PREDICTION AND DECISION PROBLEMS IN REGRESSION MODELS FROM THE BAYESIAN POINT OF VIEW*

ARNOLD ZELLNER AND V. KARUPPAN CHETTY
University of Chicago

In this paper we review the derivation of the predictive density function for the normal multiple regression model, state and prove a general theorem on optimal point prediction, and show how the predictive density can be employed in the analysis of an illustrative investment problem. Then we derive the predictive density function for the multivariate normal regression model and indicate how it can be used in the analysis of several problems.

1. INTRODUCTION

IN THIS paper we utilize a Bayesian approach to analyze several prediction and decision problems associated with normal regression models. In particular, for the multiple regression model, we review the derivation of the distribution of the next q observations and show how this distribution can be employed to make inferences about future values of the dependent variable. Then we state a rather general theorem which yields optimal points predictions for quadratic loss functions. Some of these results are utilized in an illustrative analysis of an investment decision. Here we obtain the distribution of the present value of a future income stream and show how it depends on the rate of discount and the length of the time horizon. Next, we generalize some of the results obtained for the multiple regression model to the multivariate normal regression model. We derive the distribution of a future vector observation. The properties of this distribution as well as its use in prediction and decision problems are discussed. We briefly indicate how this distribution can be employed in connection with economic policy, portfolio and farm management problems.

2. REVIEW OF PREDICTION IN THE MULTIPLE REGRESSION MODEL

2.1. *Specification of Model and Distribution of Next q Observations*

We assume that the $T\times 1$ observation vector y is generated by the usual normal multiple regression model

$$y = X\beta + u \tag{2.1}$$

where X is a $T\times K$ matrix with rank K of nonstochastic elements, β is a $K\times 1$ coefficient vector, and u is a $T\times 1$ error vector whose elements have zero means, common variance σ^2, and are normally and independently distributed. We shall further assume that our prior information is suitably represented by locally uniform and independent distributions for the elements of β and $\log \sigma$; cf. Jeffreys (1948), Savage (1961), and Box and Tiao (1962). That is, our prior distributions are:[1]

* This research was supported by the National Science Foundation under Grant GS-151. The authors wish to thank two unknown referees, Clifford Hildreth and John W. Pratt for suggestions which helped to improve the presentation. This paper appeared earlier as Systems Formulation and Methodology Workshop Paper 6403, University of Wisconsin, April, 1964.

[1] For use of non-diffused prior distributions in regression analysis see Raiffa and Schlaifer (1961) and Tiao and Zellner (1964a).

$$p(\beta) \propto \text{const.} \quad \text{and} \quad p(\sigma) \propto 1/\sigma \tag{2.2}$$

with $-\infty<\beta_\alpha<\infty$, $\alpha=1, 2, \cdots, K$, and $0<\sigma<\infty$. On combining the prior distributions in (2.2) with the likelihood function for the regression model by means of Bayes' Theorem, we obtain the following well-known posterior distribution of β and σ:

$$p(\beta, \sigma \mid y,) \propto \sigma^{-(T+1)} \exp \left\{-\frac{1}{2\sigma^2}(y - X\beta)'(y - X\beta)\right\}. \tag{2.3}$$

Now let the next q observations, $\tilde{y}' = (y_{T+1}, y_{T+2}, \cdots, y_{T+q})$ satisfy

$$\tilde{y} = \tilde{X}\beta + \tilde{u} \tag{2.4}$$

where $\tilde{X}$ is a $q\times K$ matrix of known nonstochastic elements, β is the $K\times 1$ coefficient vector appearing in (2.1), and $\tilde{u}$ is a $q\times 1$ vector of future errors, normally and independently distributed each with zero mean and common variance σ^2. We have then for the joint distribution of $\tilde{y}$, β, and σ:

$$p(\tilde{y}, \beta, \sigma \mid y) \propto \sigma^{-(T+q+1)} \exp \left\{-\frac{1}{2\sigma^2}(u'u + \tilde{u}'\tilde{u})\right\}.$$

