several time lags, and only gradually drop to zero or show a spurious pattern as the number of time periods increases.

Figure 2 shows the estimated ACF for the undifferenced data. The auto-correlations for the first twelve time lags are significantly different from zero, and the pattern gradually drops to zero rather than dropping to zero exponentially. A straight line would fit the points extremely well. This indicates that the mean of the data is non-stationary and that nonseasonal differencing is required. We will see later if estimation stage results confirm the need for differencing.

---

(1) ACF = autocorrelation function

Assuming for the moment that differncing is not needed. We proceed to examine the PACF[1] for the undifferenced data in Figure 3. We identify an AR(1) model. This is consistent with the combination of a decaying pattern in the estimated ACF and the cutoff to zero after lag 1 in the estimated PACF. Therefore, we identify the following ARIMA(1,0,0) model

$$(1 - \phi_1 B)\ \bar{Z}_t = a_t$$

where $\bar{Z}_t = Z_t - \mu$ and μ is the mean of the original time series.

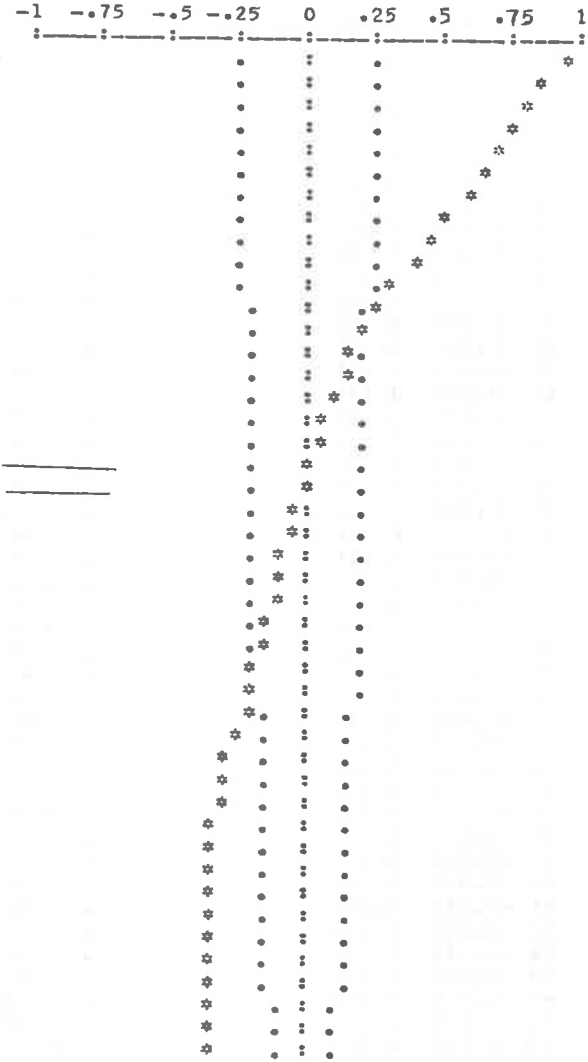
In other words the identified model in a difference equation form is given by

$$Z_t = (1 - \phi_1) + \phi_1 Z_{t-1} + a_t$$

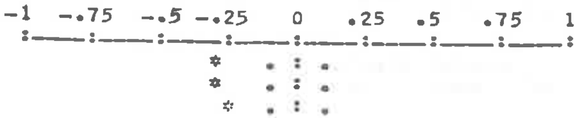
(1) PACF = partial autocorrelation function

AUTOCORRELATION FUNCTION FOR VARIABLE Y
AUTOCORRELATIONS *
TWO STANDARD ERROR LIMITS .

LAG	AUTO. CORR.	STAND. ERR.
1	0.933	0.126
2	0.866	0.125
3	0.814	0.123
4	0.758	0.122
5	0.702	0.121
6	0.639	0.120
7	0.575	0.119
8	0.520	0.117
9	0.469	0.116
10	0.387	0.115
11	0.308	0.114
12	0.252	0.112
13	0.193	0.111
14	0.157	0.110
15	0.126	0.109
16	0.091	0.107
17	0.061	0.106
18	0.038	0.104
19	0.016	0.103
20	-0.011	0.102
21	-0.028	0.100
22	-0.057	0.099
23	-0.078	0.097
24	-0.099	0.096
25	-0.125	0.094
26	-0.149	0.093
27	-0.162	0.091
28	-0.181	0.090
29	-0.202	0.086
30	-0.220	0.086
31	-0.250	0.085
32	-0.288	0.083
33	-0.305	0.081
34	-0.319	0.080
35	-0.331	0.078
36	-0.341	0.076
37	-0.348	0.074
38	-0.352	0.072
39	-0.356	0.070
40	-0.360	0.068
41	-0.358	0.066
42	-0.355	0.063
43	-0.353	0.061
44	-0.344	0.059
45	-0.331	0.056



LAG	AUTO. CORR.	STAND. ERR.
46	-0.306	0.054
47	-0.289	0.051
48	-0.263	0.048



```
PARTIAL AUTOCORRELATION FUNCTION FOR VARIABLE Y
PARTIAL AUTOCORRELATIONS *
TWO STANDARD ERROR LIMITS .

      PR-AUT  STAND.
 LAG   CORR.    ERR.   -1  -.75 -.5 -.25   0   .25  .5   .75   1
                       :----:----:----:----:----:----:----:----:
   1   0.933   0.131                  .    :    .             *
   2  -0.032   0.131                  .   *:    .
   3   0.084   0.131                  .    : *  .
   4  -0.067   0.131                  .   *:    .
   5  -0.012   0.131                  .    *    .
   6  -0.100   0.131                  .  * :    .
   7  -0.039   0.131                  .   *:    .
   8   0.011   0.131                  .    *    .
   9   0.002   0.131                  .    *    .
  10  -0.270   0.131                  *    :    .
  11  -0.030   0.131                  .   *:    .
  12   0.087   0.131                  .    : *  .
  13  -0.063   0.131                  .   *:    .
  14   0.164   0.131                  .    :  * .
  15   0.020   0.131                  .    *    .
  16  -0.024   0.131                  .    *    .
  17  -0.040   0.131                  .   *:    .
  18   0.019   0.131                  .    *    .
  19   0.030   0.131                  .    :*   .
  20  -0.074   0.131                  .   *:    .
  21  -0.005   0.131                  .    *    .
  22  -0.109   0.131                  .  * :    .
  23  -0.031   0.131                  .   *:    .
  24  -0.093   0.131                  .  * :    .
  25   0.022   0.131                  .    *    .
  26  -0.033   0.131                  .   *:    .
  27   0.070   0.131                  .    :*   .
  28  -0.070   0.131                  .   *:    .
  29  -0.007   0.131                  .    *    .
  30  -0.026   0.131                  .   *:    .
  31  -0.104   0.131                  .  * :    .
  32  -0.110   0.131                  .  * :    .
  33   0.101   0.131                  .    : *  .
  34  -0.006   0.131                  .    *    .
  35  -0.015   0.131                  .    *    .
  36  -0.059   0.131                  .   *:    .
  37   0.023   0.131                  .    *    .
  38  -0.014   0.131                  .    *    .
  39  -0.051   0.131                  .   *:    .
  40   0.039   0.131                  .    :*   .
  41   0.072   0.131                  .    :*   .
  42  -0.152   0.131                  . *  :    .
  43  -0.041   0.131                  .   *:    .
  44   0.031   0.131                  .    :*   .
  45  -0.010   0.131                  .    *    .

      PR-AUT  STAND.
 LAG   CORR.    ERR.   -1  -.75 -.5 -.25   0   .25  .5   .75   1
                       :----:----:----:----:----:----:----:----:
  46   0.096   0.131                  .    : *  .
  47  -0.021   0.131                  .    *    .
  48   0.096   0.131                  .    : *  .
  49  -0.086   0.131                  .  * :    .
```

