LINEAR APPROXIMATING MODELS USING ORTHOGONAL FUNCTIONS: ANALYSIS OF DISCREPANCY

Fatma A. AbdelAty
Faculty of Commerce
Mansoura University

ABSTRACT

This paper presents an approach for model selection methods based on the Gauss discrepancy in the case of simple regression, replicated observations. The use of approximating mean functions which are linear combinations of orthogonal functions is considered, and the analysis is done using data compiled from monthly gross evaporation at Matatiele for the seasons 1937/38 to 1956/57 (South African Department of Water Affairs, 1957). The mean of the operating model (mean monthly gross evaporation) is represented in terms of Fourier series. Least-squares estimators of the parameters in the operating model are given. The contribution to the expected discrepancy of a parameter is estimated withou bias, the analysis of the criterion is done, and the model is selected. Also, a model for the mean gross evaporation per day in each month is selected. The observed and fitted means are given.

1. INTRODUCTION

A commen approach to model fitting is to select the family of models which is estimated to be the " most appropriate " in the circumstances, namely the background assumptions, the sample size, and the specific requirements of the user. Briefly, we begin by specifying in which sense the fitted model is required to best conform to the operating model, that is we specify a discrepancy which measures the lack of fit. The approximating family which minimizes an estimate of the expected discrepancy is selected. It is not assumed that this family contains the operating model.

The problem of model selection in regression analysis is discussed. We consider cases where we have more than one observation on the response variable, y, for each observed value of the variable x. In this case the basic difficulty of specifying a suitable operating family, that we encounter, in cases where we do not have more than one observation disappears. We can make assumptions about the form of the operating mean function such that it becomes possible to estimate the expected discrepancy. The operating model is then given by

$$y_{ij} = \mu(x_i) + e_{ij} \qquad i = 1, 2, \ldots, I$$
$$= \mu_i + e_{ij} \qquad j = 1, 2, \ldots, J_i, J_i > 1, \qquad (1.1)$$

e_{ij} independently and for each i identically distributed, $Ee_{ij} = 0$, Var $e_{ij} = \sigma_i^2$

Since at least 2I observations are available the parameters μ_i and σ_i^2 can be estimated.

In this paper we discuss selection methods based on the Gauss discrepanc First, some definitions relating to discrepancies are discussed in section 2. T use of mean functions which are linear combinations of orthogonal functions i discussed in section 3. Applications based on monthly data are available in section 4.

2. SOME DEFINITIONS RELATING TO DISCREPANCIES

Suppose that we have n independent observations on k variables and that each observation can be regarded as a realization of a k-dimensional random vector having distribution function F. Let M be the set of all k-dimensional distribution functions. Each member of M is a fully specified model.

A family of models, G_θ, $\theta \in \Theta$, is a subset of M whose individual members are identified by the vector of parameters $\theta = (\theta_1, \theta_2, \ldots, \theta_p)^T$.

A fitted model, $G_{\hat{\theta}}$, is a member of a family of models G_θ, $\theta \in \Theta$, which has been selected by estimating the parameters using the observations.

A discrepancy is a functional, Δ, on MxM which has the property

$$\Delta(G,F) \geq \Delta(F,F) \text{ for } G,F \in M. \tag{2.1}$$

The discrepancy between a member G_θ of an approximating family of models and the operating F will be denoted by

$$\Delta(\theta) = \Delta(\theta, F) = \Delta(G_\theta, F). \tag{2.2}$$

The discrepancy due to approximation between an approximating family, G_θ, $\theta \in \Theta$, and an operating model, F, is given by $\Delta(\theta_o)$, where

$$\theta_o = \arg\min \left\{ \Delta(\theta) : \theta \in \Theta \right\}. \tag{2.3}$$

We will usually assume that θ_o exists and is unique. The model G_{θ_o} is called the best approximating model for the family G_θ, $\theta \in \Theta$, and the discrepancy Δ

The discrepancy due to estimation is defined as $(G_{\hat{\theta}}, G_{\theta_o})$. It expresses the magnitude of the lack of fit due to sampling variation. The overall discrepancy is a random variable defined as $\Delta(\hat{\theta}) = \Delta(G_{\hat{\theta}}, F)$. Its distribution under the operating model determines the quality of a given procedure.