Then on integrating with respect to σ and the elements of β, we obtain the following predictive probability density function for the q elements of $\tilde{y}$:

$$p(\tilde{y}\mid y) = c\{\nu + (\tilde{y} - \tilde{X}\hat{\beta})'H(\tilde{y} - \tilde{X}\hat{\beta})\}^{-(\nu+q/2)} \tag{2.5}$$

where

$$\nu = T-K,\ \hat{\beta} = (X'X)^{-1}X'y,\ H = (1/s^2)[I+\tilde{X}(X'X)^{-1}\tilde{X}']^{-1},\ \nu s^2 = (y-X\hat{\beta})'(y-X\hat{\beta})$$

and

$$c = \nu^{\nu/2}\left(\frac{\nu}{2}+\frac{q}{2}-1\right)!\,|H|^{1/2}/\pi^{q/2}\left(\frac{\nu}{2}-1\right)!\,.$$

The result in (2.5) is completely analogous to that derived in Raiffa and Schlaifer (1961, p. 345), which they call the unconditional distribution of $\tilde{y}$, except that here we work with the diffuse prior distributions in (2.2) rather than a non-diffuse natural conjugate prior distribution.[2] From (2.5), it is seen that the distribution of $\tilde{y}$ is in the multivariate Student t form—see, e.g., Dunnett and Sobel (1954), Cornish (1954), and Raiffa and Schlaifer (1961, p. 256 ff.). This implies that

$$E\tilde{y} = \tilde{X}\hat{\beta}, \tag{2.6}$$

$$E(\tilde{y} - E\tilde{y})(\tilde{y} - E\tilde{y})' = \frac{\hat{u}'\hat{u}}{T-K-2}[I + \tilde{X}(X'X)^{-1}\tilde{X}'], \tag{2.7}$$

[2] See Theil (1963) and Tiao and Zellner (1964a) for considerations bearing on the use of non-diffuse natural conjugate prior distributions in regression analysis.

and the quantity

$$\frac{\tilde{y}(i) - \tilde{X}(i)\hat{\beta}}{\sqrt{h^{ii}}} \tag{2.8}$$

has a univariate Student t distribution with $T-K$ degrees of freedom where $\tilde{y}(i)$ is the i^{th} element of $\tilde{y}$, $\tilde{X}(i)$ is the i^{th} row of $\tilde{X}$, and h^{ii} is the $(i, i)^{\text{th}}$ element of H^{-1}. Further, from the properties of the multivariate t distribution, linear combinations of the elements of $\tilde{y}$ will be distributed in the multivariate t form —see Raiffa and Schaifer (1961, p. 258). In particular, with $d' = (d_1, d_2, \cdots, d_q)$, a nonstochastic vector, $V = d'\tilde{y}$ has a univariate t distribution with $T-K$ degrees of freedom; that is,

$$t_{T-K} = \frac{V - \hat{V}}{s\{d'[I + \tilde{X}(X'X)^{-1}\tilde{X}']^{-1}d\}^{1/2}} \tag{2.9}$$

where $\hat{V} = E(V) = d'\tilde{X}\hat{\beta}$. This result will be utilized in an illustrative analysis of an investment problem in Section 2.3.

2.2. *Theorem on Optimal Point Predictions*

Let B be any positive definite $q \times q$ matrix; if $\bar{y}$ is a point prediction of $\tilde{y}$, let $L = (\tilde{y} - \bar{y})'B(\tilde{y} - \bar{y})$ be our loss function. Then the point prediction that minimizes EL is $\bar{y} = E\tilde{y} = \tilde{X}\hat{\beta}$.

Proof:

$$EL = E[(\tilde{y} - E\tilde{y}) - (\bar{y} - E\tilde{y})'B(\tilde{y} - E\tilde{y}) - (\bar{y} - E\tilde{y})$$

$$= E(\tilde{y} - E\tilde{y})'B(\tilde{y} - E\tilde{y}) + (\bar{y} - E\tilde{y})'B(\bar{y} - E\tilde{y}).$$

Given that B is p.d., this last expression will be a minimum for $\bar{y} = E\tilde{y} = \tilde{X}\hat{\beta}$, q.e.d.

