The Annals of Statistics
1985, Vol. 13, No. 2, 526–533

BEST INVARIANT ESTIMATION OF A DIRECTION PARAMETER[1]

BY T. W. ANDERSON, CHARLES STEIN, AND ASAD ZAMAN

Stanford University, Stanford University, and Columbia University

Let X be an $n \times k$ random matrix whose coordinates are independently normally distributed with common variance σ^2 and means given by $EX = e\mu' + \theta\lambda'$, where e is the vector in R^n having all coordinates equal to 1, $\theta \in R^n$, and μ, $\lambda \in R^k$ with $\sum_{j=1}^k \lambda_j^2 = 1$. The problem is to estimate λ, say by $\hat{\lambda}$, with loss function $1 - (\lambda'\hat{\lambda})^2$ when μ, θ, and σ^2 are unknown. It is shown that the largest principal component of $X'X - (1/n)X'ee'X$ is the best estimator invariant under rotations in R^k and rotations in R^n leaving e invariant and is admissible.

1. Introduction. The classical problem of a linear functional relation concerns a collection of bivariate observations whose expectations lie on a line. If the "errors" are independent and have equal variances, the usual estimate of the line is that line minimizing the sum of squared deviations of the points from the line in the direction orthogonal to the fitted line (Adcock, 1878). Suppose the observations (Y_i, Z_i) for $i \in 1, \cdots, n$ are independently normally distributed with means (η_i, ζ_i) and covariance matrix $\sigma^2 I$ and the means satisfy the linear relation

$$\zeta_i = \alpha + \beta\eta_i \quad \text{for} \quad i \in \{1, \cdots, n\}. \tag{1}$$

Then the usual estimate of (α, β), described earlier, is the maximum likelihood estimate. In this paper we find some optimum properties of this, thought of as an estimate of the angle ν that the line makes with the first coordinate axis. Of course, the slope β is the tangent of this angle. (Anderson, 1976, discusses this estimator and some of its properties.)

The covariance matrix $\sigma^2 I$ is invariant under two-dimensional rotations and reflections, and this is also true of the angle between the estimated line and the true line. The random vectors $Y = (Y_1, \cdots, Y_n)'$ and $Z = (Z_1, \cdots, Z_n)'$ are independently distributed, having spherically symmetric normal distributions with means $\eta = (\eta_1, \cdots, \eta_n)'$ and $\zeta = (\zeta_1, \cdots, \zeta_n)'$ related by $\zeta = \alpha e + \beta\eta$, where $e = (1, \cdots, 1)'$. The problem is also invariant under n-dimensional rotations and reflections leaving the vector e invariant. We shall show that the maximum likelihood estimator of the angle ν is the best estimator invariant under the 2-dimensional and n-dimensional orthogonal transformations described above and deduce that it is admissible in the class of all estimators. The loss function

Received September 1982; revised May 1984.

[1] This work was supported by National Science Foundation Grants SES-79-13976, MCS80-24649, and MCS82-19748 at Stanford University. The paper was completed while the first author was at Columbia University and IBM Systems Research Institute.

AMS 1980 *subject classifications.* Primary 62C15, secondary 62F10.

Key words and phrases. Best invariant estimation, direction parameters, linear functional relationship, factor analysis.

is taken to be the square of the sine of the difference in angles. This loss function increases with the difference in angles in $[0, \pi/2]$ and decreases in $[\pi/2, \pi]$; it is reasonable because the line is not directed.

Our general treatment includes the case of n observations in a k-dimensional space with the means on a line and the "errors" independent with equal variances. The direction of the usual estimated line is the direction of the first principal component (Pearson, 1901; Hotelling, 1933). This model is the case of factor analysis where there is one factor and the error variances are equal; the direction cosines of the line are the factor loadings. See Lawley (1953) for the derivation of maximum likelihood estimates in this model.

Although the direction of the first principal component seems a natural estimator of the direction of a line of means, the optimal properties developed in the paper seem to be the first decision-theoretic results in the area of factor analysis, linear functional relationships, and principal component analysis. Nearly all known general properties of estimators are asymptotic. (See Anderson, 1984.) It is worth remarking that the admissibility of the best invariant estimators is not a priori evident in this problem. As Berger (1976) has shown for the location parameter problem, best invariant estimators may fail to be admissible in the presence of too many nuisance parameters.

