Statistics



Netherlands School of Economics

ECONOMETRIC INSTITUTE

Report 7109

# A PRIORI FIXED COVARIANCE MATRICES OF DISTURBANCE ESTIMATORS

by C. Dubbelman

August 2, 1971

Preliminary and Confidential

# A PRIORI FIXED COVARIANCE MATRICES OF DISTURBANCE ESTIMATORS

by C. Dubbelman

Contents



SUMMARY

This article is a sequel to articles by Abrahamse and Koerts [1971] and Abrahamse and Louter [1971]. An a priori fixed covariance matrix of the disturbance estimator in the linear model is established on empirical data from the field of economic time series analysis. The empirical results agree with theoretical results of spectral analysis. In practice this appears to be too great a burden for a single fixed covariance matrix to satisfy in all test cases. Selection of a matrix from a given set of fixed covariance matrices according to a device generally gives very good results.

## 1. INTRODUCTION

Consider the linear model

$$y = X\beta + u \tag{1.1}$$

where y is an n-element vector of values taken by the dependent variable, X is an $n \times k'$ matrix of n values taken by each of the $k'$ explanatory variables, $\beta$ is a column vector of $k'$ unknown parameters, and u is an n-element column vector of non-observable disturbances. It is assumed that X has rank k, $k \leq k'$, and the distribution of u is $N(0, \Gamma)$, where $\Gamma$ is an $n \times n$ symmetric positive definite matrix. Let w be a linear unbiased estimator of u, so that $w = B'y$ with $B'$ independent of y and $B'y = B'u$, and regard a test statistic T of the form

$$T = \frac{w'Aw}{w'Cw} \tag{1.2}$$

where A and C are arbitrary $n \times n$ real symmetric nonzero matrices. The distribution function $F(t)$ of T is

$$F(t) = Pr[\frac{w'Aw}{w'Cw} \leq t] = Pr\ [u'B(A - tC)B'u \leq 0]$$

Write $\Gamma = SS'$ and define $v = S^{-1}u$, then v is $N(0, I)$ and

$$F(t) = Pr\ [v'S'B(A - tC)B'Sv \leq 0]$$

Write $S'B(A - tC)B'S = LDL'$ with $L' = L^{-1}$ and D diagonal, and define $z = L'v$, then z is $N(0, I)$ and

$$F(t) = Pr\ [v'LDL'v \leq 0] = Pr\ [z'Dz \leq 0]$$

or finally,

$$F(t) = Pr\ [\sum_{i=1}^{n} d_i z_i^2 \leq 0] \tag{1.3}$$

where $d_i$, $i = 1, 2, \ldots, n$, are the eigenvalues of $S'B(A - tC)B'S$ and the $z_i^2$ are independent $\chi^2(1)$ variables.

Since the nonzero eigenvalues of $S'B(A - tC)B'S$ are equal to those of $(A - tC)B'\Gamma B$, it is clear that the distribution of T is independent of X if and only if $B'\Gamma B$ is independent of X. Abrahamse and Koerts [1971] have solved the problem of finding an estimator w, which fulfills this independence. Notice that $B'\Gamma B$ is the covariance matrix of $W = B'u$. Abrahamse and Koerts considered the following problem. Let $w = B'y$ be a linear unbiased estimator of the disturbances in the linear model (1.1) with $\Gamma = \sigma^2 I$, such that $B'B = \Omega \equiv KK'$, $K'K = I_{(n-k)}$, which minimizes $E(w - u)'(w - u)$. The solution is

$$(1.4) \qquad w = B'y = K(K'MK)^{-\frac{1}{2}}K'My$$

where

$$(1.5) \qquad M = I - RR'$$

R is an orthonormal basis of $M(X)$, the space spanned by the columns of X. The more general estimation problem, where $\Gamma$ is not assumed to be equal to $\sigma^2 I$, where $\Omega$ is not assumed to be idempotent, and where $E(w - u)'(w - u)$ is replaced by $E(w - u)'Q(w - u)$ with Q positive definite, is solved in Dubbelman, Abrahamse and Koerts (1970).

The crucial point is the choice of $\Omega$. The purpose of this paper is to establish some idempotent $\Omega$ for some branch of science, namely economic time series analysis, on the basis of empirical data. In the following sections a "best criterion" for $\Omega$ is chosen, and a measure is derived for handling the data. The empirical results are idealized and generalized. A selection device is introduced to choose one $\Omega$ from a set of admissible $\Omega$'s. The estimator w, defined in (1.4), is used in tests for autocorrelation and heterovariance, and the powers are compared with powers of the exact Durbin-Watson test and other tests.

We denote the $n \times k$ orthonormal complement to K by P, thus having $\Omega = I - PP'$. An expression for w in X and P only is derived in the appendix.

## 2. MEASURES FOR LEAST-SQUARES APPROXIMATION

Suppose that a test, using a test statistic T as in (1.2) with w defined in (1.4), is such that the null hypothesis, $E(uu') = \sigma^2 I$, is rejected when T takes a value below some critical point $t_\alpha$, and is not rejected in the opposite case. Let the alternative hypothesis be $E(uu') = \sigma^2\Gamma$. The significance level of the test is $\alpha = F(t_\alpha)$, and $t_\alpha$ depends on $\Omega$ via the eigenvalues of $K'(A - tC)K$. The power of the test is

$$F(t_\alpha/\Gamma) = \Pr\left[\sum_{i=1}^{n} d_i' z_i^2 \leq 0\right]$$

where the $d_i'$ are the eigenvalues of $(A - t_\alpha C)B'\Gamma B$. From the point of view of the theory of statistical hypothesis testing, it is natural to call that $\Omega$ best, for which the power is maximal, given a fixed significance level $\alpha$. Though $F(t)$ is independent of X, $F(t/\Gamma)$ is not, so that maximal power for varying X would lead to varying $\Omega$. Even if the variation from X to X is negligible, derivation of $\Omega$ by maximization of a power function generally involves extremely complicated mathematics. We therefore abandon the power maximization criterion, and return to the estimation problem of w, with the hope that good estimators give good testing results.

The objective function $E(w - u)'(w - u)$ is uniquely minimized by w, given $\Omega$, the minimum value being (apart from the multiplicative scalar $\sigma^2$)

$$\mathrm{tr}\ (\Omega) + \mathrm{tr}\ (I_{(n)}) - 2\mathrm{tr}\ (K'MK)^{\frac{1}{2}} = n - k + n - 2[n - 2k + \sum_{i=1}^{k} d_i^{\frac{1}{2}}] = 3k - 2\sum_{i=1}^{k} d_i^{\frac{1}{2}}$$

where $d_i$, $0 < d_i \leq 1$, $i = 1, 2, \ldots, k$ are the eigenvalues of $P'RR'P$, see the appendix. We define

$$\psi = k - \sum_{i=1}^{k} d_i^{\frac{1}{2}}$$

and we have $0 \leq \psi < k$, while $\psi = 0$ only if $PP' = RR'$, or $\Omega = M$. Therefore when comparing alternative specifications of $\Omega$ with respect to a particular X, the specification which makes $\psi$ minimal gives the best estimator w, according to the estimation criterion. An alternative measure is defined as

$$\phi = k - \sum_{i=1}^{k} d_i$$

It measures the sum of squared differences between the elements of M and $\Omega$, this sum being $\mathrm{tr}\ [(M - \Omega)(M - \Omega)] = 2\phi$. Also, $0 \leq \phi < k$.

The measures of least-squares approximation $\psi$ and $\phi$ seem very alike, and in fact behave very alike. The latter is preferred to the former because of its manageability:

$$\phi = k - \text{tr}\,(P'RR'P) = \text{tr}\,(P'MP)$$

It can be interpreted as the sum of squared cosines of all angles between the column vectors of K on the one hand and those of R on the other hand, tr (K'RR'K). A $\phi$ equal to zero means that all cosines are zero, or $M(K) \perp M(R)$, or equivalently, $M(P) = M(R)$.