Thus, when it is appropriate to employ a quadratic loss function, our theorem tells us that the mean of the distribution of $\tilde{y}$ will be an optimal point predictor, "optimal" in the sense of minimizing our expected loss.[3]

2.3. *Illustrative Application to an Investment Decision*

We assume that the regression model in (2.1) "explains" net returns on an investment. At time T, with sample information covering periods $t = 1, 2, \cdots, T$, an investor wishes to make an inference about the present value of net returns for the next q future periods, $T+1, T+2, \cdots, T+q$. That is, he is concerned about

---

[3] It is apparent that the minimization procedure could also be carried through with inequality restrictions imposed on the elements of $\bar{y}$, in which case we would have a quadratic programming problem to solve.

$$V = \frac{\tilde{y}_{T+1}}{1+r} + \frac{\tilde{y}_{T+2}}{(1+r)^2} + \cdots + \frac{\tilde{y}_{T+q}}{(1+r)^q} \tag{2.10}$$

where $\tilde{y}_{T+i}$ is net return in the i^{th} future period, t is the investor's rate of discount,[4] and V, a random variable, is the present value of the income stream for the next q periods.

Clearly V is a linear combination of the future y's and thus, as pointed out in Section 2.2, will have a univariate Student t distribution. That is, the quantity

$$t = (V - \hat{V})/s\{d'[I_q + \tilde{X}(X'X)^{-1}\tilde{X}']d\}^{1/2} \tag{2.11}$$

has a univariate Student t distribution with $T-K$ degrees of freedom where

$$d' = [(1+r)^{-1}, (1+r)^{-2}, \cdots, (1+r)^{-q}], \qquad \hat{y} = \tilde{X}\hat{\beta},$$
$$\hat{V} = d'\hat{y}, \quad \text{and} \quad s^2 = (T-K)^{-1}(y - X\hat{\beta})'(y - X\hat{\beta}).$$

If an investor has a utility function depending on V, say $U(V)$, its expectation can generally be evaluated given that we have the distribution of V. Also, if several investments are under consideration, the expected utility associated witn each can be computed and used to rank their relative attractiveness.

3. PREDICTION IN THE MULTIVARIATE REGRESSION MODEL

3.1. *Specification of Model and Distribution of Next Vector Observation*

In this Section we consider the traditional multivariate normal regression model

$$y_\alpha = X\beta_\alpha + u_\alpha \qquad \alpha = 1, 2, \cdots, m \tag{3.1}$$

where

y_α is a $T\times 1$ vector of observations on the α^{th} "dependent" variable,
X is a $T\times K$ matrix of nonstochastic elements with rank K,
β_α is a $K\times 1$ regression coefficient vector in the α^{th} equation, and
u_α is a $T\times 1$ vector of error terms.

We assume that the elements of the u_α vectors have zero means and a multivariate normal distribution with covariance matrix $\Sigma\otimes I_T$ where Σ is an $m\times m$ positive definite matrix with elements $\sigma_\alpha{}'$. Further, it is assumed that the next vector observation[5] $y'_F = (y_{1F}, y_{2F}, \cdots, y_{mF})$ has elements generated by

$$y_{\alpha F} = X_F\beta_\alpha + u_{\alpha F} \qquad \alpha = 1, 2, \cdots, m \tag{3.2}$$

where X_F is a $1\times K$ matrix of known elements relating to period $T+1$, and the $u_{\alpha F}$ are future error terms, with zero means, distributed normally and independently of the u_α with covariance matrix Σ.

[4] It would be possible to use rates of discount specific to particular future periods if they were thought appropriate.

[5] Note that in this section the symbol y_F denotes a $m\times 1$ vector of observations on m variates for the period $T+1$.