A similar result was obtained much earlier by Blackwell (1951). His Example 2 on page 397 can be interpreted as a k-dimensional (with $k \geq 4$) location parameter problem located on the lattice of points with integral coordinates, where it is required to estimate a linear combination (with weights linearly independent over the rationals) of the coordinates of the location parameter vector. He showed that the best translation-invariant estimator is inadmissible.

2. The best invariant estimator of a direction. Let X be an $n \times k$ random matrix whose coordinates are independently normally distributed with common variance σ^2 and means given by

$$EX = e\mu' + \theta\lambda', \tag{2}$$

where e is the vector in R^n having all coordinates equal to 1, $\theta \in R^n$ and μ, $\lambda \in R^k$, with $\| \lambda \| = \sqrt{\Sigma \lambda_j^2} = 1$. In this section μ will be assumed known. Without essential loss of generality we take $\mu = 0$. With θ, λ, and σ unknown, we study the problem of estimating λ, say by $\hat{\lambda}$, with loss function L defined by

$$L((\sigma, \theta, \lambda), \hat{\lambda}) = 1 - (\lambda'\hat{\lambda})^2. \tag{3}$$

This loss is the square of the sine of the angle between the line of means and the estimated line. For the problem described in Section 1 the loss function is $(\hat{\beta} - \beta)^2(1 + \beta^2)^{-1}(1 + \hat{\beta}^2)^{-1}$. Since the loss remains unchanged when λ is replaced by $-\lambda$ the ambiguity in the identification of λ does not matter. Our aim is to prove that the usual estimator of λ, the (normalized) characteristic vector of $X'X$ corresponding to the largest characteristic root, has certain optimal properties.

The problem is invariant under orthogonal transformations on the left and on the right. More precisely, if X is distributed according to the parameters $(\sigma, \theta, \lambda)$ and p is an $n \times n$ orthogonal matrix and q a $k \times k$ orthogonal matrix, then

pXq' is distributed according to the parameters $(\sigma, p\theta, q\lambda)$. Also

$$L((\sigma, \theta, \lambda), \hat{\lambda}) = L((\sigma, p\theta, q\lambda), q\hat{\lambda}). \tag{4}$$

A (randomized) decision procedure for this problem is a measurable function δ on $R^{n\times k}$ to the set of all probability measures in S^{k-1}, the unit sphere in R^k. The interpretation is that, after observing X, the statistician takes a random action whose conditional distribution given X is $\delta(X)$. This decision procedure δ is said to be invariant under the group of transformations (p, q) introduced above if for all such p, q and all $n \times k$ matrices x and all measurable $B \subseteq S^{k-1}$

$$\delta(pxq')(qB) = \delta(x)(B), \tag{5}$$

where qB denotes the set of all $q\hat{\lambda}$ with $\hat{\lambda} \in B$. Let δ_0 be the (invariant) decision procedure defined by

$$\delta_0(x)\{\psi_0(x)\} = \delta_0(x)\{-\psi_0(x)\} = ½, \tag{6}$$

where $\psi_0(x)$ and $-\psi_0(x)$ are the two possible choices of the characteristic vector of $x'x$ corresponding to the largest characteristic root.

THEOREM 1. *The decision procedure δ_0 is the best invariant decision procedure for this problem in the sense that for any essentially different invariant procedure δ, and any σ, $\tau > 0$ the risk of δ is greater than that of δ_0 at $(\sigma, \theta, \lambda)$ for any θ, λ with $\| \theta \| = \sqrt{\theta'\theta} = \tau$. Note that the risk of any invariant procedure depends only on σ, τ.*

PROOF. From (5) with $q = I$ we see that for all orthogonal $n \times n$ matrices p, $\delta(px) = \delta(x)$. However, when $x'_{(1)}x_{(1)} = x'_{(2)}x_{(2)}$ (with $x_{(1)}$ and $x_{(2)}$ $n \times k$ matrices) there exists p such that $x_{(1)} = px_{(2)}$. Thus, we can write

$$\delta(x) = \delta^*(x'x), \tag{7}$$

and it follows from (5) that for all orthogonal $k \times k$ matrices q

$$\delta^*(qx'xq')(qB) = \delta^*(x'x)(B). \tag{8}$$

In particular if, for given x, we choose a diagonal matrix ℓ with diagonal elements in decreasing order and an orthogonal matrix q such that