4- ESTIMATION

The whole results of estimation and diagnostic checking for the ARIMA (1,o,o) model are shown in Figure 4 (a,b, and c).

We begin with the estimation results in Figure 4-a. Since the mean of the series Z is equal to 307370 while its standard deviation is equal to 303920, which means that they are nearly equal, therefore using the t test[1] we accept the hypothesis that the actual mean, μ, of the original time series is not different from zero. Thus the constant term of the ARIMA (1,0,0) model, as given by

$\mu(1-\phi_1)$ is dropped, and only ϕ_1 is estimated.

The estimated model is therefore as follows:

$$Z_t = \underset{(52.125)}{1.0248}\, Z_{t-1} + a_t$$

The t value shown in parantheses under the estimated ϕ_1 coefficient indicates that the latter is significantly different from zero at all levels of significance. However, there are

(1) t = 307370/303920 = 1.01

two striking facts in the estimation results. First, the model does not satisfy the stationarity condition, since $\hat{\phi}_1 = 1.0248 > 1$. This confirms that the time series is non-stationary in the mean. Second, the estimated coefficient $\hat{\phi}_1 = 1.0248$ is virtually 1.0. Testing the null hypothesis that $\hat{\phi}_1 = 1$ gives the following t statistic

$$t = \frac{\phi_1 - 1}{S(\hat{\phi})}$$

$$= \frac{1.0248}{0.01966} = \frac{.0248}{.01966} = 1.26$$

Since $|t| < 2$, we fail to reject the hypothesis that $\phi_1 = 1$. This might indicate that taking a first difference of the original time series may be appropriate for inducing stationarity.

Since the ARIMA (1,0,0) model is not good in the sense of not satisfying the stationairty condition, no attempt, therefore, has been made to test the statistical adequacy of the model. the diagnostic cheking results shown in Figures 4-b and 4-c, however, are kept for completeness and for later reference.

VARIABLE Y CONTAINS THE TIME SERIES

DEGREE OF NONSEASONAL DIFFERENCING - 0

DEGREE OF SEASONAL DIFFERENCING - 0

SEASONAL SPAN - 1

MEAN VALUE OF THE PROCESS
0.30737E 06

STANDARD DEVIATION OF THE PROCESS
0.30392E 06

NONLINEAR ESTIMATION RESULTS

PAR	LAG	ESTIMATE	STD ERROR	T RATIO
AR	1	1.0248	0.19661E-01	52.125

COVARIANCE MATRIX OF THE ESTIMATES

PAR	LAG	
AR	1	0.38657E-03

CORRELATION MATRIX OF THE ESTIMATES

PAR	LAG	
AR	1	1.00000

MEAN VALUE OF RESIDUAL SERIES
0.48532E 04

STANDARD DEVIATION OF RESIDUAL SERIES
0.64498E 05

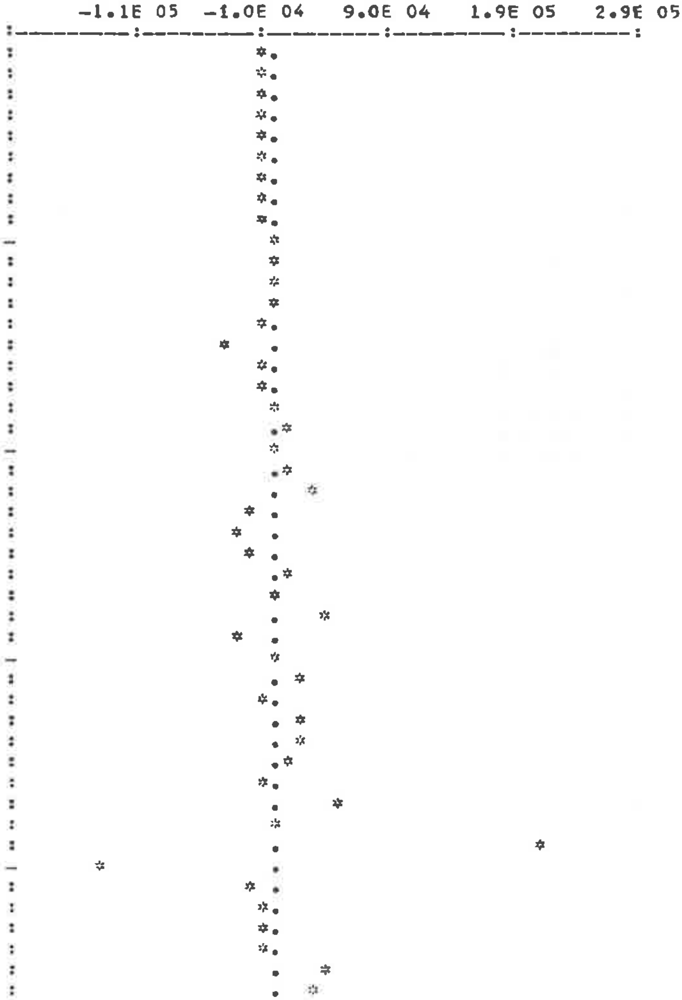
VARIANCE OF RESIDUAL SERIES
0.41600E 10

DIAGNOSTIC CHI-SQUARE STATISTICS FOR RESIDUAL SERIES OF VARIABLE Y

LAG	CHI-SQ	D.F	PROB
6	5.73	5	0.3332
12	13.46	11	0.2644
18	16.79	17	0.4687
24	27.67	23	0.2284
30	33.50	29	0.2581
36	37.44	35	0.3576
42	38.46	41	0.5843
43	39.39	47	0.7771
49	39.52	48	0.8033

GRAPHIC DISPLAY OF RESIDUAL SERIES FOR VARIABLE Y
DATA – *
MEAN – .