## 3. A BEST Ω FROM EMPIRICAL DATA

Suppose $X_1, X_2, \ldots, X_N$ is the (hypothetical) population of all X matrices with rank k and n rows, occurring in some branch of research. Let Z* denote the matrix consisting of all N orthonormal bases to the X-matrices,

$$Z^* = [R_1 \vdots R_2 \vdots \ldots \vdots R_N]$$

We want $\Omega = I - PP'$ with PP' close to $R_iR_i'$, $i = 1, 2, \ldots, N$. Write

$$\phi_i = \text{tr}\,[P'(I - R_iR_i')P]$$

and adopt minimization of $\bar{\phi} = \frac{1}{N}\sum_{i=1}^{N} \phi_i$ as a criterion for finding the best Ω,

$$\bar{\phi} = \frac{1}{N}\sum_{i=1}^{N} \text{tr}\,[I_{(k)} - P'R_iR_i'P] = k - \frac{1}{N}\,\text{tr}\,[P'Z^*Z^{*\prime}P]$$

Hence, we look for P which maximizes tr $[P'Z^*Z^{*\prime}P]$, with the provision $P'P = I_{(k)}$. This is the well-known principal components problem, see Anderson [1958], Chapter 11. Writing $\lambda_i^*$ for the i-th eigenvalue of $Z^*Z^{*\prime}$, $\lambda_1^* \geq \lambda_2^* \geq \ldots \geq \lambda_n^* \geq 0$, and $h_i^*$ for the eigenvector of $Z^*Z^{*\prime}$ defined by $\lambda_i^*$, gives $h_i^*$ as the i-th principal component of Z*. The solution to our problem of minimization of $\bar{\phi}$ is $P = [h_1^* \vdots h_2^* \vdots \ldots \vdots h_k^*]$

We collected sixty vectors, each consisting of $n = 15$ subsequent annual observations on economic variables. We combined two of them and added a constant term vector, so that we got $m = 30$ X-matrices with $k = 3$ columns. The data include both stock and flow variables, deflated and undeflated, price indices and ratios, and also logarithmic series. Every vector covers a 15-year period between 1920 and 1969, and contains either American, English or Dutch data. First and higher order differences are excluded. Then a matrix Z

$$Z = [R_1 \vdots R_2 \vdots \cdots \vdots R_{30}]$$

can be constructed, as indicated above. We computed all 15 eigenvalues $\lambda_i$ and eigenvectors $h_i$ of ZZ', $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{15} \geq 0$. Defining Z* as the hypothetical set of all 15 × 3 X-matrices occurring in economic time series analysis, we regard Z as a sample from Z*, and we regard $h_i$ as an estimate of $h_i^*$. The eigenvalue $\lambda_i$ is often interpreted as an indicator of the contribution of $h_i$ to the explanation of Z. A convenient definition of the percentage contribution of g to the explanation of Z appears to be

$$\frac{100g'ZZ'g}{\text{tr } [ZZ']g'g}$$

Then

$$\mu_i \equiv 100\lambda_i / \sum_{j=1}^{15} \lambda_j \tag{3.1}$$

is the percentage contribution of $h_i$ to the explanation of Z. In Table 1 the first six principal components of Z are presented, together with their explanatory contributions. The vectors are graphically displayed in Figure 1.

The empirical matrix P is $[h_1 \vdots h_2 \vdots h_3]$ and it explains 33.3 + 27.1 + 17.5 = 77.9 per cent of Z. This corresponds to

$$\bar{\phi} = k - \frac{1}{m} \text{tr } [P'ZZ'P] = \{1 - \frac{\text{tr } [P'ZZ'P]}{\text{tr } [ZZ']}\}k = \{1 - 0.779\}3 = 0.663 \tag{3.2}$$

Notice that 33.3 is a maximum for $\mu_i$, since for any real n-vector g we have

$$0 \leq g'M_i g = g'(I - R_i R_i')g$$

and hence

$$g'R_i R_i' g / g'g \leq 1$$

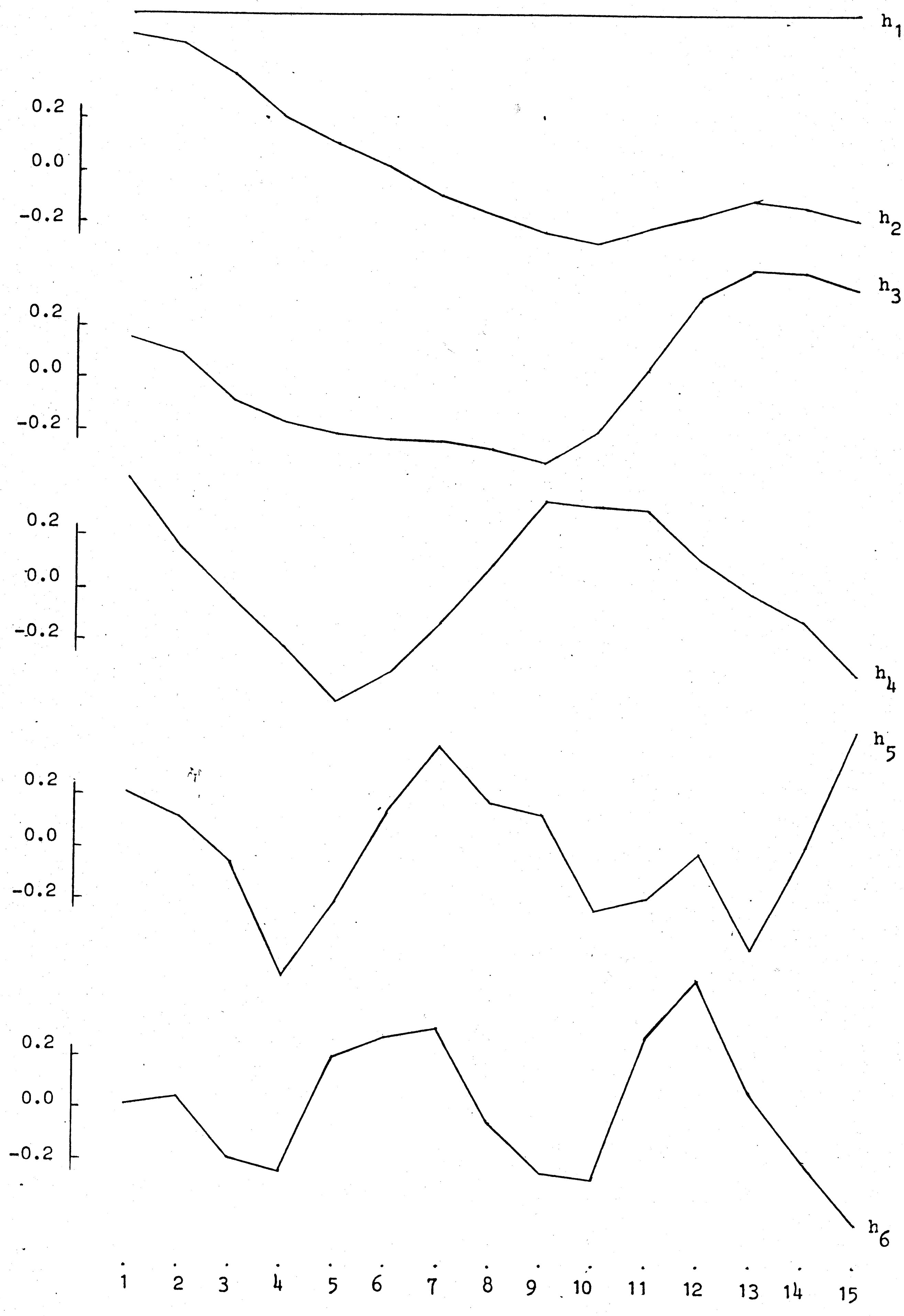
so that

$$\mu = \frac{100 g'ZZ'g}{tr\,[ZZ']g'g} = \frac{100 \sum_{i=1}^{m} g'R_i R_i' g}{tr\,[\sum_{i=1}^{m} R_i R_i']g'g} = \frac{100}{mk} \sum_{i=1}^{m} \frac{g'R_i R_i' g}{g'g} \leq \frac{100}{k}$$

TABLE 1. THE FIRST SIX PRINCIPAL COMPONENTS OF Z

| | $h_1$ | $h_2$ | $h_3$ | $h_4$ | $h_5$ | $h_6$ |
|---|---|---|---|---|---|---|
| 1 | 0.258 | 0.512 | 0.145 | 0.407 | 0.212 | 0.006 |
| 2 | 0.258 | 0.475 | 0.094 | 0.148 | 0.121 | 0.040 |
| 3 | 0.258 | 0.359 | 0.087 | -0.049 | -0.062 | -0.192 |
| 4 | 0.258 | 0.197 | -0.170 | -0.234 | -0.493 | -0.249 |
| 5 | 0.258 | 0.098 | -0.216 | -0.440 | -0.211 | 0.190 |
| 6 | 0.258 | 0.018 | -0.236 | -0.328 | 0.132 | 0.270 |
| 7 | 0.258 | -0.097 | -0.240 | -0.137 | 0.387 | 0.307 |
| 8 | 0.258 | -0.178 | -0.275 | 0.083 | 0.170 | -0.057 |
| 9 | 0.258 | -0.245 | -0.327 | 0.325 | 0.116 | -0.245 |
| 10 | 0.258 | -0.279 | -0.202 | 0.309 | -0.239 | -0.270 |
| 11 | 0.258 | -0.226 | 0.041 | 0.293 | -0.184 | 0.271 |
| 12 | 0.258 | -0.178 | 0.312 | 0.110 | -0.015 | 0.499 |
| 13 | 0.258 | -0.116 | 0.414 | -0.025 | -0.380 | 0.068 |
| 14 | 0.258 | -0.147 | 0.404 | -0.129 | -0.005 | -0.197 |
| 15 | 0.258 | -0.194 | 0.342 | -0.336 | 0.452 | -0.441 |
| $\mu_i$ | 33.3 | 27.1 | 17.5 | 8.5 | 4.1 | 3.6 |

FIGURE 1. DIAGRAM OF THE FIRST SIX PRINCIPAL COMPONENTS OF Z



# 4. IDEALIZATION OF THE EMPIRICAL RESULTS

We regard Z as a sample from Z*. It stands to reason that a different sample would yield different numerical outcomes. Both from this point of view and from the standpoint of tractability the vectors $h_i$ are streamlined in this section.