The expected discrepancy $E_F\Delta(\hat{\theta})$ depends on the operating model and its estimator is called a criterion.

A consistent estimator of $\Delta(\theta)$ is called an empirical discrepancy and is denoted by $\Delta_n(\theta)$. A suitable $\Delta_n(\theta)$ for $\Delta(\theta) = \Delta(\theta, F)$ is usually $\Delta(\theta, F_n)$, where F_n is the empirical distribution function, Boos (1981,1982).

3. THE USE OF APPROXIMATING MEAN FUNCTIONS

We assume that J observations are available for each value of x and that the operating model is

$$y_{ij} = \mu_i + e_{ij} \quad , \qquad i = 1, 2, \ldots, I, \quad j = 1, 2, \ldots, J, \tag{3.1}$$

$$Ee_{ij} = o \quad \text{and} \qquad E\bar{e}_{i.}\,\bar{e}_{j.} = \sigma_{ij}.$$

An important special case occurs if the e_{ij} are uncorrelated. In such a case $\sigma_{ij} = \delta_{ij}\, \sigma_i^2 / J$, where σ_i^2 is the variance of e_{ij}, Draper (1981).

The Gauss discrepancy is the square of a distance in R_I and there are systems of functions which are orthogonal under the corresponding inner product. A typical example are polynomials of degree 0, 1, 2, ..., I-1, denoted by $P_1, P_2, \ldots, P_I$, which are constructed so that

$$\sum_{r=1}^{I} P_i(x_r)\, P_j(x_r) = \delta_{ij}, \tag{3.2}$$

where the x_r are (equidistant) values of a variable x. For every subset of $P_1, P_2, \ldots, P_I$ an approximating model can be constructed with mean function

$$h(x, \theta^S) = \sum_{i \in s} \theta_i\, P_i(x) \, . \tag{3.3}$$

Here S denotes a subset of $\{1, 2, ..., I\}$ with $1 \leq P \leq I$ elements and θ^s has elements θ_i for $i \in S$ and zero for $i \notin S$. For P=1 we shall use θ for θ^s.

Hall (1983a) show how the most appropriate approximating family (i.e., the most appropriate set S) can be found by determining the contribution to the expected discrepancy of the individual parameters separately. To find the set S it is not necessary to calculate the criterion for each of the $2^I - 1$ different sets; it is sufficient to calculate the I contributions to the criterion of the parameters θ_i.

The orthogonal functions generate orthogonal base vectors (Linhart,1984a)

$$P_r = (P_r(x_1), ..., P_r(x_I))', \; r = 1, 2, ..., I, \tag{3.4}$$

which define a new coordinate system in R_I. If ζ denotes the coordinates of a point in the new system and z the coordinates of the point in the old system, we have

$$Z_i = \Sigma_r \; \zeta_r \; P_r(x_i) \text{ and } \zeta_i = \Sigma_r \; z_r \; P_i(x_r). \tag{3.5}$$

3.1 The Discrepancy Due to Approximation

Each mean vector

$$h(\theta^S) = (h(x_1, \theta^S), ..., h(x_I, \theta^S))' \tag{3.6}$$

defines a point in R_I and since

$$h(x_i, \theta^S) = \Sigma_{r \in S} \; \theta_r \; P_r(x_i) \tag{3.7}$$

this point has the coordinates θ^s in the new system. It follows immediately that this point is in the p-dimensional subspace spanned by the p vectors $\{P_i : i \in S\}$.