OBS.	DATA
1	-6130.31
2	-5981.69
3	-5836.63
4	-5695.13
5	-5557.06
6	-5422.38
7	-5290.94
8	-5162.63
9	-5037.44
10	-4915.31
11	-4796.13
12	-4679.88
13	-4566.44
14	-11353.2
15	-44650.1
16	-10950.1
17	-9606.18
18	4608.56
19	7978.59
20	-910.262
21	14846.4
22	25476.7
23	-18540.9
24	-28905.3
25	-23926.9
26	14614.8
27	287.078
28	37232.2
29	-26288.2
30	-1167.63
31	22721.0
32	-7564.77
33	18997.4
34	21576.3
35	6121.44
36	-9748.81
37	45437.9
38	-2364.13
39	210508.
40	-140147.
41	-18037.6
42	-10631.3
43	-11734.4
44	-7223.88
45	41050.4
46	26283.6



```
OBS.     DATA            -1.1E 05  -1.0E 04    9.0E 04   1.9E 05   2.9E 05
                    :---------:---------:---------:---------:---------:
  47 -76597.9       :            *       .
  48 -22795.3       :                  * .
  49  62571.5       :                    .     *
  50  10140.9       -                    .*
  51  3759.38       :                    *
  52  14800.1       :                    .*
  53 -5576.25       :                   *.
  54  48363.1       :                    .    *
  55  22198.8       :                    . *
  56  14879.8       :                    .*
  57  37353.3       :                    .   *
```

```
RESIDUAL AUTOCORRELATION FUNCTION FOR VARIABLE Y
AUTOCORRELATIONS *
TWO STANDARD ERROR LIMITS .

      AUTO. STAND.
LAG  CORR.   ERR. -1  -.75 -.5 -.25   0   .25   .5  .75   1
                   :----:----:----:----:----:----:----:----:
  1 -0.262  0.115                 *    :    .
  2 -0.051  0.114                 .   *:    .
  3 -0.025  0.114                 .   *:    .
  4  0.070  0.113                 .    :*   .
  5 -0.018  0.112                  .   *   .
  6  0.033  0.111                  .   :*  .
  7 -0.014  0.110                  .   *   .
  8 -0.077  0.109                  . * :   .
  9  0.165  0.108                  .   :  *.
 10 -0.179  0.107                  *   :   .
 11  0.147  0.106                  .   :  *.
 12 -0.072  0.105                  .  *:   .
 13 -0.100  0.104                  . * :   .
 14 -0.024  0.104                  .   *   .
 15  0.136  0.103                  .   :  *.
 16 -0.069  0.102                  .  *:   .
 17 -0.047  0.101                  .  *:   .
 18 -0.022  0.100                  .   *   .
 19  0.131  0.099                  .   :  *.
 20 -0.167  0.098                  .*  :   .
 21  0.239  0.097                  .   :   .*
 22  0.003  0.096                  .   *   .
 23 -0.050  0.094                  .  *:   .
 24 -0.050  0.093                  .  *:   .
 25  0.007  0.092                  .   *   .
 26  0.056  0.091                  .   :*  .
 27 -0.035  0.090                  .  *:   .
 28  0.019  0.089                  .   *   .
 29 -0.111  0.088                  . * :   .
 30  0.173  0.087                   .  :  *
 31 -0.149  0.086                   *  :  .
 32  0.069  0.085                   .  :* .
 33 -0.024  0.083                   .  *  .
 34 -0.026  0.082                   . *:  .
 35 -0.043  0.081                   . *:  .
 36 -0.008  0.080                   .  *  .
 37 -0.007  0.078                   .  *  .
 38  0.045  0.077                   .  :* .
 39  0.020  0.076                   .  *  .
 40 -0.056  0.075                   . *:  .
 41  0.015  0.073                   .  *  .
 42 -0.017  0.072                   .  *  .
 43 -0.027  0.070                   . *:  .
 44 -0.035  0.069                   . *:  .
 45 -0.039  0.068                   . *:  .

      AUTO. STAND.
LAG  CORR.   ERR. -1  -.75 -.5 -.25   0   .25   .5  .75   1
                   :----:----:----:----:----:----:----:----:
 46  0.002  0.066                   .  *  .
 47  0.032  0.065                   .  :* .
 48 -0.011  0.063                   .  *  .
 49 -0.023  0.061                    . * .
```

5- DIFFERENCING AND FURTHER IDENTIFICATION

To induce stationairty, we difference the series $\{Z_t\}$ until a stationary series, $\{W_t\}$ say, is obtained. We experiment with first differences on the basis of the estimation results of the last section.

The first regular differences of the time series are plotted in Figure 5 and appear to be stationary in the mean. This is confirmed by the estimated ACF of the first regular differences, which is given in Figure 6. In this figure the working series is given by

$$W_t = (1 - B) Z_t$$

i.e., the working series W_t consists of the first regular differences of the original time series.

The estimated ACF of the first differences has the following characteristics: (1) The auto-correlation coefficient at the first lag is nearly significantly different from zero, since it lies on the lower vertical axis of the 95%

confidence limits.[1] (2) All the autocorrelation coefficients lie within the confidence limits except the first auto-correlation coefficient which lies on the lower axis and the 21th autocorrelation coefficient which lies outside the confidence limits but near the upper axis. This may be attributed to chance. On the basis of these characteristics we may assume either that the first auto-correlation is different from zero, and in this case, the estimated ACF cuts off to zero after the first lag; or that the first autocorrelation coefficient is not statistically different from zero, and in this case, all the estimated autocorrelation coefficients lie between the two confidence limits except that one that is ubnormally large.