It is interesting to consider a successive computation of the principal components, which were actually computed simultaneously. To derive $h_1$ we write Z as

$$Z = h_1h_1'Z + (I - h_1h_1')Z = h_1h_1'Z + Z_1$$

The part $Z_1$ lies in $M(h_1)^{\perp}$, hence no part of $Z_1$ can be spanned, or explained, by $h_1$. Therefore we call $Z_1$ the remainder of Z after explanation by $h_1$. The method of principal components makes the sum of squared elements of $Z_1$ minimal, or $h_1'ZZ'h_1$ maximal. The solution is, that $h_1$ is the eigenvector of ZZ' defined by its largest eigenvalue. (We disregard multiple eigenvalues). The second principal component, $h_2$, is the eigenvector of $Z_1Z_1'$ defined by the largest eigenvalue of $Z_1Z_1'$, which is the second largest eigenvalue of ZZ'. Naturally $h_2'h_1 = 0$, since $h_2$ lies in $M(Z_1)$ and $Z_1$ lies in $M(h_1)^{\perp}$. In the same way $h_3$ results from $Z_2$, the remainder of Z after explanation by both $h_1$ and $h_2$,

$$Z_2 = (I - h_2h_2')Z_1 = (I - h_1h_1' - h_2h_2')Z$$

and so on.

If Z is not an ideal sample from Z* then each $h_i$ generally deviates from its ideal counterpart $h_i^*$. But that is not all. From the above considerations it is clear that $h_i$ is affected by all its predecessors $h_j$, $j = 1, 2, \ldots, i-1$, since $h_i$ is allowed to explain the part of Z that should have been explained by $h_1$ through $h_{i-1}$ and $h_i$ cannot explain its own part of Z insofar it has been explained already by its predecessors. This makes $h_i$ more and more unreliable as i increases.

Returning to Figure 1, we notice a remarkable regularity. The structure of $h_1$ is fixed, given the inclusion of a constant-term vector in all X-matrices. Column vector $h_2$ shows mainly a trend. There is no reason to assume that economic time series diverge from a perfect trend in the way $h_2$ does. As a first step, we idealize $h_2$ to a perfect trend, and $h_3$ through $h_{15}$ are replaced by the principal components of the remainder of Z after explanation by $h_1$ and the idealized $h_2$. The result is visualized in Figure 2 (again the first six vectors only). The explanatory contributions of these vectors are 33.3, 24.2, 20.0, 8.6, 4.1, and 3.8. A matrix P, consisting of the first three vectors, explains 77.5 per cent of Z, a loss of only 0.4 per cent compared with P based on Table 1. Therefore $h_2$ can be replaced by a perfect trend at very low cost. The third vector is no longer disturbed by its predecessors, provided that the idealization is correct. It shows a much more pronounced downward-upward movement, which is stylized in Figure 3. The figure is completed in the same sense as Figure 2. The explanatory contributions now are 33.3, 24.2, 18.1, 8.6, 5.0, and 4.1. When comparing Figures 1 and 3, we conclude that $(h_2, h_3)$ can be replaced at low cost as has been done in Figure 3, and that $h_4$, $h_5$, and $h_6$ are very stable, $h_5$ and $h_6$ being interchanged.

We discontinue this type of idealization. Some aspects of the principal components are illustrated and data for comparitive purposes are provided. In accordance with most economic theories, we look for curved idealizations. We considered sets of polynomials, see Abramovitz and Stegun [1965], p. 778. The Chebyshev polynomials look appropriate. Such a polynomial of degree k is defined as

$$T_k(x) = \cos\,(k \arccos x)$$

This polynomial is zero for $x_j = \cos\,[(j + \frac{1}{2})\pi/k]$, $j = 0, 1, 2, \ldots, k - 1$. It is known that Chebyshev polynomials are orthogonal in the following sense.

$$(4.1)\qquad \sum_{j=0}^{n-1} T_k(x_j)T_\ell(x_j) \begin{cases} = 0 & \text{if } k \neq \ell \\ = n & \text{if } k = \ell = 0 \\ = n/2 & \text{if } k = \ell \neq 0 \end{cases}$$

where $x_j$, $j = 0, 1, 2, \ldots, n - 1$, are the values for which $T_n(x)$ is zero, and $k, \ell = 0, 1, 2, \ldots, n - 1$.

See Hildebrand [1956], p. 390. Since $x_{j-1} = \cos\,[(j - \frac{1}{2})\pi/n]$, we have $T_{i-1}(x_{j-1}) = \cos\,[(i - 1)(j - \frac{1}{2})\pi/n]$.