$$x'x = q'\ell q, \tag{9}$$

we have, for all B,

$$\delta^*(x'x)(B) = \phi(\ell)(qB), \tag{10}$$

where we have written $\phi(\ell)$ for $\delta^*(\ell)$. Let $q^{(i)}$ be the $k \times k$ diagonal matrix, all of whose diagonal elements are 1 except for the ith which is -1. For diagonal matrices ℓ

$$q^{(i)\prime}\ell q^{(i)} = \ell. \tag{11}$$

Substitute $x'x = \ell$ and $q = q^{(i)}$ in (9) and (10) to obtain

$$\phi(\ell)(B) = \phi(\ell)(q^{(i)}(B)), \tag{12}$$

that is, $\phi(\ell)$ is invariant under reflections in the coordinate axes. Summarizing, the invariant procedures δ are those expressible in the form (7) with δ^* satisfying (10) (with q, ℓ, and x related by (9)), and $\phi(\ell)$ satisfying (12) for diagonal ℓ.

Our aim now is to prove that the best choice ϕ_0 of ϕ is given by

$$\phi_0(\ell)\{e^{(1)}\} = \phi_0(\ell)\{-e^{(1)}\} = \tfrac{1}{2}, \tag{13}$$

where $e^{(1)} = (1, 0, \cdots, 0)'$. It suffices to prove that δ_0^* is the essentially unique Bayes procedure against the invariant prior distribution

$$\Sigma(d\theta)\xi(d\lambda), \tag{14}$$

where Σ is the orthogonally invariant probability measure on the sphere $\| \theta \| = \tau$ in R^n and ξ the orthogonally invariant probability measure on the unit sphere in R^k. Without essential loss of generality we take $\sigma = 1$. The posterior density of the parameters with respect to the prior distribution (14) is given by

$$p(\theta, \tau, \lambda \mid X) = \exp[-\tfrac{1}{2}\mathrm{tr}\,(X - \theta\lambda')'(X - \theta\lambda')]\rho(X) = \exp(\theta' X\lambda)\rho^*(X), \tag{15}$$

where ρ and ρ^* represent functions whose precise form is of no relevance to the argument.

One minus the posterior risk of the randomized decision rule δ related to ϕ by (7)–(12) is proportional to

$$I(\phi) = \int\int\int (\lambda' q' z)^2 \exp(\theta' X\lambda)\Sigma(d\theta)\xi(d\lambda)\phi(\ell)(dz). \tag{16}$$

Let t be an orthogonal matrix with first column proportional to $X\lambda$. Changing variables from θ to $t\theta$ and expanding the exponential into its Taylor series gives

$$\begin{aligned} I(\phi) &= \int\int\int (\lambda' q' z)^2 \textstyle\sum_{j=0}^{\infty} \dfrac{\theta_1^j \, \| X\lambda \|^j}{j!} \, \xi(d\lambda)\Sigma(d\theta)\phi(\ell)(dz) \\ &= \textstyle\sum_{j=0}^{\infty} \left[\displaystyle\int \frac{\theta_1^{2j}}{(2j)!}\,\Sigma(d\theta)\right]\left[\int\int (\lambda' q' z)^2 \, \| X\lambda \|^{2j}\xi(d\lambda)\phi(\ell)(dz)\right]. \end{aligned} \tag{17}$$

The odd terms drop out since θ_1 has a symmetric distribution. Now note that since ϕ is reflection-symmetric, $\int z_i z_j \phi(\ell)(dz) = \int (-z_i)z_j\phi(\ell)(dz)$, and hence z_i and z_j are uncorrelated if $i \neq j$. Therefore

$$\begin{aligned} &\int\int (\lambda' q' z)^2(\lambda' X' X\lambda)^j \xi(d\lambda)\phi(\ell)(dz) \\ &\qquad = \int\int (v' z)^2 (v' q X' X q' v)^j \xi(dv)\phi(\ell)(dz) \\ &\qquad = \textstyle\sum_{m=1}^{k} \left[\displaystyle\int z_m^2 \phi(\ell)(dz)\right]\left[\int v_m^2 \left(\textstyle\sum_{h=1}^{k} \ell_h v_h^2\right)^j \xi(dv)\right], \end{aligned} \tag{18}$$

where $\lambda = q'v$. This is a weighted sum of the terms

$$A_m = \int v_m^2 \left(\textstyle\sum_{h=1}^{k} \ell_h v_h^2\right)^j \xi(dv). \tag{19}$$

We shall show that A_1 is the largest among the A_m for all values of j and hence we can maximize the weighted sum simultaneously for all j by choosing ϕ so that $\int z_1^2\phi(\ell)(dz) = 1$. Since ϕ_0 is the unique symmetric measure on the unit sphere for which this is true, this will prove the theorem.