As regards the sample PACF of the first differences which is shown in Figure 7, it has exactly the same behavior as that of the sample ACF which is described above. Again, we may assume either that the first partial autocorrelation

---

(1) From figure 5, the t value of the first autocorrelation coefficeint is given by

$$t_{r_1} = \frac{r_1}{S(r_1)} = \frac{-.242}{.127} = -1.9$$

coefficient is different from zero, and in this case, the estimated PACF cuts off to zero after the first lag, or that the first partial autocorrelation coefficient is not significantly different from zero, and in that case, all the estimated partial autocorrelation coefficients lie between the confidence limits except that one that is exceptionally large.

The most plausible assumed behavior which is consistent with theory for the estimated ACF and PACF of the first differences is the one which is compatible with the assumption that all the estimated autocorrelation and partial autocorrelation coefficients are not significantly different from zero. According to theory, this leads to an ARIMA (0,1,0) random walk model. From figure 5, the sample mean of the first regular differences is 14542, which is not large compared with the sample standard deviation of the first regular differences, where the latter is equal to 71035. Therefore, no constant term is included in the model. Thus the identified random walk model is given by

$$W_t = a_t$$

where

$$W_t = (1 - B) Z_t$$

In other words, the identified random walk model in a difference equation form is as follows:

$$Z_t = Z_{t-1} + a_t$$

It is possible, however, to expand the random walk model and identify the following two further autoregressive and moving average models successively

$$(1 - \phi_1 B) W_t = a_t$$

and

$$W = (1 - \theta B) a_t$$

where

$$W_t = (1 - B) Z_t$$

In terms of the original values of the time series, the last two identified models are successively as follows:

$$Z_t = (\phi_1 + 1) Z_t - \phi_1 Z_{t-2} + a_t$$

$$Z_t = Z_{t-1} - \theta_1 a_{t-1} + a_t$$

The first ARIMA (1,1,0) model is consistent with a hypothesis of a significant spike at the first lag in the estimated PACF and no significant spikes in the ACF. For the second ARIMA (0,1,1) model, it is consistent with a hypothesis of a significant spike at the first lag in the ACF and no significant spikes in the PACF. This is admittedly not exactly the same as theory says. They are introduced, however, to increase the hope of obtaining a good model(1).

In summary, the identified models in this section are as follows:

ARIMA (0,1,0): $W_t = a_t$

ARIMA (1,1,0): $W_t = \phi_1 W_{t-1}\ a_t$

ARIMA (0,1,1): $W_t = a_t - \theta_1 a_{t-1}$

where

$W_t = (1 - B)\ Z_t$

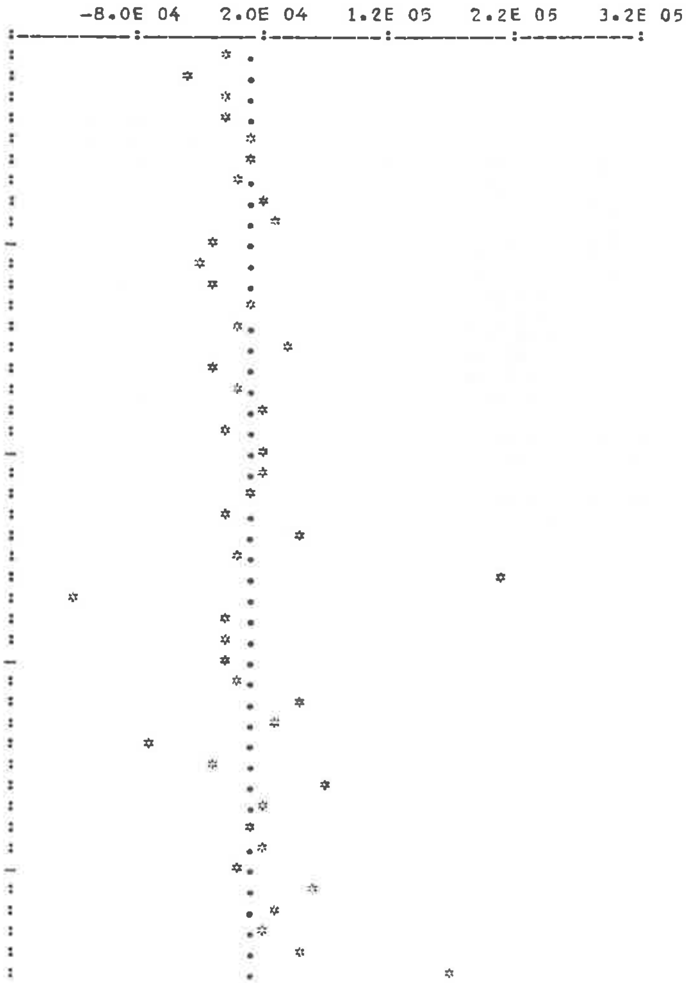
---

(1) Although it was possible to identify a further ARIMA-(1,1,1) model. This was not tried for two reawons. First, we have to wait to see if the pure AR and MA parts of this mixed model as represented by ARIMA (1,1,0) and ARIMA (0,1,1) models are statistically adequate at the estimation and diagnostic cheking stages. Second, the estimated ACF and PACF do not closely resemble their corresponding theoritical counterparts for the ARIMA (1,1,1) model.

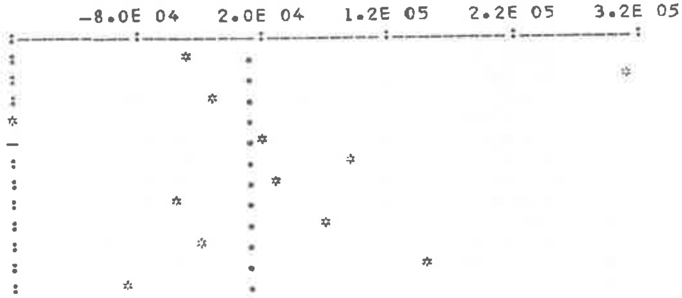
SAMPLE RUN FOR SPSS

GRAPHIC DISPLAY OF DIFFERENCED SERIES FOR VARIABLE Y
DEGREE OF NONSEASONAL DIFFERENCING - 1 DEGREE OF SEASONAL DIFFERENCING 0
DATA - *
MEAN - .