In the present analysis, we follow the invariance theory of Jeffreys (1961, p. 179ff.) and use for our prior distribution of $\beta=(\beta_1, \beta_2, \cdots, \beta_m)$ and Σ

$$p(\beta, \Sigma) = p(\beta)p(\Sigma) \tag{3.3}$$

with

$$p(\beta) \propto \text{constant} \tag{3.4}$$

and

$$p(\Sigma) \propto |\Sigma|^{-(m+1)/2} \tag{3.5}$$

As pointed out in Tiao and Zellner (1964b), (3.5) implies

$$p(\Sigma^{-1}) \propto |\Sigma^{-1}|^{-(m+1)/2}, \tag{3.6}$$

a prior distribution derived and employed by Savage (1961) in the analysis of a set of means; see also Geisser and Cornfield (1963).

On combining these prior distributions with the likelihood function for the model in (3.1), namely,

$$l(\beta, \Sigma^{-1} | y) \propto |\Sigma^{-1}|^{T/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}A\}, \tag{3.7}$$

where $y=(y_1, y_2, \cdots, y_m)$ and

$$A = \{u_\alpha' u_l\}, \qquad \alpha, l = 1, 2, \cdots, m,$$

we obtain the following joint posterior distribution for β and Σ^{-1}:

$$p(\beta, \Sigma^{-1} | y) \propto |\Sigma^{-1}|^{T-(m+1)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}A\}. \tag{3.8}$$

Further, we have that

$$\begin{aligned} p(y_F, \beta, \Sigma^{-1} | y) &= p(y_F | \beta, \Sigma^{-1}, y)p(\beta, \Sigma^{-1} | y) \\ &\propto |\Sigma^{-1}|^{(T-m)/2} \exp\{-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1}(A + A_F)\} \end{aligned} \tag{3.9}$$

where the matrix A_F is given by,

$$A_F = \{u_{\alpha F} u_{lF}\}, \qquad \alpha, l = 1, 2, \cdots, m. \tag{3.10}$$

For integration over the distinct elements of Σ^{-1}, we utilize the properties of the Wishart distribution, see, e.g., Anderson (1958, p. 154), to obtain

$$\begin{aligned} p(y_F, \beta | y) &\propto |A + A_F|^{-(T+1)/2} \\ &\propto |S + B|^{-(T+1)/2} \end{aligned} \tag{3.11}$$

where $S=\{\bar{u}_\alpha \bar{u}_l\}$ and $B=\{(\beta_\alpha-\hat{\beta}_\alpha-\tilde{\beta}_\alpha)'Z'Z(\beta_l-\hat{\beta}_l-\tilde{\beta}_l)\}$ with $\bar{u}_\alpha=w_\alpha-Z\tilde{\beta}_\alpha$, $w_\alpha' = [(y_\alpha - X\hat{\beta}_\alpha) \vdots y_{\alpha F} - X_F\hat{\beta}_\alpha]$, $Z' = (X' \vdots X_F)$, $\tilde{\beta}_\alpha = (Z'Z)^{-1}Z'w_\alpha$ and $\hat{\beta}_\alpha=(X'X)^{-1}X'y_\alpha$. Then following the analysis in Tiao and Zellner (1964b), we find

$$\begin{aligned} p(y_F, \beta | y) \propto\ & |\bar{S}|^{-1/2} |\bar{S} + \bar{B}|^{-T/2} |Z'D_mZ|^{-1/2} \\ & \cdot \{c_m + (\beta_m - \eta_m)'Z'D_mZ(\beta_m - \eta_m)\}^{-(T+1)/2} \end{aligned} \tag{3.12}$$

where $\overline{S}+\overline{B}$ is the $(m-1)\times(m-1)$ matrix formed from $S+B$ by deleting the m^{th} column and row, c_m is the reciprocal of the $(m, m)^{\text{th}}$ element of S^{-1},

$$D_m = I - \gamma(\overline{S} + \overline{B})^{-1}\gamma', \quad \gamma = (\gamma_1, \gamma_2, \cdots, \gamma_{m-1}), \quad \gamma_\alpha = Z\ (\beta_\alpha - \hat{\beta}_\alpha - \tilde{\beta}_\alpha),$$

$$\eta_m = \hat{\beta}_m + \tilde{\beta}_m + d\overline{S}^{-1}s, \quad d = (\beta_1 - \hat{\beta}_1 - \tilde{\beta}_1 \cdots \beta_{m-1}\ \hat{\beta}_{m-1} - \tilde{\beta}_{m-1})$$