LEMMA 1. *Let $V = (V_1, \cdots, V_k)$ be a random variable uniformly distributed on the surface of the unit sphere in R^k. Then*

$$E \prod_{i=1}^k V_i^{2\alpha_i} = \frac{\Gamma(k/2)}{\Gamma(1/2)^k} \frac{\Gamma(\alpha_1 + 1/2) \cdots \Gamma(\alpha_k + 1/2)}{\Gamma(k/2 + \sum_{i=1}^k \alpha_i)}. \tag{20}$$

PROOF. Let $Y = (Y_1, \cdots, Y_k)$ be distributed according to $N(0, I_k)$. Then V can be represented as $V_i = Y_i/\| Y \|$, $i = 1, \cdots, p$. Let $Z_i = Y_i^2/\| Y \|^2$, $i = 1, \cdots, p$. Then $Z_1, \cdots, Z_{k-1}$ have the Dirichlet density

$$\frac{\Gamma(p_1 + \cdots + p_k)}{\prod_{i=1}^k \Gamma(p_i)} z_1^{p_1-1} \cdots z_k^{p_k-1}, \quad z_k \equiv 1 - \sum_{i=1}^{k-1} z_i,$$

for $z_1 \geq 0, \cdots, z_k \geq 0$, with $p_1 = \cdots = p_k = 1/2$. This follows from the fact that $Y_1^2, \cdots, Y_k^2$ are independent χ^2-variables with one degree of freedom. (See Exercise 11, page 215, Rao (1973), for example; note that Rao's constant should be inverted.) Then we have

$$E \prod_{i=1}^k V_i^{2\alpha_i} = E \prod_{i=1}^k Z_i^{\alpha_i},$$

which is (20). □

Using the multinomial expansion we can write A_m as

$$\begin{aligned} A_m &= \sum_{j_1+\cdots+j_k=j} \frac{j!}{j_1! \cdots j_k!} \ell_1^{j_1} \cdots \ell_k^{j_k} \int v_m^2(v_1^{2j_1} \cdots v_k^{2j_k})\xi(dv) \\ &= \sum \frac{j!}{j_1! \cdots j_k!} \ell_1^{j_1} \cdots \ell_k^{j_k} \frac{\Gamma(k/2)}{\Gamma(1/2)^k} \frac{\Gamma(j_1 + 1/2) \cdots \Gamma(j_k + 1/2)\Gamma(j_m + 3/2)}{\Gamma(j + (k/2) + 1)\Gamma(j_m + 1/2)}. \end{aligned}$$

From this it follows that

$$A_1 - A_m = \sum (j_1 - j_m) \frac{j!}{j_1! \cdots j_k!} \ell_1^{j_1} \cdots \ell_k^{j_k} \frac{\Gamma(1/2 k)}{\Gamma^k(1/2)} \frac{\Gamma(j_1 + 1/2) \cdots \Gamma(j_k + 1/2)}{\Gamma(j + 1/2 + 1)}. \tag{21}$$

Since $\ell_1 > \ell_m$, each term with $j_1 > j_m$ is positive and larger in absolute value than the corresponding term with j_1 and j_m reversed. It follows that $A_1 - A_m > 0$, proving the theorem.