OBS.	DATA
1	-9098.00
2	-42621.0
3	-9980.00
4	-8884.00
5	5110.00
6	8607.00
7	-68.0000
8	15687.0
9	26707.0
10	-16647.0
11	-27425.0
12	-23128.0
13	14839.0
14	880.000
15	37847.0
16	-24733.0
17	-227.000
18	23656.0
19	-6042.00
20	20370.0
21	23455.0
22	8583.00
23	-7074.00
24	47937.0
25	1326.00
26	214231.
27	-131101.
28	-12249.0
29	-5147.00
30	-6378.00
31	-2026.00
32	46198.0
33	32579.0
34	-69493.0
35	-17417.0
36	67517.0
37	16764.0
38	10799.0
39	22108.0
40	2281.00
41	56277.0
42	31511.0
43	24975.0
44	48069.0
45	165843.



OBS.	DATA
46	-37427.0
47	311022.
48	-24204.0
49	-175533.
50	20279.0
51	90917.0
52	32284.0
53	-49628.0
54	66476.0
55	-25813.0
56	151505.
57	-85339.0

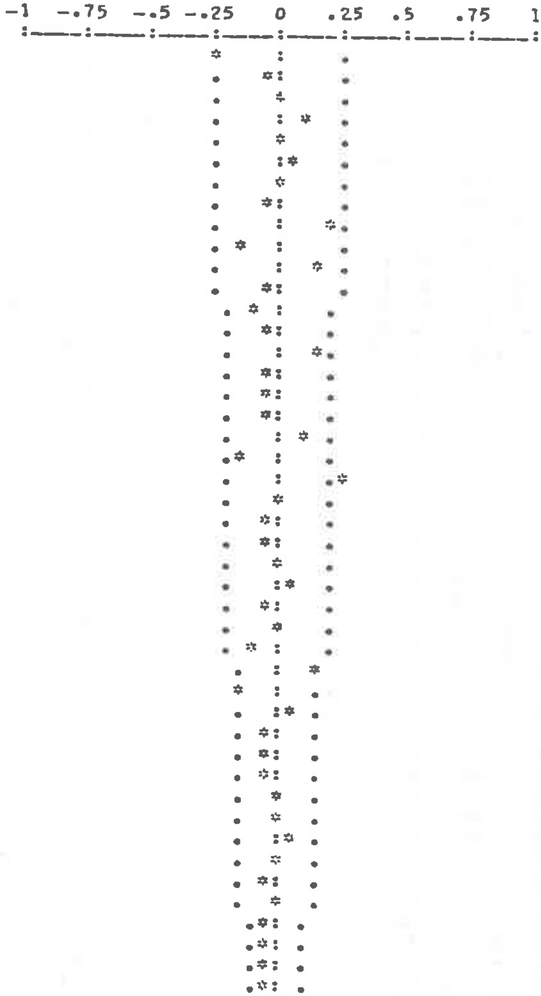


MEAN VALUE OF THE PROCESS
0.14542E 05

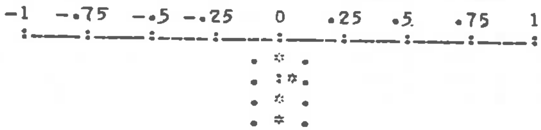
STANDARD DEVIATION OF THE PROCESS
0.71035E 05

CORRELATION FUNCTION FOR VARIABLE Y
CORRELATIONS *
STANDARD ERROR LIMITS

AUTO. CORR.	STAND. ERR.
0.242	0.127
0.042	0.126
0.015	0.124
0.076	0.123
0.008	0.122
0.043	0.121
0.006	0.119
0.068	0.118
0.181	0.117
0.158	0.116
0.150	0.114
0.066	0.113
0.099	0.112
0.026	0.110
0.135	0.109
0.070	0.108
0.051	0.106
0.025	0.105
0.124	0.103
0.170	0.102
0.231	0.101
0.006	0.099
0.058	0.098
0.050	0.096
0.004	0.094
0.063	0.093
0.035	0.091
0.017	0.090
0.111	0.083
0.169	0.086
0.144	0.084
0.064	0.083
0.027	0.081
0.037	0.079
0.055	0.077
-0.017	0.075
-0.018	0.073
0.032	0.071
0.001	0.069
-0.071	0.067
0.003	0.065
-0.026	0.062
-0.037	0.060
-0.042	0.057
-0.039	0.055



AUTO. CORR.	STAND. ERR.
-0.002	0.052
0.049	0.049
0.006	0.046
-0.020	0.042



```
PARTIAL AUTOCORRELATION FUNCTION FOR VARIABLE Y
PARTIAL AUTOCORRELATIONS *
TWO STANDARD ERROR LIMITS .

    PR-AUT STAND.
LAG  CORR.   ERR.  -1   -.75 -.5  -.25 0    .25  .5   .75  1
                   :----:----:----:----:----:----:----:----:
 1 -0.242  0.132                  *    :    .
 2 -0.107  0.132                  .  * :    .
 3 -0.056  0.132                  .   *:    .
 4  0.057  0.132                  .    :*   .
 5  0.025  0.132                  .    :*   .
 6  0.063  0.132                  .    :*   .
 7  0.027  0.132                  .    :*   .
 8 -0.065  0.132                  .   *:    .
 9  0.159  0.132                  .    :  * .
10 -0.099  0.132                  .  * :    .
11  0.123  0.132                  .    : *  .
12 -0.013  0.132                  .    *    .
13 -0.140  0.132                  . *  :    .
14 -0.071  0.132                  .   *:    .
15  0.065  0.132                  .    :*   .
16 -0.025  0.132                  .    *    .
17 -0.036  0.132                  .   *:    .
18 -0.086  0.132                  .  * :    .
19  0.161  0.132                  .    :  * .
20 -0.201  0.132                  .*   :    .
21  0.269  0.132                  .    :    *
22  0.091  0.132                  .    : *  .
23 -0.017  0.132                  .    *    .
24 -0.059  0.132                  .   *:    .
25 -0.041  0.132                  .   *:    .
26  0.002  0.132                  .    *    .
27 -0.014  0.132                  .    *    .
28 -0.036  0.132                  .   *:    .
29  0.019  0.132                  .    *    .
30 -0.080  0.132                  .  * :    .
31 -0.019  0.132                  .    *    .
32  0.019  0.132                  .    *    .
33 -0.011  0.132                  .    *    .
34 -0.016  0.132                  .    *    .
35 -0.015  0.132                  .    *    .
36 -0.111  0.132                  .  * :    .
37 -0.085  0.132                  .  * :    .
38  0.061  0.132                  .    :*   .
39  0.082  0.132                  .    : *  .
40 -0.112  0.132                  .  * :    .
41 -0.032  0.132                  .   *:    .
42 -0.068  0.132                  .   *:    .
43 -0.111  0.132                  .  * :    .
44 -0.069  0.132                  .   *:    .
45 -0.036  0.132                  .   *:    .

    PR-AUT STAND.
LAG  CORR.   ERR.  -1   -.75 -.5  -.25 0    .25  .5   .75  1
                   :----:----:----:----:----:----:----:----:
46  0.027  0.132                  .    :*   .
47 -0.033  0.132                  .   *:    .
48  0.050  0.132                  .    :*   .
49 -0.016  0.132                  .    *    .
```