The coordinates of the operating mean in the new system are denoted by $\theta_o = (\theta_{o1}, \theta_{o2}, \ldots, \theta_{oI})'$:

$$\mu_i = h(x_i, \theta_o) = \Sigma_r \theta_{or} P_r(x_i),$$

$$\theta_{oi} = \Sigma_r \mu_r P_i(x_r). \tag{3.8}$$

Since the discrepancy used here is the squared Euclidean distance, the "best" θ^S, denoted by θ_o^S, is the projection of θ_o into the subspace spanned by $\{P_i : i \in S\}$, that is, the elements of θ_o^S are θ_{oi} for $i \in S$ and zero for $i \notin S$. The squared distance between θ_o and θ_o^S is then the discrepancy due to approximation:

$$\Delta(\theta_o^S) = \Sigma_{i \notin S} \theta_{oi}^2. \tag{3.9}$$

If an element r is added to S the discrepancy due to approximation decreases by θ_{or}^2. The contribution of each parameter to the discrepancy due to approximation can be indiviually assessed (Linhart, 1984b).

3.2 The Empirical Discrepancy and the Discrepancy due to estimation:

The discrepancy used here is the squared distance between the operating and approximating mean points in R_I:

$$\Sigma_i (\mu_i - h(x_i, \theta^S))^2 = \Sigma_{i \in S} (\theta_{oi} - \theta_i)^2 + \Sigma_{i \notin S} \theta_{oi}^2. \tag{3.10}$$

The vector $\bar{y} = (\bar{y}_{1.}, \dots, \bar{y}_{I.})'$ estimates μ. An empirical discrepancy is the squared distance between $\bar{y}$ and $h(\theta^S)$. If the coordinates of $\bar{y}$ in the ne system are denoted by $\hat{\theta}$, with elements

$$\hat{\theta}_i = \Sigma_r \bar{y}_{r.} P_i(x_r), \tag{3.11}$$

the empirical discrepancy can be expressed as

$$\sum_{i \in s} (\hat{\theta}_i - \theta_i)^2 + \sum_{i \notin s} \hat{\theta}_i^2 .$$

It can be seen that this is minimized by $\theta_i = \hat{\theta}_i$ for $i \in S$ (Sahler,1970). The minimum discrepancy estimator $\hat{\theta}^S$ has elements $\hat{\theta}_i$ for $i \in S$ and zero for $i \notin S$. In other words, $\hat{\theta}^S$ is the projection of $\hat{\theta}$ into the space spanned b $\{P_i : i \in S\}$.

The discrepancy due to estimation is the squared distance between θ_o^S and $\hat{\theta}_o^S$ and is given by

$$\sum_{i \in s} (\hat{\theta}_i - \theta_{oi})^2,$$

(Robertson, 1972).

3.3 The Expected Discrepancy:

The overall discrepancy is the squared distance between θ_o and $\hat{\theta}^S$:

$$\Delta(\hat{\theta}^S) = \sum_{i \in s} (\hat{\theta}_i - \theta_{oi})^2 + \sum_{i \notin s} \theta_{oi}^2 \tag{3.12}$$

and is the sum of the discrepancies due to estimation and due to approximation. The expected discrepancy is

$$E\Delta(\hat{\theta}^S) = \sum_{i \in S} \mathrm{Var}(\hat{\theta}_i) + \sum_{i \notin S} \theta_{oi}^2 . \tag{3.13}$$

By adding an element r to S the expected discrepancy changes by

$$E(\theta_{or} - \hat{\theta}_r)^2 - \theta_{or}^2 = \mathrm{Var}\,\hat{\theta}_r - \theta_{or}^2$$

$$= \sum_{ij} \sigma_{ij} P_r(x_i) P_r(x_j) - \theta_{or}^2 . \tag{3.14}$$

The best set S contains all elements r corresponding to parameters which lead to negative contributions.

3.4 The Criterion:

The contribution to the expected discrepancy of a parameter θ_r, that is $\mathrm{Var}\,\hat{\theta}_r - \theta_{or}^2$, is estimated without bias by

$$2\,\widehat{\mathrm{Var}}\,\hat{\theta}_r - \hat{\theta}_r^2,$$

where

$$\widehat{\mathrm{Var}}\,\hat{\theta}_r = \sum_{ij} \hat{\sigma}_{ij} P_r(x_i) P_r(x_j), \tag{3.15}$$

and $\hat{\sigma}_{ij}$ is an unbiased estimator of σ_{ij}. Only those parameters whose estimated contributions are negative should be fitted. So, we use a parameter θ_r only if

$$F = \frac{\hat{\theta}_r^2}{\widehat{\mathrm{Var}}(\hat{\theta}_r)} > 2. \tag{3.16}$$