COROLLARY 1. *The rule δ_0 is admissible for every fixed τ and σ, and thus also when τ and σ are not fixed.*

3. The case of unknown μ.

In order to obtain a result analogous to Corollary 1 for the case where μ in (2) is unknown, we shall need to apply a simple decision-theoretic lemma. Before introducing this, let us look at an analogue of Theorem 1. With q a $k \times k$ orthogonal matrix, p an $n \times n$ orthogonal matrix such that $pe = e$ and $\xi \in R^k$, we consider transformations (p, q, ξ) operating on the sample space by

$$(p, q, \xi)x = pxq' + e\xi'. \tag{22}$$

If X is distributed according to the parameter point $(\sigma, \theta, \lambda, \mu)$, then the transformed random matrix $pxq' + e\xi'$ is distributed according to $(\sigma, p\theta, q\lambda, q\mu + \xi)$. The condition for a decision function δ to be invariant is now

$$\delta(pxq' + e\xi')(B) = \delta(x)(B) \tag{23}$$

rather than (5). Let δ_1 be the particular invariant decision procedure defined by

$$\delta_1(x)\{\psi_1(x)\} = \delta_1(x)\{-\psi_1(x)\} = 1/2, \tag{24}$$

where $\psi_1(x)$ is the characteristic vector of $x'x - (1/n)x'ee'x$ corresponding to the largest characteristic root.

THEOREM 2. *The decision procedure δ_1 is the best invariant decision procedure for this problem in the sense that for any essentially different invariant procedure δ and any $\sigma, \tau > 0$ the risk of δ is greater than that of δ_1 at $(\sigma, \theta, \lambda, \mu)$ for any θ, λ, μ with*

$$\| \theta - (1/n)ee'\theta \| \doteq \tau. \tag{25}$$

PROOF. The result is reduced to Theorem 1. Let

$$\alpha = \begin{pmatrix} e'/\sqrt{n} \\ \alpha_{(2)} \end{pmatrix} \tag{26}$$

be an $n \times n$ orthogonal matrix and let

$$Y = \alpha X. \tag{27}$$

Partition Y as

$$Y = \begin{pmatrix} Y_{(1)} \\ Y_{(2)} \end{pmatrix}, \tag{28}$$

where $Y_{(1)}$ is the first row of Y. Then $Y_{(2)}$ is distributed as X in Section 2 with n replaced by $n - 1$ and θ replaced by $\alpha_{(2)}\theta$. Expressed in terms of Y, the invariant decision procedures are exactly the invariant procedures depending only on $Y_{(2)}$. Since δ_1 is related to $Y_{(2)}$ in the same way that δ_0 was related to X in Section 2, the desired result follows.

Because the group of transformations considered here is not compact, the admissibility of δ_1 does not follow as easily as the admissibility of δ_0 in Corollary 1. Instead, we shall need a decision-theoretic lemma which may be related to Lemma 3.1 of Kiefer and Schwartz (1965). Although the lemma we shall need has a simple proof, the statement is long and not easy to grasp. Thus a few preliminary remarks may be helpful. The sample space is given explicitly as a Cartesian product $\mathscr{V} \times \mathscr{W}$. In effect, condition (i) expresses the parameter space as a Cartesian product, but this is not expressed explicitly because it would appear unnatural in the applications. Conditions (ii) and (iii) indicate the relation between the two Cartesian product structures. A restriction is imposed on the loss function in (29).

LEMMA 2. *Let V and W be random variables (not necessarily real-valued, taking values in sets $\mathscr{V}$ and $\mathscr{W}$), jointly distributed according to P_ζ, where $\zeta \in \mathscr{Z}^*$ is an unknown parameter point. Suppose there are sets $\mathscr{V}^*$, $\mathscr{W}^*$, and functions α: $\mathscr{Z}^* \to \mathscr{V}^*$ and β: $\mathscr{Z}^* \to \mathscr{W}^*$ such that*

(i) *for every $(\alpha_1, \beta_1) \in \mathscr{V}^* \times \mathscr{W}^*$ there exists a unique $\zeta \in Z^*$ such that $\alpha(\zeta) = \alpha_1$ and $\beta(\zeta) = \beta_1$,*

(ii) *the distribution of V is determined by $\alpha(\zeta)$, and*

(iii) *the conditional distribution of W given V is determined by $\beta(\zeta)$.*

Suppose also that, with an action space $\hat{\mathscr{V}}$, the loss function L has the form

$$L(\zeta, a) = L^*(\alpha(\zeta), a) \tag{29}$$

for $\zeta \in \mathscr{Z}^*$. If δ_1 is a decision procedure depending on the sample point (V, W) only through V, admissible among such procedures, then δ_1, is admissible among all decision procedures.