6- FURTHER ESTIMATION

Estimation results for the ARIMA (1,1,0) and ARIMA (0,1,1) models identified in section 5, are given in figures 8-a and 9-a respectively. The estimated AR coefficient in the ARIMA (1,1,0) model ($\hat{\phi}$ =-0.19214) is not even significantly different from zero at 10% level of significance (t-value = -1.6059). The same is exactly true for the estimated MA coefficient in the ARIMA (0,1,1) model ($\hat{\theta}_1$ = 0.18799 and t-value =1.6011). This confirms the validity of the identification of the random walk model as a good representation of the unknown, underlying process of Hajj as well as a good respresentation of the available pilgrimage data. This is so because if either of the estimated coefficients of the ARIMA (1,1,0) and ARIMA (0,1,1) models is insignificantly different from zero, then both of the models reduce to a random walk (0,1,0) model.

Since the ARIMA (1,1,0) and ARIMA (0,1,1) models are not good in the sense of not having significantly estimated coefficients, there is no point, therefore, to test their statistical adequacy. The diagnostic checking results shown in figures 8-b and 8-c for the estimated ARIMA (1,1,0) model and in figures 9-b and 9-c for the estimated ARIMA (0,1,1) model are kept for the sake of completeness.

VARIABLE Y CONTAINS THE TIME SERIES

DEGREE OF NONSEASONAL DIFFERENCING - 1

DEGREE OF SEASONAL DIFFERENCING - 0

SEASONAL SPAN - 1

MEAN VALUE OF THE PROCESS
0.14542E 05

STANDARD DEVIATION OF THE PROCESS
0.71035E 05

NONLINEAR ESTIMATION RESULTS

PAR	LAG	ESTIMATE	STD ERROR	T RATIO
AR	1	-0.19214	0.11965	-1.6059

COVARIANCE MATRIX OF THE ESTIMATES

PAR	LAG	
AR	1	0.14316E-01

CORRELATION MATRIX OF THE ESTIMATES

PAR	LAG	
AR	1	1.00000

MEAN VALUE OF RESIDUAL SERIES
0.14585E 05

STANDARD DEVIATION OF RESIDUAL SERIES
0.63047E 05

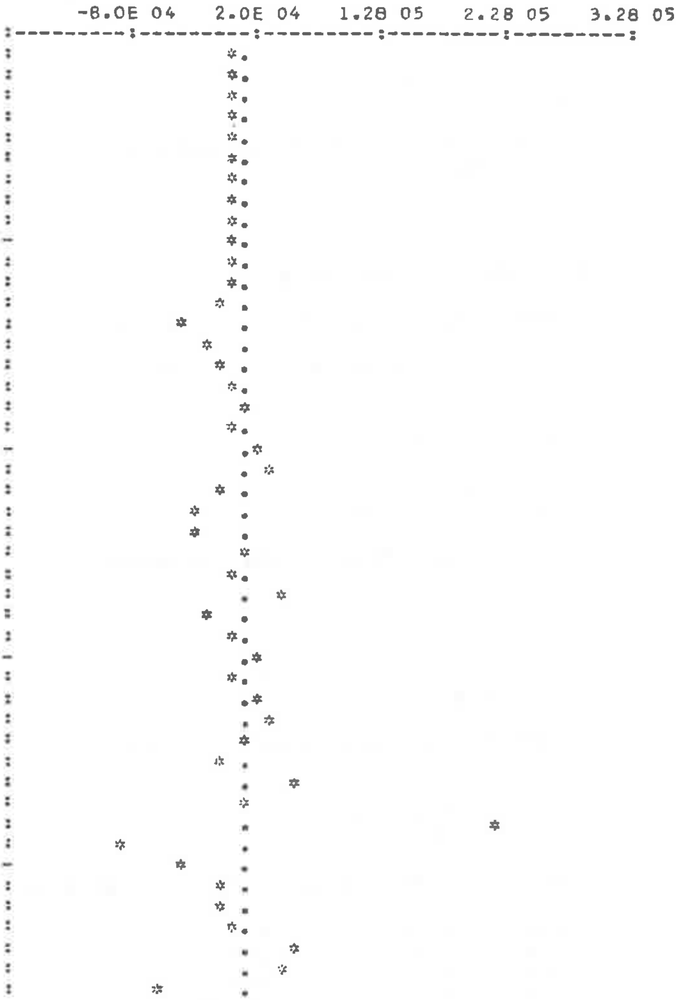
VARIANCE OF RESIDUAL SERIES
0.39749E 10

DIAGNOSTIC CHI-SQUARE STATISTICS FOR RESIDUAL SERIES OF VARIABLE Y

LAG	CHI-SQ	D.F	PROB
6	1.76	5	0.8817
12	6.51	11	0.8371
18	9.70	17	0.9158
24	18.23	23	0.7449
30	21.72	29	0.8315
36	24.97	35	0.8954
42	26.52	41	0.9610
48	29.15	47	0.9809
49	29.60	48	0.9830

GRAPHIC DISPLAY OF RESIDUAL SERIES FOR VARIABLE Y
DATA - *
MEAN - .

OBS.	DATA
1	-0.221895E-04
2	0.115483E-03
3	-0.601022E-03
4	0.312797E-02
5	-0.162792E-01
6	0.847238E-01
7	-0.440938
8	2.29482
9	-11.9432
10	62.1574
11	-323.493
12	1683.59
13	-8762.11
14	-44369.1
15	-18169.4
16	-10801.6
17	3402.99
18	9588.86
19	1585.79
20	15673.9
21	29721.2
22	-11515.4
23	-30623.6
24	-28397.6
25	10395.1
26	3731.23
27	38016.1
28	-17460.9
29	-4979.31
30	23612.4
31	-1496.63
32	19209.1
33	27369.0
34	13089.8
35	-5424.82
36	46577.8
37	10536.8
38	214486.
39	-89937.6
40	-37439.4
41	-7500.58
42	-7366.96
43	-3251.50
44	45808.7
45	41455.7
46	-63233.1



```
OBS.      DATA                  -8.0E 04   2.0E 04   1.2E 05   2.2E 05   3.2E 05
                        :---------:---------:---------:---------:---------:
 47   -30769.7          :              *   .
 48    64170.4          :                  .    *
 49    29737.0          :                  . *
 50    14020.1          -                  *
 51    24183.0          :                  .*
 52    6523.93          :                  *
 53    56715.3          :                  .    *
 54    42324.3          :                  .  *
 55    31029.7          :                  . *
 56    52867.8          :                  .   *
```

RESIDUAL AUTOCORRELATION FUNCTION FOR VARIABLE Y
AUTOCORRELATIONS *
TWO STANDARD ERROR LIMITS .