and s is the mth column of S excluding the last element. Noting that (3.12) is in the multivariate t form, the integration over β_m can be performed. Then, by similar operations, the other elements of β can be integrated out to yield

$$p(y_F \mid y) \alpha \qquad \left| \{\hat{u}'_\alpha\, \hat{u}_e\} + q\{\hat{u}_{\alpha F}\, \hat{u}_{eF}\} \right|^{-(T-1-K)/2} \tag{3.13}$$

where $\hat{u}_{\alpha F} = y_{\alpha F} - X_F\hat{\beta}_\alpha$, $\hat{u}_\alpha = y_\alpha - X\hat{\beta}_\alpha$, and $q = [1 + X_F(X'X)^{-1}X'_F]^{-1}$. By some further algebra, the density function in (3.13) can be expressed as follows:[6]

$$\begin{aligned} p(y_F \mid y) \propto \{\nu + (y_{1F} - X_F\hat{\beta}_1 \cdots y_{mF} - X_F\hat{\beta}_m) \\ H(y_{1F} - X_F\hat{\beta}_1 \cdots y_{mF} - X_F\hat{\beta}_m)\}^{-(\nu+m)/2} \end{aligned} \tag{3.14}$$

where $\nu = T - K - (m-1)$ and $H = \nu q\{\hat{u}'_\alpha \hat{u}_l\}$. Then it is seen that the m elements of y_F are distributed in the multivariate t form which implies:

$$Ey_{\alpha F} = X_F\hat{\beta}_\alpha \qquad \alpha = 1, 2, \cdots, m \tag{3.15}$$

and

$$\begin{aligned} E(y_{\alpha F} - X_F\hat{\beta}_\alpha)(y_{lF} - X_F\hat{\beta}_l) &= \frac{\nu}{\nu - 2} H^{-1} \\ &= \frac{1}{\nu - 2}\{\hat{u}'_\alpha\, \hat{u}_e\}[1 + X_F(X'X)^{-1}X_F\ ]. \end{aligned} \tag{3.16}$$

Also, as is well-known, the marginal distribution of a subset of the elements of y_F will be in the multivariate t form and that of a single element will be in the univariate t form—cf., e.g., Raiffa and Schlaiffer (1961, p. 258). In addition, as mentioned above, linear combinations of variables distributed in the multivariate t form will themselves be so distributed. In particular, for nonstochastic n_α, $\alpha = 1, 2, \cdots, m$, the quantity W given by

$$W = n_1y_{1F} + n_2y_{2F} + \cdots + n_my_{mF} \tag{3.17}$$

is distributed in the univariate t form; that is, the quantity

$$t = \frac{W - \hat{W}}{[n'H^{-1}n]^{1/2}} \tag{3.18}$$

has a univariate t distribution with $\nu = T - K - (m-1)$ degrees of freedom where

$$\hat{W} = EW = \sum_{\alpha=1}^{m} n_\alpha X_F\hat{\beta}_\alpha \quad \text{and} \quad n' = (n_1n_2, \cdots, n_m).$$

[6] After this result was obtained and reported in the University of Wisconsin Systems Formulation and Methodology Workshop Paper 6403, April 20, 1964, alternative methods for deriving it have been suggested by a referee and friends. Also cf. Geisser (1964) for another derivation reported in a manuscript dated August, 1964.