When the e_{ij} are uncorrelated then:

$$\text{Var } \hat{\theta}_r = \Sigma_i \left(\frac{\sigma_i^2}{J} \right) P_r^2(x_i),$$

$$\widehat{\text{Var }} \hat{\theta}_r = \Sigma_i \left(\frac{\hat{\sigma}_i^2}{J} \right) P_r^2(x_i), \tag{3.17}$$

where $\hat{\sigma}_i^2$ is estimated by the sample variance of the observations y_{ij}, j= 1, 2 ..., J:

$$\hat{\sigma}_i^2 = \frac{\Sigma_i (y_{ij} - \bar{y}_{i.})^2}{J - 1} \tag{3.18}$$

If we assume that $\sigma_i^2 = \sigma^2$, then $\text{Var } \hat{\theta}_r = \sigma^2/J$ and we use the parameter θ_r only if

$$F = \frac{J \hat{\theta}_r^2}{\hat{\sigma}^2} = \frac{J \hat{\theta}_r^2}{MSE} > 2. \tag{3.19}$$

In analysis of variance $J \hat{\theta}_r^2$ is known as the sum of squares for θ_r and

$$F = \frac{J \hat{\theta}_r^2}{MSE} = \frac{SS \theta_r}{MSE}, \tag{3.20}$$

is the statistic used to test the hypothesis that $\theta_r = o$. Here we test a different hypothesis with the same statistic and we therefore require other critical values.

4. APPLICATION TO MONTHLY DATA

Table A.1 in the Appendix gives the monthly gross evaporation at Matatiele for the seasons 1937/38 to 1956/57. (Linhart, 1986). The season begins in October (month 1) and ends in Septemper (month 12). We represent these data using the operating model

$$y_{ij} = \mu_i + e_{ij} \quad , \qquad \begin{array}{l} i = 1, 2, \dots, 12, \\ j = 1, 2, \dots, 20, \end{array} \tag{4.1}$$

where y_{ij} is the gross evaporation (mm) in month i of year j (where j=1 represents 1937/38), and μ_i is the mean gross evaporation in month i. We assume that the e_{ij} are independently distributed with $Ee_{ij} = o$ and $\text{Var } e_{ij} = \sigma_i^2$.

We can estimate the mean monthly gross evaporation μ_i by $\bar{y}_{i.}$. However, with this type of data we can often improve the estimates by making use of the knowledge that the μ_i follow an approximately sinusoidal pattern. It is known that in such situations truncated Fourier series often lead to good approximation models.

We begin by representing the μ_i in terms of their Fourier series (chatfield 1982):

$$\mu_i = \alpha_o \; P_1(i) + \sum_{r=1}^{5} (\alpha_r P_{2r}(i) + \beta_r \; P_{2r+1}(i)) + \alpha_6 \; P_{12}(i) \tag{4.2}$$

where $\alpha^{'s}$ and $\beta^{'s}$ are the Fourier coefficients and

$$P_1(i) = (\frac{1}{12})^{1/2},$$

$$P_{2r}(i) = (\frac{2}{12})^{1/2} \cos w_r i,$$

$$P_{2r+1}(i) = (\frac{2}{12})^{1/2} \sin w_r i, \quad r = 1, 2, ..., 5, \tag{4.3}$$

$$P_{12}(i) = (\frac{1}{12})^{1/2} (-1)^i,$$

$$w_r = \frac{2\pi r}{12}$$

Least-squares estimators of the parameters in the operating model are given by

$$\hat{\alpha}_o = (\frac{1}{12})^{1/2} \Sigma_i \bar{y}_{i},$$

$$\hat{\alpha}_r = (\frac{2}{12})^{1/2} \Sigma_i \bar{y}_{i.} \cos w_r i,$$

$$\hat{\beta}_r = (\frac{2}{12})^{1/2} \Sigma_i \bar{y}_{i.} \sin w_r i, \quad r = 1, 2, ..., 5 \tag{4.4}$$

$$\hat{\alpha}_6 = (\frac{1}{12})^{1/2} \Sigma_i (-1)^i \bar{y}_{i.},$$

where the pairs of parameters (α_r, β_r), $r = 1, 2, ..., 5$, are considered jointly because they are coefficients belonging to the same frequency.