LAG	AUTO. CORR.	STAND. ERR.
1	-0.062	0.116
2	-0.074	0.115
3	0.004	0.114
4	0.097	0.113
5	0.028	0.112
6	0.062	0.112
7	0.009	0.111
8	-0.026	0.110
9	0.152	0.109
10	-0.888	0.108
11	0.153	0.107
12	-0.046	0.106
13	-0.100	0.105
14	-0.007	0.104
15	0.137	0.103
16	-0.049	0.102
17	-0.063	0.101
18	-0.005	0.100
19	0.104	0.099
20	-0.105	0.098
21	0.225	0.097
22	0.038	0.096
23	-0.072	0.095
24	-0.069	0.094
25	0.009	0.093
26	0.055	0.091
27	-0.027	0.090
28	-0.015	0.089
29	-0.086	0.088
30	0.130	0.087
31	-0.114	0.086
32	0.049	0.085
33	-0.025	0.083
34	-0.052	0.082
35	-0.068	0.081
36	-0.029	0.080
37	-0.014	0.078
38	0.044	0.077
39	-0.001	0.076
40	-0.072	0.074
41	-0.017	0.073
42	-0.040	0.071
43	-0.057	0.070
44	-0.068	0.069
45	-0.064	0.067

LAG	AUTO. CORR.	STAND. ERR.
46	-0.022	0.065
47	0.016	0.064
48	-0.022	0.062
49	-0.043	0.061

VARIABLE Y CONTAINS THE TIME SERIES

DEGREE OF NONSEASONAL DIFFERENCING - 1

DEGREE OF SEASONAL DIFFERENCING - 0

SEASONAL SPAN - 1

MEAN VALUE OF THE PROCESS
0.14542E 05

STANDARD DEVIATION OF THE PROCESS
0.71035E 05

SAMPLE RUN FOR SPSS

NONLINEAR ESTIMATION RESULTS

PAR	LAG	ESTIMATE	STD ERROR	T RATIO
MA	1	0.18799	0.11741	1.6011

COVARIANCE MATRIX OF THE ESTIMATES

PAR	LAG	
MA	1	0.13785E-01

CORRELATION MATRIX OF THE ESTIMATES

PAR	LAG	
MA	1	1.00000

MEAN VALUE OF RESIDUAL SERIES
0.14831E 05

STANDARD DEVIATION OF RESIDUAL SERIES
0.62511E 05

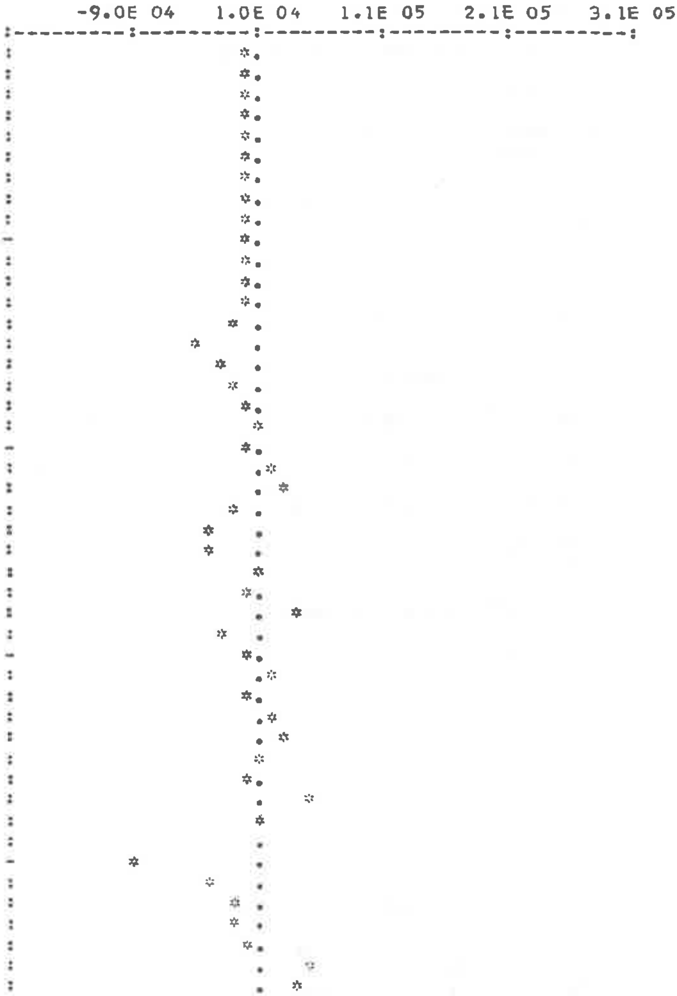
VARIANCE OF RESIDUAL SERIES
0.39076E 10

DIAGNOSTIC CHI-SQUARE STATISTICS FOR RESIDUAL SERIES OF VARIABLE Y

LAG	CHI-SQ	D.F	PROB
6	1.55	5	0.9070
12	6.71	11	0.8221
18	9.62	17	0.9186
24	18.14	23	0.7500
30	21.83	29	0.8272
36	25.21	35	0.8887
42	26.59	41	0.9601
48	29.39	47	0.9793
49	29.88	48	0.9813

GRAPHIC DISPLAY OF RESIDUAL SERIES FOR VARIABLE Y
DATA - *
MEAN - .