FIGURE 2. STYLIZED $h_1$ AND $h_2$, AND THE FIRST FOUR PRINCIPAL COMPONENTS OF $Z_2$

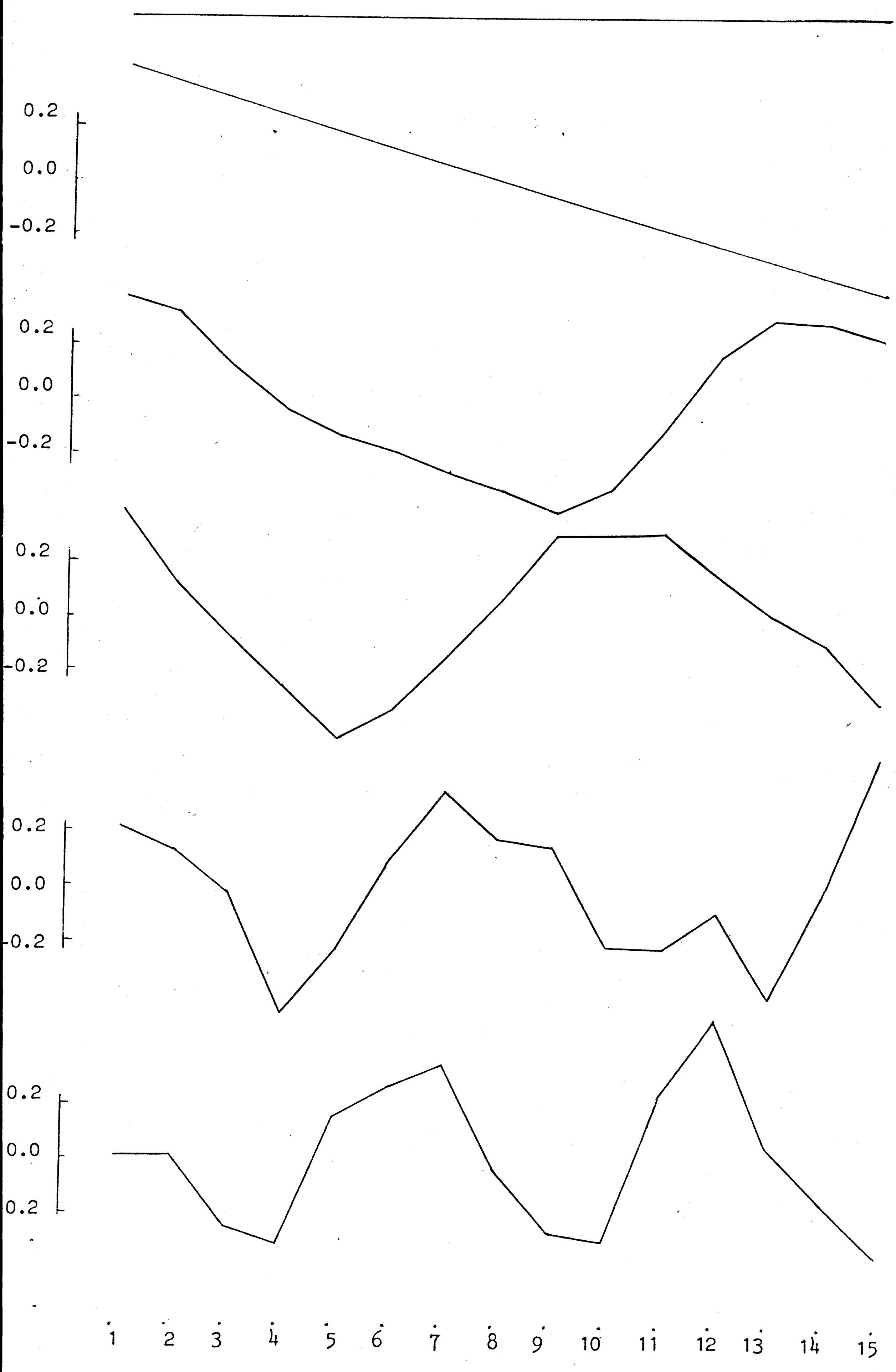
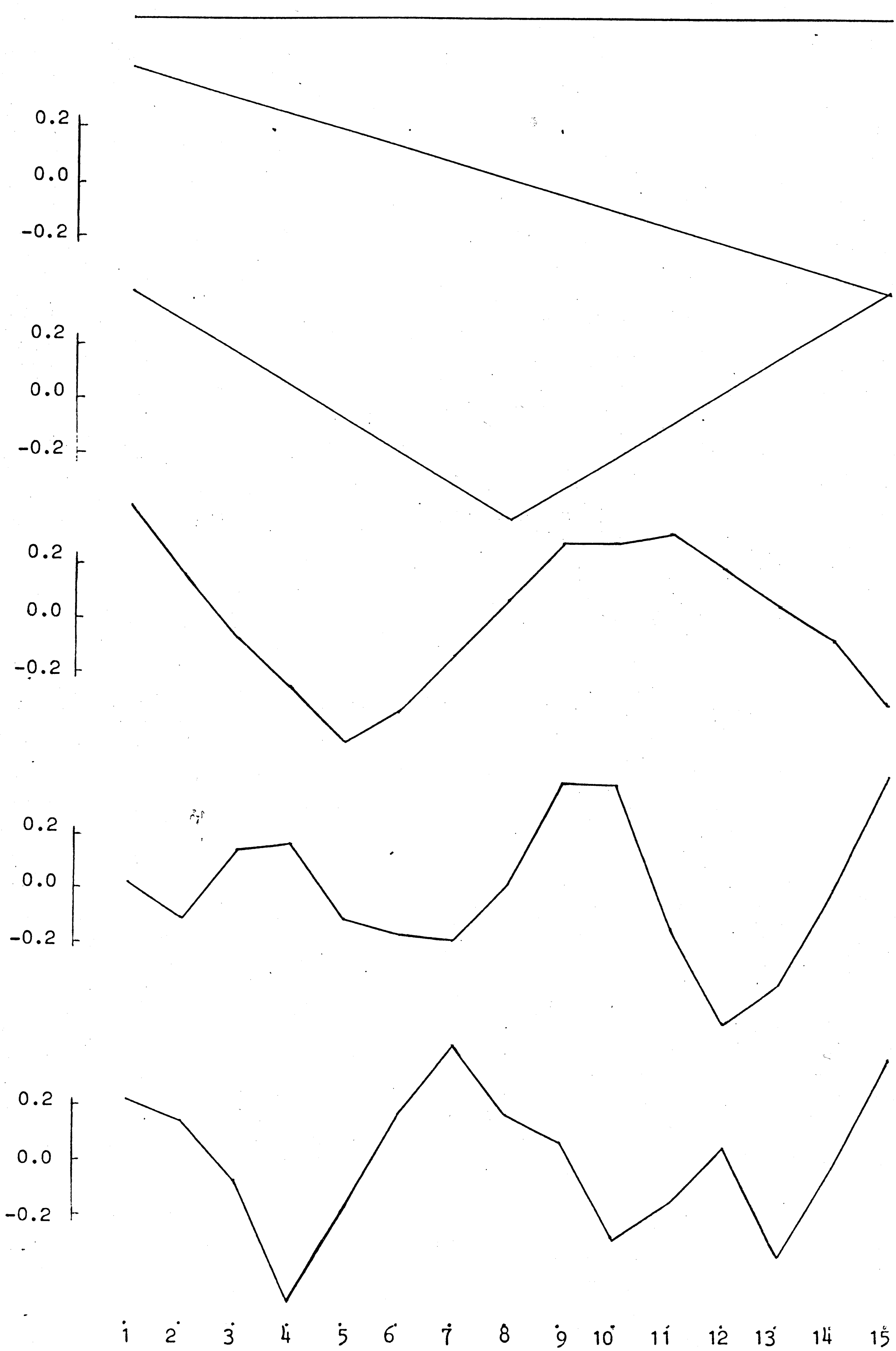


FIGURE 3. STYLIZED $h_1$, $h_2$, AND $h_3$, AND THE FIRST THREE PRINCIPAL COMPONENTS OF $Z_3$



We define the j-th element of $h_i^*$, which is the i-th column of $H^*$, as

(4.2) $$h_{ji}^* = c \cos [(i - 1)(j - \tfrac{1}{2})\pi/n] \qquad c = n^{-\frac{1}{2}} \text{ if } i = 1$$
$$= (\tfrac{n}{2})^{-\frac{1}{2}} \text{ if } i \neq 1$$

The n × n matrix $H^*$ is orthonormal, which follows from the orthogonality theorem (4.1). Figure 4 shows $h_1^*$ through $h_6^*$ for n = 15. They explain 33.3, 24.3, 18.7, 8.2, 3.8, and 4.2 per cent of Z, respectively. The first three vectors together explain 76.3 per cent of Z, even more than in the case of Figure 3.

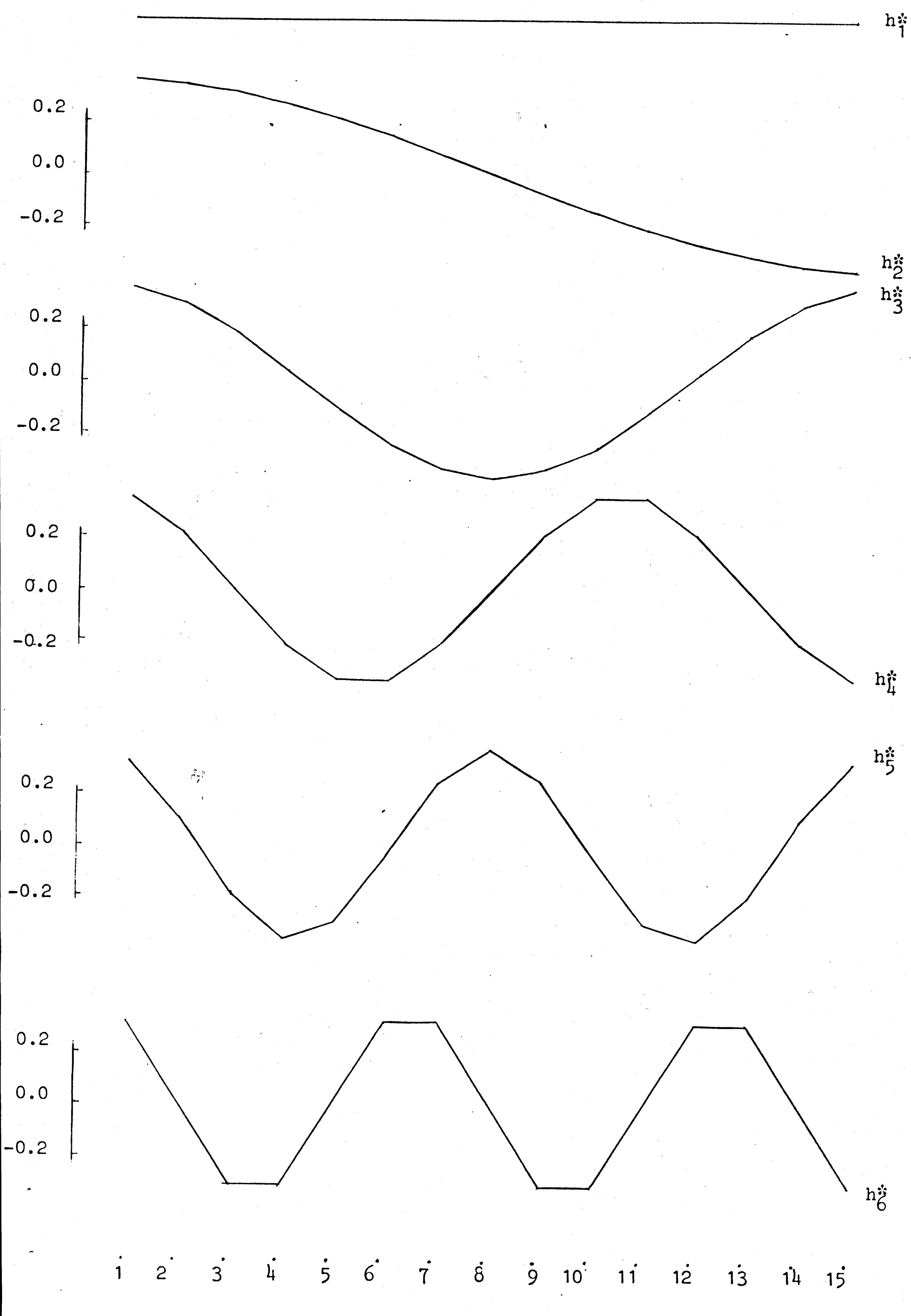
The vectors $h_i^*$ happen to be eigenvectors of A, the matrix occurring in the numerator of the Von Neumann ratio,

(4.3) $$A = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ 0 & -1 & 2 & \cdots & 0 & 0 \\ \vdots & & & & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}$$

The i-th eigenvector $h_i^*$ is defined by the eigenvalue $2[1 - \cos (i - 1)\pi/n]$ The vectors $h_1^*$, $h_2^*$, and $h_3^*$ were adopted and applied by Abrahamse and Louter [1971] on the basis of remarks made by Hannan [1960].