PROOF BY CONTRADICTION. Suppose a decision procedure δ is everywhere as good as δ_1 and strictly better at ζ_0. We define a random variable W^*, whose conditional distribution given V (at every parameter point) is the same as the conditional distribution of W given V at the parameter point ζ_0. Let δ^* be the decision procedure defined by

$$\delta^*(V) = \delta(V, W^*). \tag{30}$$

We shall see that δ^* is strictly better than δ_1.

Identifying ζ with $(\alpha(\zeta), \beta(\zeta))$, which is legitimate because of condition (i), we have

$$\begin{aligned} &E_{\alpha(\zeta),\beta(\zeta)} L^*(\alpha(\zeta), \delta^*(V)) \\ &\qquad = E_{\alpha(\zeta),\beta(\zeta_0)} L^*(\alpha(\zeta), \delta(V, W^*)) \\ &\qquad = E_{\alpha(\zeta),\beta(\zeta_0)} L^*(\alpha(\zeta), \delta(V, W)) \le E_{\alpha(\zeta),\beta(\zeta_0)} L^*(\alpha(\zeta), \delta_1(V)) \\ &\qquad = E_{\alpha(\zeta),\beta(\zeta)} L^*(\alpha(\zeta), \delta_1(V)). \end{aligned} \tag{31}$$

The first equality uses the definition (30) of δ^* and the fact that the distribution of V is determined by $\alpha(\zeta)$. The second equality uses the fact that at $(\alpha(\zeta), \beta(\zeta_0))$, (V, W) has the same distribution as (V, W^*). The final equality again uses the fact that the distribution of V is determined by $\alpha(\zeta)$. Since the inequality in (31) was assumed at the beginning of the proof to hold strictly at ζ_0, this shows that δ^* is strictly better than δ_1, completing the proof.

Now we can prove the analogue of Corollary 1 for the case of unknown μ.

COROLLARY 2. *The decision procedure δ_1 defined in Theorem 2 is admissible for fixed σ and τ and thus also when σ and τ are not fixed.*

PROOF. By Corollary 1 and the construction used in (27) and (28) for the proof of Theorem 2, δ_1 is admissible among procedures depending only on $Y_{(2)}$. Because α is orthogonal, $\alpha_{(2)}e = 0$. If we identify the V, W, α, and β of Lemma 1 with $Y_{(2)}$, $Y_{(1)}$, $EY_{(2)} = \alpha_{(2)}$, $EX = \alpha_{(2)}\theta\lambda'$, and

$$EY_{(1)} = \sqrt{n}\mu' + (1/\sqrt{n})(e'\theta)\lambda'. \tag{32}$$

respectively, the conclusion follows.

Acknowledgement. The authors are indebted to a referee for pointing out serious errors in Lemmas 1 and 2.

REFERENCES

ADCOCK, R. J. (1878). A problem in least squares. *The Analyst* **5** 53–54.

ANDERSON, T. W. (1976). Estimation of linear functional relationships: approximate distributions and connections with simultaneous equations in econometrics. *J. Roy. Statist. Soc. Ser. B* **38** 1–36.

ANDERSON, T. W. (1984). Estimating linear statistical relationships. *Ann. Statist.* **12** 1–45.

BERGER, J. O. (1976). Inadmissibility results for generalized Bayes estimators of coordinates of a location vector. *Ann. Statist.* **4** 302–333.

BLACKWELL, D. (1951). On the translation parameter problem for discrete variables. *Ann. Math. Statist.* **22** 393–399.

HOTELLING, H. (1933). Analysis of a complex of statistical variables into principal components. *J. Educational Psych.* **24** 417–441, 498–520.

KIEFER, J. and SCHWARTZ, R. (1965). Admissible Bayes character of T^2, R^2, and other fully invariant tests for classical multivariate normal problems. *Ann. Math. Statist.* **36** 747–770.

LAWLEY, D. N. (1953). A modified method of estimation in factor analysis and some large sample results. *Uppsala Symp. Psych. Factor Anal. 17–19 March 1953.* Almqvist and Wiksell, Uppsala.

PEARSON, K. (1901). On lines and planes of closest fit to systems of points in space. *Philosophical Magazine* **2** (sixth series), 559–572.

RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, second edition. Wiley, New York.

T. W. ANDERSON
C. STEIN
DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305

A. ZAMAN
DEPARTMENT OF ECONOMICS
COLUMBIA UNIVERSITY
NEW YORK, NEW YORK 10027