The pair (α_r, β_r) should be retained in the model if

$$2\,\widehat{\mathrm{Var}}\;\hat{\alpha}_r - \hat{\alpha}_r^2 + 2\,\widehat{\mathrm{Var}}\;\hat{\beta}_r - \hat{\beta}_r^2$$

$$= \frac{2\,\Sigma_i\, P_{2r}^2(i)\,\hat{\sigma}_i^2}{J} - \hat{\alpha}_r^2 + \frac{2\,\Sigma_i\, P_{2r+1}^2(i)\,\hat{\sigma}_i^2}{J} - \hat{\beta}_r^2 < o, \tag{4.5}$$

that is, if

$$\frac{J(\hat{\alpha}_r^2 + \hat{\beta}_r^2)}{\Sigma_i[\,P_{2r}^2(i) + P_{2r+1}^2(i)\,]\;\hat{\sigma}_i^2} > 2, \tag{4.6}$$

where J is the number of observations in each month; here J = 20.

Since $P_{2r}^2(i) + P_{2r+1}^2\;(i) = \frac{2}{12}$ for all i, j, it follows that the denominator is twice the residual mean square

$$\mathrm{MSE} = \frac{1}{12}\Sigma_i\,\hat{\sigma}_i^2 = \frac{1}{12(J-1)}\Sigma_{ij}\,(y_{ij} - \bar{y}_{i.})^2. \tag{4.7}$$

The pair (α_r, β_r) should thus be retained in the approximating model if

$$F = \frac{J(\hat{\alpha}_r^2 + \hat{\beta}_r^2)/2}{\mathrm{MSE}} = \frac{\mathrm{MS}\;\alpha_r,\,\beta_r}{\mathrm{MSE}} > 2. \tag{4.8}$$

The parameters α_o and α_6 have to dealt with separately. There is usually no question of omitting α_o from the approximating model. On the other hand, α_6 should be retained if

$$F = \frac{J\,\hat{\alpha}_6^2}{\mathrm{MSE}} = \frac{\mathrm{MS}\;\alpha_6}{\mathrm{MSE}}\,2. \tag{4.9}$$

The estimates of the parameter in the operating model are given by

$$\hat{\alpha}_0 = 464.348,$$

$$\begin{aligned} \hat{\alpha}_1 &= 6.406, & \hat{\beta}_1 &= 133.728, \\ \hat{\alpha}_2 &= 7.583, & \hat{\beta}_2 &= 1.714, \\ \hat{\alpha}_3 &= 7.001, & \hat{\beta}_3 &= -10.043, \\ \hat{\alpha}_4 &= 4.421, & \hat{\beta}_4 &= 1.998, \\ \hat{\alpha}_5 &= -8.692, & \hat{\beta}_5 &= 1.145, \\ \hat{\alpha}_6 &= 2.526 . \end{aligned} \quad (4.10)$$

The basic analysis of variance results in MSE = 319.32.

Table 4.1. The analysis of the Criterion, with MSE = 319.32.

Parameter	d.f.	SS	F
α_0	1	4,312,381.302	13,504.889
α_1, β_1	2	358,484.296	561.325
α_2, β_2	2	1,847,586	2.893
α_3, β_3	2	2,997.517	4.694
α_4, β_4	2	470.745	.737
α_5, β_5	2	1,537.238	2.407
α_6	1	127.614	.399

According to the criterion shown in Table 4.1 only (α_4, β_4) and α_6 should be omitted from the approximating model.