Availability of a generation formula for $h^*$-vectors is very advantageous. A simple computer subroutine is needed to generate $h_i^*$ for any i, $i \leq n$, and for any n. We adopt $h_i^*$ as the i-th principal component of $Z^*$.

FIGURE 4. THE FIRST SIX $h^*$-VECTORS FOR n = 15



## 5. GENERALIZATION FOR k AND n

Write Z (m; n, k) for the n × (mk) matrix consisting of m R-matrices, each of order n × k. In the preceding sections we investigated Z (30; 15, 2 + 1), where 2 + 1 indicates that the original X-matrices contain two time series columns and a constant term column. By rearrangement of our sixty data vectors we got Z (20; 15, 3) and Z (20; 15, 3 + 1), the only difference being the addition of a constant term to all X-matrices; Z (15; 15, 4) and Z (15; 15, 4 + 1) with the same difference; and Z (12; 15, 5). Write $\mu_i^*$ for the percentage contribution of $h_i^*$, defined in (4.2), to the explanation of Z (m; n, k); for $\mu_i$ see (3.1)

TABLE 2. PERCENTAGE CONTRIBUTIONS TO THE EXPLANATION OF Z (m; n, k) BY PRINCIPAL COMPONENTS OF Z (m; n, k) AND h*-VECTORS

| i | Z(30;15,2+1) | | Z(20;15,3+1) | | Z(15;15,4+1) | | Z(20;15,3) | | Z(15;15,4) | | Z(12;15,5) | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | $\mu_i$ | $\mu_i^*$ | $\mu_i$ | $\mu_i^*$ | $\mu_i$ | $\mu_i^*$ | $\mu_i$ | $\mu_i^*$ | $\mu_i$ | $\mu_i^*$ | $\mu_i$ | $\mu_i^*$ |
| 1 | 33.3 | 33.3 | 25.0 | 25.0 | 20.0 | 20.0 | 33.1 | 33.1 | 24.9 | 24.8 | 20.0 | 20.0 |
| 2 | 27.1 | 24.3 | 22.0 | 21.6 | 18.8 | 18.4 | 25.5 | 25.2 | 21.4 | 21.1 | 18.5 | 18.3 |
| 3 | 17.5 | 18.7 | 18.1 | 17.2 | 17.2 | 16.0 | 17.6 | 16.1 | 20.7 | 18.7 | 17.1 | 15.8 |
| 4 | 8.5 | 8.2 | 11.8 | 11.4 | 13.4 | 13.2 | 8.4 | 8.4 | 13.4 | 13.3 | 12.8 | 12.9 |
| 5 | 4.1 | 3.8 | 7.2 | 6.3 | 8.8 | 7.9 | 5.0 | 3.8 | 5.2 | 3.4 | 9.2 | 7.0 |
| 6 | 3.6 | 4.2 | 5.9 | 6.4 | 6.1 | 6.5 | 4.0 | 5.0 | 4.1 | 5.8 | 7.0 | 8.1 |
| $\sum_1^k$ | 77.9 | 76.4 | 76.9 | 75.2 | 78.2 | 75.6 | 76.3 | 74.3 | 80.5 | 77.9 | 77.5 | 74.0 |

It is seen that the subcolumns below $\mu_i$ and $\mu_i^*$ are pairwise very much alike. The conclusion is that P, consisting of the first k h*-vectors a good idealization of P, consisting of the first k principal components of Z (m; n, k) for n = 15, at least for k = 3, 4, 5. The idealized P roughly explains 75 per cent of Z in all cases. This means, see (3.2),

$$\bar{\phi} = (1 - 0.75)k = 0.25k$$

Before stating a general hypothesis, we investigate the effect of some variations of n.

Reconsider the X-matrices underlying Z (30; 15, 2 + 1). We examine four types of submatrices: delete (a) the first three, (b) the first five, (c) the last three, and (d) the last five rows of each X. For each type we computed a matrix Z (30; n, 2 + 1): $Z_a$, $Z_b$, $Z_c$, and $Z_d$, say, with n = 12, 10, 12, 10, respectively.

TABLE 3. PERCENTAGE CONTRIBUTIONS TO THE EXPLANATION OF $Z_j$ BY PRINCIPAL COMPONENTS OF $Z_j$ AND h*-VECTORS

| i | $Z_a$ | | $Z_b$ | | $Z_c$ | | $Z_d$ | |
|---|---|---|---|---|---|---|---|---|
| | $\mu_i$ | $\mu_i^*$ | $\mu_i$ | $\mu_i^*$ | $\mu_i$ | $\mu_i^*$ | $\mu_i$ | $\mu_i^*$ |
| 1 | 33.3 | 33.3 | 33.3 | 33.3 | 33.3 | 33.3 | 33.3 | 33.3 |
| 2 | 23.7 | 23.4 | 24.8 | 24.5 | 30.3 | 29.1 | 30.7 | 30.4 |
| 3 | 20.4 | 18.5 | 18.6 | 15.9 | 17.3 | 16.7 | 17.4 | 17.2 |
| 4 | 8.6 | 8.6 | 10.3 | 12.6 | 8.5 | 6.9 | 6.8 | 6.6 |
| 5 | 5.6 | 6.3 | 5.3 | 3.7 | 4.6 | 6.4 | 4.8 | 4.2 |
| 6 | 2.9 | 2.7 | 3.1 | 4.0 | 2.1 | 2.3 | 2.8 | 1.5 |
| $\sum_1^3$ | 77.4 | 75.3 | 76.7 | 73.7 | 80.6 | 79.1 | 81.4 | 80.9 |

As in Table 2 we find a high correlation between the subcolumns of Table 3, suggesting the following hypothesis.

In economic time series analysis the best idempotent Ω of order n × n and rank (n - k) is I - PP', where P is the n × k leading submatrix of the n × n matrix H*, defined by (4.2).

The adjective "best" means: on the average, Ω is as close as possible to M in a least-squares sense, while the only thing we know about X is, that it contains annual data.

## 6. TEST CASES; A SELECTION DEVICE

We computed powers of tests for both autocorrelation and heterovariance on each of the six 15 × 3 matrices $X_C$, $X_K$, $X_S$, $X_L$, $X_T$, and $X_A$, described below, with alternative specifications of $\Omega$. At the suggestion of Durbin [1970], we compared his alternative exact test for autocorrelation, which uses the disturbance estimator z,

$$z = [\Omega - \Omega\tilde{X}(\tilde{X}'\Omega\tilde{X})^{-1}\tilde{X}'\Omega + \Omega\tilde{X}P_1P_2'\tilde{P}'M]y$$

where $\Omega = I_{(n)} - PP'$, $P'P = I_{(k)}$, $P = [h_1^* \vdots \tilde{P}]$, $\tilde{X}$ is the $n \times (k-1)$ matrix obtained from X by deleting the constant term vector, and $P_1$ and $P_2$ are the lower triangular matrices such that $P_1P_1' = (\tilde{X}'\Omega\tilde{X})^{-1}$ $P_2P_2' = (\tilde{P}'M\tilde{P})^{-1}$. Both estimators, z and w, are linear in y and unbiased, having the same covariance matrix, $\sigma^2\Omega$, when taking the same P. Durbin takes $\tilde{P} = [h_2^* \vdots h_3^*]$ in his applications. He does not indicate in which sense z is best, if any. It is not clear why orthonormal transformations of $P_1$ and $P_2$ are disregarded.

All six X-matrices contain a constant term vector. The time series data are taken from literature, as follows.