4. DISCUSSION OF RESULTS AND APPLICATIONS

The distribution of y_F in (3.14) with mean vector and covariance matrix as given in (3.15) and (3.16), respectively, can be employed to make inferences about the "next" vector observation for multivariate normal regression models such as certain "unrestricted reduced form" systems or "reduced form" systems associated with "just identified" econometric models.[7] For these systems, this distribution will also play an important role in analyzing decision problems. That is, if $\bar{y}_F$ represents a set of known nonstochastic target values, as defined by Tinbergen (1956), for y_F, we clearly can determine the distribution of $y_F-\bar{y}_F$ from $p(y_F|y)$ in (3.19). Further, if we have a loss function which depends on $y_F-\bar{y}_F$, it seems possible to determine the X_F vector which minimizes the mathematical expectation of a loss function—cf. Theil (1961a), Fisher (1962), and the references cited in these works for some previous work on this problem. To illustrate how this problem can be approached utilizing our results, we consider a quadratic loss function $L=(y_F-\bar{y}_F)'G(y_F-\bar{y}_F)$, where $G=\{g_{\alpha l}\}$ is positive definite with known elements. Then,

$$\begin{aligned}
EL &= \int (y_F - \bar{y}_F)'G(y_F - \bar{y}_F)p(y_F \mid y)dy_F \\
&= \int (y_F - Ey_F)'G(y_F - Ey_F)p(y_F \mid y)dy_F + (\bar{y}_F - Ey_F)'G(\bar{y}_F - Ey_F) \\
&= \int \left[\sum_{\alpha,l} g_{\alpha l}(y_{\alpha F} - Ey_{\alpha F})(y_{lF} - Ey_{lF}) \right] p(y_F | y)dy_F \qquad (3.19) \\
&\quad + \sum_{\alpha,l} g_{\alpha l}(\bar{y}_{\alpha F} - Ey_{\alpha F})(\bar{y}_{lF} - Ey_{lF}) \\
&= \sum_{\alpha,l} g_{\alpha l}\{ \bar{s}_{\alpha l}[1 + X_F(X'X)^{-1}X_F'] + (\bar{y}_{\alpha F} - X_F\hat{\beta}_\alpha)(\bar{y}_{lF} - X_F\hat{\beta}_l)\}
\end{aligned}$$

where (3.15) and (3.16) have been used and $\bar{s}_{\alpha l}\equiv\hat{u}_\alpha\hat{u}_l/(\nu-2)$. Then, on differentiating EL with respect to X_F we obtain

$$\frac{\partial EL}{\partial X_F} = \sum_{\alpha,l} g_{\alpha l}\{2\bar{s}_{\alpha l}(X'X)^{-1} + 2\hat{\beta}_\alpha\hat{\beta}_l'\}X_F' - \sum_{\alpha,l} g_{\alpha l}(\hat{\beta}_\alpha\bar{y}_{lF} + \hat{\beta}_l\bar{y}_{\alpha F}).$$

The value of X_F which sets this derivative equal to zero, say $\overline{X}_F$, is given by

$$\overline{X}_F = \frac{1}{2}\left[\sum_{\alpha,l} g_{\alpha l}\{ \bar{s}_{\alpha l}(X'X)^{-1} + \hat{\beta}_\alpha\hat{\beta}_l' \right]^{-1} \sum_{\alpha,l} g_{\alpha l}(\hat{\beta}_\alpha\bar{y}_{lF} + \hat{\beta}_l\bar{y}_{\alpha F}) \qquad (3.20)$$

which can be computed easily given that the $g_{\alpha l}$ are known.

Another result which seems useful in connection with applications is that given in equation (3.18) which gives the distribution of the linear combination of next period's y's shown in (3.17). If $y_{\alpha F}$ in (3.17) is interpreted as the net yield per dollars invested in the αth investment alternative and n_α as the number of dollars invested in the αth alternative, then W is the return of an investor's

[7] For results pertaining to the prediction problem for "fully recursive" models, see Zellner (1964).

portfolio in the future period $T+1$. Since the distribution of W is known for given $n_1, n_2, \cdots, n_m$, it is not difficult to compute the distribution of W for various sets of n's and thus give an investor a chance to observe the distributions of the future return of his portfolio as its composition is varied. Further, if the investor has fixed resources, say W_T, then the n's are constrained to satisfy

$$W_T = \sum_{\alpha=1}^{m} n_\alpha.$$

If one postulates a utility function for the investor, $U(W)$, it appears possible to find the n's which maximize $EU(W)$ subject to this constraint.