The selected model is complex. In particular, the inclusion of the highrequency component corresponding to (α_5, β_5) is unexpected. However, since he data consist of monthly totals and the number of days in each month aries, it is probable that this high-frequency oscillation is present in the operating model. To correct for the effect of different numbers of days in each month we can divide each observation by the number of days in the corresponding month and select a model for the mean gross evaporation per lay in each month. So, we get the estimators

$$
\begin{aligned}
&\hat{\alpha}_0 = 15.277, && \\
&\hat{\alpha}_1 = 0.082, && \hat{\beta}_1 = 4.429, \\
&\hat{\alpha}_2 = 0.367, && \hat{\beta}_2 = -0.018, \\
&\hat{\alpha}_3 = 0.212, && \hat{\beta}_3 = -0.122, \qquad (4.11)\\
&\hat{\alpha}_4 = 0.062, && \hat{\beta}_4 = -0.148, \\
&\hat{\alpha}_5 = 0.047, && \hat{\beta}_5 = 0.017, \\
&\hat{\alpha}_6 = -0.023. &&
\end{aligned}
$$

Table 4.2. The analysis of the Criterion, with MSE = 0.343.

Parameter	d.f.	SS	F
α_0	1	4667.735	13,608.556
α_1, β_1	2	392.455	572.092
α_2, β_2	2	2.700	3.936
α_3, β_3	2	1.197	1.744
α_4, β_4	2	.518	.754
α_5, β_5	2	.049	.073
α_6	1	.011	.031

The selected model as shown in Table 4.2 has no high-frequency component and contains the five parameters α_o, α_1, β_1, α_2, and β_2. The observed and fitted means are given below and are illustrated in Figure 4.1.

	Observed	Fitted
Oct	5.19	5.41
Nov	5.86	5.91
Dec	6.16	6.07
Jan	5.89	5.89
Feb	5.38	5.35
Mar.	4.44	4.53
Apr	3.55	3.54
May	2.86	2.75
Jun	2.43	2.45
Jul	2.65	2.79
Aug	3.69	3.62
Sep	4.72	4.59

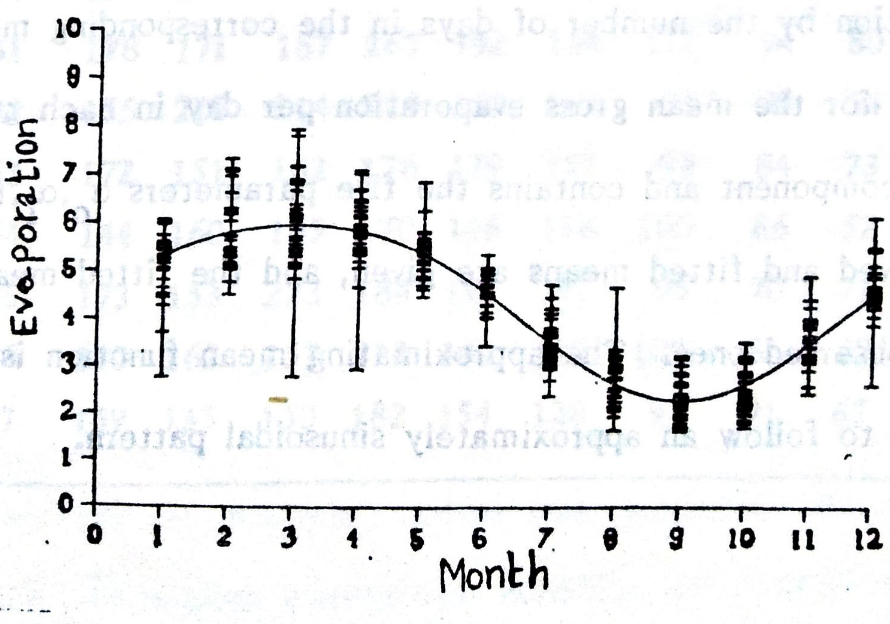

Figure 6.1. Monthly average evaporation per day at Matatiele, October (month1 1937 to 1957, and approximating mean function.

5. CONCLUTION

An approach to model selection methods for monthly data is presented in his paper, which uses operating model with mean functions follow an approxinately sinusoidal pattern, and the method of least squares to estimate the ourier coefficients. Only those parameters whose estimated contributions to he expected discrepancy are negative should be fitted.

According to the analysis of the criterion of the model using mean monhly gross evaporation, the high-frequency component corresponding to (α_5, β_5) hould be included, shich is unexpected. It is probable that this high-freuency oscillation is present in the operating model since the data consist f monthly totals and the number of days in each month varies.