- $X_C$. Chow [1957], Table 1, log automobile stock per capita and log personal money stock per capita for the United States, 1921-1935.
- $X_K$. Klein [1950], p.135, profits and wages for the United States, 1923-1937.
- $X_S$. Sato [1970], p. 203, capital and man hours for the United States, 1946-1960.
- $X_L$. Log transform of the Sato data.
- $X_T$. Theil and Nagar [1961], Table 4, log real income per head and log relative price for the Netherlands, 1923-1937.
- $X_A$. Koerts and Abrahamse [1969], p. 153-154, artificial data.

For all tests the maintained hypothesis is that u is normally distributed with zero mean and the null hypothesis is $E(uu') = \sigma_0^2 I$. The alternative hypothesis is $E(uu') = \sigma_1^2\Gamma_{(15)}$ with $\rho = 0.8$ when testing for autocorrelation, and it is $E(uu') = \sigma_1^2{}_{(15)}$ with $\gamma = 0.5$ when testing for heterovariance,

$$\Gamma_{(n)} = \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \vdots & & & & \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}, \qquad \Lambda_{(n)} = \begin{bmatrix} 1^\gamma & 0 & 0 & \dots & 0 \\ 0 & 2^\gamma & 0 & \dots & 0 \\ \vdots & & & & \cdot \\ 0 & 0 & 0 & \dots & n^\gamma \end{bmatrix}$$

We use the autocorrelation test statistic Q and the heterovariance test statistic T

$$Q = \frac{w'Aw}{w'w} \qquad\qquad T = \frac{w'Cw}{w'(I_{(15)} - C)w}$$

where A is the 15 × 15 matrix given in (4.3) and C is the 15 × 15 matrix with leading submatrix $I_{(7)}$, all other elements being zero. The statistic T has been proposed by Theil [1968]. Both test statistics are of the form (1.2), so using an estimator w = B'u with B'B = I - PP', their distributions under the null hypothesis depend on P alone.

Though we did not use a power maximization criterion, but, instead, an estimation criterion to arrive at $P = [h_1^* \vdots h_2^* \vdots h_3^*]$, testing procedures nevertheless should mainly be evaluated by means of their relative powers. Relative, since we are generally unaware of the maximum attainable. On the other hand, a test with relatively high powers may be disqualified by computational inconvenience.

To each of the six X-matrices some alternative specifications of P are taken as follows.

(I) P equals R. Then Ω = M, w reduces to the ordinary least-squares estimator and Q becomes the Durbin-Watson test statistic. Durbin and Watson [1971] prove that their test for autocorrelation is locally most powerful invariant.

(II) $P = [h_1^* \vdots h_2^* \vdots h_3^*]$. This is the specification according to the hypothesis at the end of the previous section.

(III) $P = [h_1^* \vdots h_2^* \vdots h_3^*]$. Same specification as II, but now Durbin's estimator z is used.

(IV) $P = [e_7 \vdots e_8 \vdots e_9]$, where $e_i$ is a 15 element vector with 1 at place i, all other elements zero. Here $\Omega$ becomes a diagonal matrix, with diagonal elements equal to unity apart from the three central elements, which are zero. Therefore w is a modified BLUS-estimator, modified in the sense that its three central elements are not deleted.

(V) $P = [h_1^* \vdots h_i^* \vdots h_j^*]$. The subscripts i and j depend on X and are determined by a selection device, defined below.

From Table 4 we learn the following:

(a) $\phi$ versus $\psi$. Ranking the rows of Table 7 according to increasing $\phi$ is identical to ranking these rows according to increasing $\psi$. This is what we meant by saying that the behaviour of $\phi$ and $\psi$ is very alike, see Section 2.

(b) Durbin's estimator (z) versus BLUS estimator ($e_7e_8e_9$). Apart from the autocorrelation test on $X_A$, where the difference is relatively small, the BLUS estimator scores higher powers in all cases.

(c) Durbin's estimator versus w ($h_1^*h_2^*h_3^*$). Apart from the autocorrelation test on the Sato data, where the difference is relatively small, the Durbin estimator scores lower powers in all cases.

(d) BLUS estimator versus w. Apart from the tests on the Sato data and the heterovariance test on $X_T$, where the difference is relatively small, the BLUS estimator scores lower powers in all cases.

(e) Least-squares estimator (R) versus both Durbin's estimator and BLUS estimator. L-s estimator scores higher powers in all cases, particularly in the autocorrelation tests.

(f) L-s estimator versus w. In the heterovariance tests there is no systematical winner; in the autocorrelation tests the l-s estimator scores higher powers in all cases, though the difference is small for $X_K$, $X_T$, and $X_A$.

(g) There is no simple relation between the approximation measure values and the powers computed.

The power levels of the tests are not essential for our purpose: the comparison of different P's. For instance, the powers for Q are determined by $\alpha$ (significance level), $\rho$ (specification of the alternative hypothesis), R (least-squares regression space), and P (a priori fixed regression space).

TABLE 4. APPROXIMATION MEASURES, AND POWERS FOR Q AND T AT SIGNIFICANCE LEVEL $\alpha$

| X | P | Approx. measures $\phi$ | $\psi$ | Powers for Q at $\alpha$=0.05 | $\alpha$=0.10 | Powers for T at $\alpha$=0.05 | $\alpha$=0.10 |
|---|---|---|---|---|---|---|---|
| $X_C$ | $R_C$ | 0.000 | 0.000 | 0.536 | 0.665 | 0.131 | 0.230 |
| | $h_1^* h_2^* h_3^*$ | 0.440 | 0.238 | 0.466 | 0.595 | 0.132 | 0.231 |
| | (z) | 0.440 | 0.238 | 0.372 | 0.495 | 0.115 | 0.206 |
| | $e_7 e_8 e_9$ | 2.492 | 2.071 | 0.424 | 0.564 | 0.127 | 0.224 |
| $X_K$ | $R_K$ | 0.000 | 0.000 | 0.629 | 0.734 | 0.142 | 0.245 |
| | $h_1^* h_2^* h_3^*$ | 1.186 | 1.027 | 0.605 | 0.715 | 0.142 | 0.245 |
| | (z) | 1.186 | 1.027 | 0.371 | 0.501 | 0.121 | 0.215 |
| | $e_7 e_8 e_9$ | 2.485 | 1.972 | 0.501 | 0.631 | 0.139 | 0.240 |
| | $h_1^* h_2^* h_4^*$ | 0.467 | 0.250 | 0.610 | 0.719 | 0.138 | 0.239 |
| $X_S$ | $R_S$ | 0.000 | 0.000 | 0.628 | 0.738 | 0.144 | 0.248 |
| | $h_1^* h_2^* h_3^*$ | 0.969 | 0.795 | 0.355 | 0.479 | 0.137 | 0.238 |
| | (z) | 0.969 | 0.795 | 0.367 | 0.490 | 0.134 | 0.233 |
| | $e_7 e_8 e_9$ | 2.730 | 2.248 | 0.436 | 0.586 | 0.141 | 0.243 |
| | $h_1^* h_2^* h_8^*$ | 0.663 | 0.416 | 0.625 | 0.734 | 0.135 | 0.235 |
| $X_L$ | $R_L$ | 0.000 | 0.000 | 0.636 | 0.744 | 0.143 | 0.246 |
| | $h_1^* h_2^* h_3^*$ | 0.990 | 0.850 | 0.357 | 0.482 | 0.137 | 0.238 |
| | (z) | 0.990 | 0.850 | 0.374 | 0.500 | 0.132 | 0.231 |
| | $e_7 e_8 e_9$ | 2.715 | 2.226 | 0.455 | 0.603 | 0.139 | 0.240 |
| | $h_1^* h_2^* h_8^*$ | 0.645 | 0.397 | 0.632 | 0.740 | 0.134 | 0.233 |
| $X_T$ | $R_T$ | 0.000 | 0.000 | 0.508 | 0.633 | 0.143 | 0.247 |
| | $h_1^* h_2^* h_3^*$ | 0.341 | 0.182 | 0.499 | 0.625 | 0.138 | 0.241 |
| | (z) | 0.341 | 0.182 | 0.335 | 0.464 | 0.103 | 0.186 |
| | $e_7 e_8 e_9$ | 2.349 | 1.962 | 0.360 | 0.510 | 0.141 | 0.243 |
| $X_A$ | $R_A$ | 0.000 | 0.000 | 0.628 | 0.726 | 0.141 | 0.245 |
| | $h_1^* h_2^* h_3^*$ | 0.699 | 0.440 | 0.626 | 0.725 | 0.143 | 0.246 |
| | (z) | 0.699 | 0.440 | 0.545 | 0.652 | 0.103 | 0.186 |
| | $e_7 e_8 e_9$ | 2.304 | 1.924 | 0.532 | 0.646 | 0.134 | 0.233 |
| | $h_1^* h_3^* h_4^*$ | 0.380 | 0.209 | 0.628 | 0.726 | 0.135 | 0.233 |

We carried out a lot of computations, and the results suggest that power ratios of the form

$$\frac{\text{Power for Q, given } \alpha, \rho, R, P_0}{\text{Power for Q, given } \alpha, \rho, R, P_1}$$

are approximately constant for varying α, ρ, and R.