$$W_T = \sum_{\alpha=1}^{m} n_\alpha.$$

Another possible application of our results is suggested if we regard $y_{\alpha F}$ as the yield per acre of the α^{th} crop in period $T+1$ and write

$$R = n_1 p_1 y_{1F} + n_2 p_2 y_{2F} + \cdots + n_m p_m y_{mF} \tag{3.21}$$

where p_α is the net unit price for the α^{th} crop, assumed known, and n_α is the number of acres devoted to the α^{th} crop. Then R is the net return to a farmer and is a random variable since it depends on the random, as yet unobserved, $y_{\alpha F}$, $\alpha=1, 2, \cdots, m$. Given a choice of the n's, R is distributed in the univariate t form and thus the distribution of R can be studied as a function of the choice of the n's. If the farmer has a fixed number of acres, say N, then of course we must have

$$N = \sum_{\alpha=1}^{m} n_\alpha.$$

Further if one posits a utility function for the farmer, say $\mathrm{U}(\mathrm{R})$, it seems possible to find the n's which maximize $\mathrm{E}\,\mathrm{U}(\mathrm{R})$ subject to the aforementioned constraint[8] and $n_\alpha \geq 0$, $\alpha=1, 2, \cdots, m$.

REFERENCES

Anderson, T. W. *An Introduction to Multivariate Statistical Analysis*, New York: J. Wiley and Sons, Inc., 1958.

Box, G. E. P. and Tiao, G. C. "A Further Look at Robustness via Bayes' Theorem," *Biometrika*, 49, 1962, 419–32.

Cornish, E. A. "The Multivariate t-Distribution Associated with a Set of Normal Sample Deviates," *Australian Journal of Physics*, 7, 1954, 531–542.

Dunnett, C. W. and Sobel, M. "A Bivariate Generalization of Student's t-Distribution with Tables for Special Cases," *Biometrika*, 41, 1954, 153–69.

Fisher, W. D. "Estimation in the Linear Decision Model," *International Economic Review*, 3, 1962, 1–29.

Geisser, S. and Cornfield, J. "Posterior Distribution for Multivariate Normal Parameters," *Journal of the Royal Statistical Society*, Series B, 25, 1963, 368–76.

[8] This problem has been formerly considered by Tintner (1955), Theil (1961b), and others from other points of view.

Geisser, S. "Bayesian Estimation in Multivariate Analysis," U. S. Army Mathematics Research Center, University of Wisconsin, Report No. 503, August, 1964.

Jeffreys, H. *Theory of Probability*, (2nd ed.), Oxford: Clarendon Press 1948.

——— *Theory of Probability*, (3rd ed.), Oxford: Clarendon Press 1961.

Raiffa, H. and Schlaifer, R. *Applied Statistical Decision Theory*, Harvard University, 1961.

Savage, L. J. *The Subjective Basis of Statistical Practice*, Manuscript, University of Michigan, 1961.

Theil, H. *Economic Forecasts and Policy*, (2nd ed.), Amsterdam: North-Holland Publishing Co., 1961.

——— "Some Reflections on Static Programming Under Uncertainty," *Weltwirtschafliches Archiv*, 87, 1961, 124–38.

——— "On the Use of Incomplete Prior Information in Regression Analysis," *Journal of the American Statistical Association*, 58, 1963, 401–14.

Tiao, G. C. and Zellner, A. "Bayes's Theorem and the Use of Prior Knowledge in Regression Analysis," *Biometrika*, 44, 1964a, 315–27.

Tiao, G. C. and Zellner, A. "On the Bayesian Estimation of Multivariate Regression," *Journal of the Royal Statistical Society*, Series B, 26, 1964b, 277–85.

Tinbergen, J. *Economic Policy: Principles and Design*, Amsterdam: North-Holland Publishing Co., 1956.

Tintner, G. "Stochastic Linear Programming with Application to Agricultural Economics," in *Proceedings of the Second Symposium on Linear Programming*, Washington, D.C., 1, 1955, 197–228.

Zellner, A. "On Prediction in 'Fully Recursive' Econometric Models," Systems Formulation and Methodology Workshop Paper 6402, University of Wisconsin 1964).