To correct for the effect of different numbers of days in each month we livide each observation by the number of days in the corresponding month. The selected model for the mean gross evaporation per day in each month has no high-frequency component and contains the five parameters $\alpha_o, \alpha_1, \beta_1, \alpha_2$, and β_2. The observed and fitted means are given, and the fitted means are very close to the observed ones. The approximating mean function is illustrated and it seems to follow an approximately sinusoidal pattern.

APPENDIX

Table A.1. Monthly Gross Evaporation (mm) at Matatiele

Season	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP
1937/38	185	212	169	191	159	152	120	100	63	88	118	165
1938/39	165	199	192	170	129	140	133	98	69	62	123	122
1939/40	162	154	189	194	175	137	114	103	86	97	122	131
1940/41	172	187	177	189	146	141	126	144	97	96	123	184
1941/42	170	181	246	168	152	139	108	85	95	90	99	122
1942/43	133	151	170	165	157	124	102	70	85	99	96	136
1943/44	172	164	160	197	136	139	123	104	79	100	152	135
1944/45	168	209	242	191	158	133	102	97	91	109	137	147
1945/46	173	214	222	157	156	132	110	73	70	89	124	163
1946/47	171	189	211	185	162	153	91	77	49	84	121	129
1947/48	170	162	164	144	136	106	81	60	66	61	112	164
1948/49	171	181	207	208	136	137	126	91	71	87	132	137
1949/50	157	162	195	178	156	137	88	67	72	74	81	141
1950/51	178	171	187	167	192	144	112	94	80	83	97	134
1951/52	155	220	214	220	150	153	95	92	54	56	107	141
1952/53	172	151	182	176	129	157	99	84	73	66	91	128
1953/54	144	160	165	170	146	138	100	86	52	84	110	163
1954/55	173	153	223	184	141	137	95	70	71	69	120	133
1955/56	140	162	157	218	144	128	109	86	69	76	121	129
1956/57	149	135	150	182	154	128	97	91	67	73	106	128

REFERENCES

- Bishop, Y.M.M., S.E.Fienberg, and P.W.Holland (1975). Discrete Multivariate Analysis MIT Press, Cambridge,MA.

- Bloomfield, P.(1973). An exponential model for the spectrum of a scalar time series. Biometrika, 60, 217-226.

- Boos, D.D. (1981). Minimum distance estimators for location and goodness of fit .J. Am .Statist. Assoc., 76, 663-670.

- Boos, D.D. (1982). Minimum Anderson-Darling estimation. Commun. Statist Theor. Meth., 11, 2747-2774.

Chatfield, C. (1982). The analysis of Time Series: An Introduction. Champan and Hall.

Draper, N.R. and H. Smith (1981). Applied Regression Analysis, 2nd ed. Wiley, New York.

Geisser, S. and W.F. Eddy (1979). A predictive approach to model selectic .J. Am. Statist. Assoc., 74, 153-160.

Hall, P. (1983a). Orthogonal series methods for both qualitative and quantitative data .Ann. Statist., 11, 1004-1007.

Linhart, H. and W. Zucchini (1984a). An orthogonal Class of models for time series. S. Afr. Statist. J., 18, 59-67.

Linhart, H. and W. Zucchini (1985). Selection of approximating models by chi-squared discrepancies if the operating model is multinomial. Statis. & Dec., Suppl. Issue No. 2, 375-380.

Linhart, H. and P. Volkers (1985). On a criterion for the selection of models for stationary time series. Metrika, 32, 181-196.

Roberston, C.A. (1972). On minimum discrepancy estimators. Sankhya A, 34, 133-144.

Sahler, W. (1970). Estimation by minimum-discrepancy methods. Metrika, 16, 85-106.

Serfling, R.J. (1980). Approximating Theorems of Mathematical Statistics Wiley, New York.

Sparks, R.S., W. Zucchini, and D. Coutsourides (1985). On Variable Selec tion in multivariate regression. Comm. Statist. - Theor. Meth., A14, 1569-1587.