From Table 4 it is evident that the spaces $M(R_S)$ and $M(R_L)$ differ too much from $M(h_1^*, h_2^*, h_3^*)$, in one respect or another, relevant to the test performed. The latter space is recommended as being best on the average, as long as X is unknown. Now we must allow for other P-matrices. The selection of a specification from a given set of P-matrices in a particular case then depends on X. Taking into account that monthly or quarterly economic time series data, for instance, usually show a different fluctuation pattern over time, we propose to admit P to consist of higher frequency $h^*$-vectors.

We adopt ϕ as a criterion for the selection of k $h^*$-vectors. For a given X-matrix of order n × k, we choose that P for which ϕ is minimal. This leads to the following selection device

> Compose P of the k $h^*$-vectors corresponding to the k smallest values of $h_i^{*\prime}Mh_i^*$, i = 1, 2, ..., n.

If X contains a constant term, the device prescribes inclusion of $h_1^*$, for $h_1^*Mh_1^* = 0$. Exclusion would mean P'X singular, in which case w does not exist. Application of this device selects $P = [h_1^*h_2^*h_3^*]$ for $X_C$ and $X_T$, $P = [h_1^*h_2^*h_4^*]$ for $X_K$, $P = [h_1^*h_2^*h_8^*]$ for $X_S$ and $X_L$, $P = [h_1^*h_3^*h_4^*]$ for $X_A$. Powers are also presented in Table 4 for these specifications. It appears that the heterovariance tests are not improved, while the powers for Q are very close to the Durbin-Watson power.

We computed ψ for all admissible P-matrices for each of the six X-matrices. The P for which ψ is minimal with respect to X coincides with the P indicated by the selection device, in all six cases. For all admissible P-matrices we computed powers for Q in the case of $X_C$ and $X_T$. In the case of $X_T$ the highest power is scored for $P = [h_1^*h_2^*h_3^*]$. In the case of $X_C$ the highest power scores are presented in Table 5, together with the powers for T.

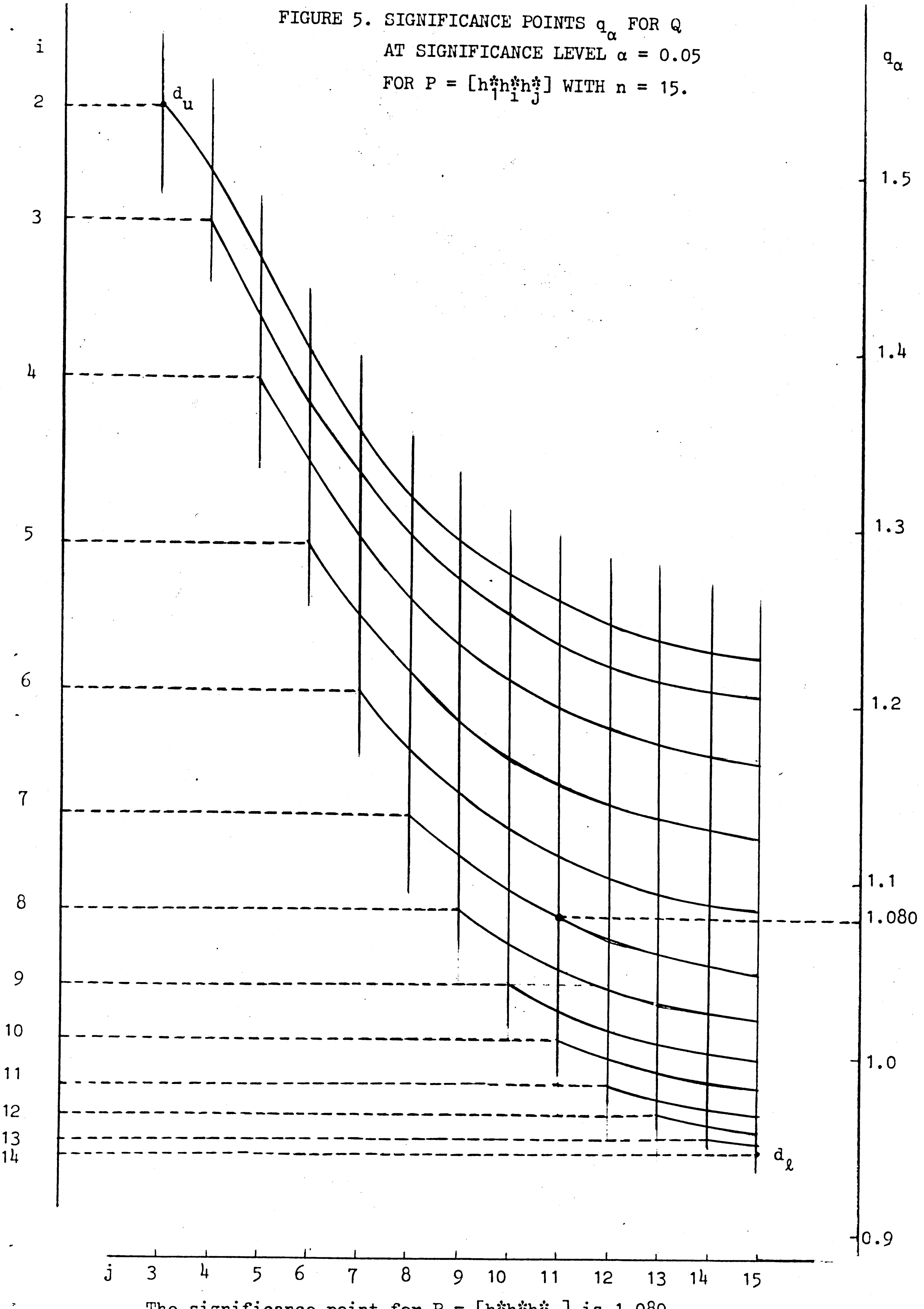
TABLE 5. POWERS FOR Q AND T AT $\alpha = 0.05$; $X = X_C$ AND $P = [h_1^* h_i^* h_j^*]$

| i j | $\phi$ | Power for Q | Power for T |
|---|---|---|---|
| 4 8 | 1.781 | 0.566 | 0.108 |
| 3 8 | 0.996 | 0.563 | 0.120 |
| 5 8 | 1.855 | 0.563 | 0.100 |
| 6 8 | 1.816 | 0.555 | 0.107 |
| (1-s) | 0.000 | 0.536 | 0.131 |
| 3 8 | 1.011 | 0.533 | 0.128 |
| 7 8 | 1.879 | 0.525 | 0.100 |
| 4 6 | 1.830 | 0.522 | 0.112 |
| 4 5 | 1.869 | 0.522 | 0.124 |
| 2 8 | 1.211 | 0.519 | 0.117 |
| 3 5 | 1.084 | 0.509 | 0.118 |
| 4 9 | 1.867 | 0.503 | 0.109 |
| 4 7 | 1.893 | 0.502 | 0.105 |

Table 5 shows that $\phi$ is not a power indicator, that the Durbin-Watson test is overpowered four times, and that there is no high-power-for-Q-high-power-for-T relation.

The conclusion of this paper is that adoption of $P = [h_1^* h_2^* h_3^*]$ yields comparatively good powers when testing for heterovariance with T, and that adoption of the selection device generally yields powers when testing for autocorrelation with Q which are competitive to the powers of the exact Durbin-Watson test, at least when computational convenience is taken into account. When no exact test for autocorrelation is needed, the $\beta$-approximation to the exact Durbin-Watson test gives good results: for our X-matrices the computer time needed is somewhat less compared with the computation of w, while the exact significance levels, corresponding to the $\beta$-approximated 5 per cent significance points, range from 0.0486 to 0.0507.

FIGURE 5. SIGNIFICANCE POINTS $q_\alpha$ FOR Q AT SIGNIFICANCE LEVEL $\alpha = 0.05$ FOR $P = [h^*_1 h^*_i h^*_j]$ WITH $n = 15$.




The significance point for $P = [h^*_1 h^*_7 h^*_{11}]$ is 1.080.

The selection device involves tabulation of a large number of significance points. The regular pattern of the 91 significance points at $\alpha = 0.05$, which are computed for Table 5, suggest that some formula must exist which makes tabulation unnecessary (see Figure 5).