OBS.	DATA
1	0.0
2	0.0
3	0.0
4	0.0
5	0.0
6	0.0
7	0.0
8	0.0
9	0.0
10	0.0
11	0.0
12	0.0
13	3292.29
14	-8479.09
15	-44215.0
16	-18291.9
17	-12322.6
18	2793.50
19	9132.14
20	1648.73
21	15996.9
22	29714.2
23	-11061.1
24	-29504.3
25	-28674.4
26	9448.56
27	2656.21
28	38346.3
29	-17524.4
30	-3521.36
31	22994.0
32	-1719.41
33	20046.8
34	27223.5
35	13700.7
36	-4498.44
37	47091.3
38	10178.6
39	216144.
40	-90468.5
41	-29255.9
42	-10646.8
43	-8379.46
44	-3601.23
45	45521.0
46	41136.4



```
OBS.     DATA             -9.0E 04   1.0E 04   1.1E 05   2.1E 05   3.1E 05
                    :---------:---------:---------:---------:---------:
  47  -61759.9      :            *      .
  48  -29027.1      :               *   .
  49   62060.3      :                   .    *
  50   28430.6      -                   . *
  51   16143.6      :                   .*
  52   25142.8      :                   . *
  53   7007.53      :                   *
  54   57594.3      :                   .    *
  55   42338.0      :                   .  *
  56   32934.0      :                   . *
  57   54260.2      :                   .   *
```

RESIDUAL AUTOCORRELATION FUNCTION FOR VARIABLE Y
AUTOCORRELATIONS *
TWO STANDARD ERROR LIMITS .

```
                      -1  -.75  -.5 -.25   0   .25   .5   .75   1
                       :----:----:----:----:----:----:----:----:
```

LAG	AUTO. CORR.	STAND. ERR.
1	-0.067	0.115
2	-0.037	0.114
3	0.005	0.114
4	0.097	0.113
5	0.031	0.112
6	0.064	0.111
7	0.015	0.110
8	-0.026	0.109
9	0.163	0.108
10	-0.093	0.107
11	0.151	0.106
12	-0.046	0.105
13	-0.092	0.104
14	-0.006	0.104
15	0.134	0.103
16	-0.049	0.102
17	-0.054	0.101
18	-0.010	0.100
19	0.111	0.099
20	-0.106	0.098
21	0.226	0.097
22	0.032	0.096
23	-0.060	0.094
24	-0.062	0.093
25	0.005	0.092
26	0.054	0.091
27	-0.031	0.090
28	-0.006	0.089
29	-0.091	0.088
30	0.133	0.087
31	-0.117	0.086
32	0.047	0.085
33	-0.030	0.083
34	-0.052	0.082
35	-0.068	0.081
36	-0.031	0.080
37	-0.017	0.078
38	0.030	0.077
39	0.004	0.076
40	-0.070	0.075
41	-0.017	0.073
42	-0.042	0.072
43	-0.059	0.070
44	-0.071	0.069
45	-0.066	0.068

```
                      -1  -.75  -.5 -.25   0   .25   .5   .75   1
                       :----:----:----:----:----:----:----:----:
```

LAG	AUTO. CORR.	STAND. ERR.
46	-0.020	0.066
47	0.017	0.065
48	-0.026	0.063
49	-0.045	0.061

7- DIAGNOSTIC CHECKING

Since the identified model is the random walk model

$$Z_t = Z_{t-1} + a_t$$

it appears at once that the estimated residuals are given by the first differences of the original time series, i.e.,

$$a_t = Z_t - Z_{t-1}$$

Thus checking the statistical adequacy of the random walk model boils down to testing the statistical independence of the values of the first differences of the original time series through the application of the t or χ^2 test on their autocorrelation function.

Figures 5 and 6 given before show the first regular differences of the time series and their estimated ACF. The characteristics of this ACF which were discussed in section 5, and the estimation results presented in section 6, show that all the autocorrelation coefficients except the 21th autocorrelation coefficient[1] were not significantly differ-

---

(1) This as referred to in section 5 may be attributed to chance

ent from zero using the t test. Thus we conclude that the shocks of the random walk model are independent and that this model is a statistically adequate representation of the available data.

8- FORECASTING

Writing the random walk model, i.e., the ARIMA (0,1,0) model in a difference equation form, we obtain

$$Z_t = Z_{t-1} + a_t$$

From this form, we see that the one-step-ahead forecast (Z_t) for Hajj data is simply the last observed value: $Z_t = Z_{t-1}$. Thus the best forecast for the next year's Hajj (and the Hajj for all future years) is just simply the present year's Hajj. Since we do not know a or have an estimate of it, we set a_t to its expected value of zero. However, we do know the last observed Z_t(i.e.,Z_{t-1}) and that becomes our forecast.

The above model contains no constant term. This is because Z_t is not stationary; it is not varying around a fixed mean. Furthermore, the first differenced series ($W_t = Z_t - Z_{t-1}$) appears to have a fixed mean of about zero, so there is no deterministic (constant)trend element in Z_t.

9- CONCLUSIONS AND COMMENTS

This research has the following important findings and comments:

1- The results obtained indicate that the random walk model is a good representation of the unknown, underlying process of Hajj as well as a good representation of the actual available data on Hajj.

2- The one-step-ahead forecast for the next year's Hajj from the random walk model is simply the present year's Hajj.

3- The question of seasonality is out of consideration. The Box-Jenkins approach is not suitable for performing seasonality analysis on the available time series on Hajj. This is due to the fact that the span of the monthly seasonal cycle in Hajj is 32 or 33 years. The available data on Hajj which covers only 58 years are not sufficient to show the seasonal effects through the autocorrelation coefficients of the Box-Jenkins method.

4- The residual ACF in Figure 4-c for the AR(1) model is quite similar for the first differences in Figure 6. This should not be surprising. The residual ACF in Figure 4-c represents the behavior of the original data after they have been filtered through the operator $(1-\phi_1 B)$, with $\phi_1 = 1.0248$. The ACF in figure 6 represents the behavior of the original data after they have been filtered through the same operator $(1-\phi_1 B)$, but with $\phi_1=1$. since we have applied the same operator $(1-\phi_1 B)$ to the same data, and since ϕ_1 in this operator is nearly the same in both cases (1.0248 vs.1.0), we should expect the two resulting series to behave similarly. The estimated ACF's should look much the same.

REFERENCES

1- Box (G.E.P) and Jenkins (G.M.(1976): Time Series Analysis - Forecasting and Control (Holden day).

2- Nelson (C.R.)(1973): Applied Time Series Analysis for Managerial Forecasting (Holden day).

3- Pasha (H.A.) and Jamal (H.)(1983): Projections of the Number of Hajjis upto 1420 A.H., Volume 1, External Hajjis (Ummal-Qura University, Hajj Research Centre).

4- Pindyck (R.S.) and Rubinfeld (D.L.)(1976): Econometric Models and Economic Forecasts (Mc Graw-Hill).