## REFERENCES


Abrahamse, A.P.J., and J. Koerts (1971), "New Estimators of Disturbances in Regression Analysis". Journal of the American Statistical Association 66, pp. 71-4.

Abrahamse, A.P.J., and A.S. Louter (1971), "On a New Test for Auto-correlation in Least-Squares Regression". Biometrika 58, pp. 53-60.

Abramowitz, M., and I.A. Stegun (1965), Handbook of Mathematical Functions. Washington: National Bureau of Standards.

Anderson, T.W. (1958), An Introduction to Multivariate Statistical Analysis. New York: Wiley.

Chow, G.C. (1957), Demand for Automobiles in the United States. Amsterdam: North Holland Publishing Company.

Dubbelman, C., A.P.J. Abrahamse and J. Koerts (1970), "A New Class of Disturbance Estimators in the General Linear Model". Report 7008 of the Econometric Institute of the Netherlands School of Economics, Rotterdam.

Durbin, J. (1970), "An Alternative to the Bounds Test for Testing for Serial Correlation in Least-Squares Regression. Econometrica 38, pp. 422-429.

Durbin, J. and G.S. Watson (1971), "Testing for Serial Correlation in Least-Squares Regression III". Biometrika 58, pp. 1-19.

Hannan, E.J. (1960), Time Series Analysis. London: Methuen.

Hildebrand, F.B. (1956), Introduction to Numerical Analysis. New York: McGraw-Hill.

Klein, L.R. (1950), Economic Fluctuations in the United States 1921-1941. New York: Wiley.

Koerts, J., and A.P.J. Abrahamse (1969), On the Theory and Application of the General Linear Model. University of Rotterdam Press.

Sato, R. (1970), "The Estimation of Biased Technical Progress and the Production Function." International Economic Review 11, pp. 179-208.

Theil, H. (1968), "A Simplification of the BLUS Procedure for Analysing Regression Disturbances". Journal of the American Statistical Association 63, 242-251.

Theil, H., and A.L. Nagar (1961), "Testing the Independence of Regression Disturbances". Journal of the American Statistical Association 56, pp. 793-806.

# APPENDIX

## B' Expressed in Terms of X and P only

(a) Existence of $(K'MK)^{-\frac{1}{2}}$

The matrix $(K'MK)^{-\frac{1}{2}}$ exists (and has real elements) if and only if all eigenvalues of K'MK are strictly positive. Denote the eigenvalues of K'MK by $\lambda_i$, $i = 1, 2, \ldots, n - k$, so that the eigenvalues of K'RR'K are $(1 - \lambda_i)$, $i = 1, 2, \ldots, n - k$, and $0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_{n-k} \leq 1$. Denote the eigenvalues of R'PP'R by $d_i$, $i = 1, 2, \ldots, k$, so that the eigenvalues of R'KK'R are $(1 - d_i)$, $i = 1, 2, \ldots, k$, and $0 \leq d_1 \leq d_2 \leq \cdots \leq d_k \leq 1$. Since the nonzero eigenvalues of K'RR'K and the nonzero eigenvalues of R'KK'R are the same, we have $\lambda_i = d_i$ for $i = 1, 2, \ldots, k$ and $\lambda_i = 1$ for $i = k + 1, k + 2, \ldots, n - k$ in the case $k < n - k$; $\lambda_i = d_i$ for $i = 1, 2, \ldots, k$ in the case $k = n - k$; and $\lambda_i = d_i$ for $i = 1, 2, \ldots, n - k$ and $d_i = 1$ for $i = n - k + 1, n - k + 2, \ldots, k$ in the case $k > n - k$. Therefore, $(K'MK)^{-\frac{1}{2}}$ exists if and only if $d_1 > 0$, or, what amounts to the same thing, P'R is nonsingular. Besides, $\mathrm{tr}\,(K'MK)^{\frac{1}{2}} = \sum_{i=1}^{n-k} \lambda_i^{\frac{1}{2}} = n - 2k + \sum_{i=1}^{k} d_i^{\frac{1}{2}}$.

(b) $(K'MK)^{-\frac{1}{2}} = I + (K'MPLD^{-\frac{1}{2}})(D + D^{\frac{1}{2}})^{-1}(K'MPLD^{-\frac{1}{2}})'$

We write $P'RR'P = LDL'$, where $L' = L^{-1}$ and D is diagonal. Then

$$R'PP'R = (R'PLD^{-\frac{1}{2}})D(R'PLD^{-\frac{1}{2}})'$$

and

$$R'KK'R = (R'PLD^{-\frac{1}{2}})(I - D)(R'PLD^{-\frac{1}{2}})'$$

or

$$R'KK'R(R'PLD^{-\frac{1}{2}}) = (R'PLD^{-\frac{1}{2}})(I - D)$$

and

$$K'RR'K(K'MPLD^{-\frac{1}{2}}) = (K'MPLD^{-\frac{1}{2}})(I - D)$$

where we replaced RR' by M = I - RR', since K'P = 0. Evidently,

$$(K'MPLD^{-\frac{1}{2}})'(K'MPLD^{-\frac{1}{2}}) = I - D$$

so that $(K'MPLD^{-\frac{1}{2}})$ can be considered as consisting of unscaled eigenvectors of K'RR'K. Suppose the j-th diagonal element of D, $d_j$, is equal to unity, so that $(I - D)^{-1}$ does not exist. Then the j-th column of $(K'MPLD^{-\frac{1}{2}})$ is a zero column. Let us write $\Lambda$ for the diagonal matrix obtained from D by deleting row j and column j for every j for which $d_j = 1$, and let H denote the matrix obtained from $(K'MPLD^{-\frac{1}{2}})$ by deleting column j for all j for which $d_j = 1$. Now we have

$$H'H = I - \Lambda$$

and

$$[H(I - \Lambda)^{-\frac{1}{2}}]'[H(I - \Lambda)^{-\frac{1}{2}}] = I$$

Then

$$K'R[R'KK'R]R'K = (K'MPLD^{-\frac{1}{2}})(I - D)(K'MPLD^{-\frac{1}{2}})' = H(I - \Lambda)H'$$

$$= [H(I - \Lambda)^{-\frac{1}{2}}](I - \Lambda)^2[H(I - \Lambda)^{-\frac{1}{2}}]'$$

and hence

$$K'RR'K = [H(I - \Lambda)^{-\frac{1}{2}}](I - \Lambda)[H(I - \Lambda)^{-\frac{1}{2}}]'$$

Let G denote the orthonormal complement to $H(I - \Lambda)^{-\frac{1}{2}}$, then

$$K'RR'K = [H(I - \Lambda)^{-\frac{1}{2}} \vdots G] \begin{bmatrix} I - \Lambda & \vdots & 0 \\ \cdots\cdots & \vdots & \cdots \\ 0 & \vdots & 0 \end{bmatrix} [H(I - \Lambda)^{-\frac{1}{2}} \vdots G]'$$

and

$$K'MK = [H(I - \Lambda)^{-\frac{1}{2}} \vdots G] \begin{bmatrix} \Lambda & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & I \end{bmatrix} [H(I - \Lambda)^{-\frac{1}{2}} \vdots G]'$$

Hence,

$$(K'MK)^{-\frac{1}{2}} = H(I - \Lambda)^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}(I - \Lambda)^{-\frac{1}{2}}H' + GG'$$

$$= I + H(I - \Lambda)^{-\frac{1}{2}}(\Lambda^{-\frac{1}{2}} - I)(I - \Lambda)^{-\frac{1}{2}}H'$$

$$= I + H(\Lambda + \Lambda^{\frac{1}{2}})^{-1}H'$$

$$= I + (K'MPLD^{-\frac{1}{2}})(D + D^{\frac{1}{2}})^{-1}(K'MPLD^{-\frac{1}{2}})'$$

The effect of the last step is the addition of half of the outer products of zero vectors.

(c) $$B' = I - TD^{-1}T' - S(I + D^{\frac{1}{2}})^{-1}S' - TD^{-\frac{1}{2}}S'$$

Substitution of the expression for $(K'MK)^{-\frac{1}{2}}$ into the formula

$$B' = K(K'MK)^{-\frac{1}{2}}K'M$$

may give the formula

$$B' = I - TD^{-1}T' - S(I + D^{\frac{1}{2}})^{-1}S' - TD^{-\frac{1}{2}}S'$$

where $S = MPL$, $T = PL - S$, $D$ is diagonal, and $L$ is the $k \times k$ orthonormal matrix such that $P'MP = L(I - D)L'$. Use of this formula reduced the computer time needed to compute $B'$ for the case $n = 15$ and $k = 3$ by some 80 per cent. If $X$ has not full column rank, then $M = I - RR'$ where $R$ is an orthonormal matrix spanning $\mathcal{M}(X)$, otherwise use $M = I - X(X'X)^{-1}X'$.