COWLES FOUNDATION DISCUSSION PAPER NO. 77

The Standard Error of Forecast for Interdependent Reduced Form Equations

John W. Hooper, Yale University

and

Arnold Zellner, University of Washington

July 29, 1959

The Standard Error of Forecast for Interdependent Reduced Form Equations

John W. Hooper, Yale University

and

Arnold Zellner, University of Washington

I. Introduction

In this paper we present a generalized standard error of forecast for a set of interdependent endogenous variables appearing in a multi-equation econometric model.* Since in most practical situations econometricians are usually concerned with forecasting the future values of a set of endogenous variables, it is important to be able to construct confidence regions for and test hypotheses about a joint forecast. In Section III, we derive an expression for the generalized standard error of forecast and show how it can be used to construct forecast regions and test hypotheses; an application of these methods to a two equation model is then made in Section IV. Finally, in Section V some further uses and limitations of the generalized standard error of forecast are discussed.

* The generalized standard error of forecast which we develop is a straightforward extension of the standard error of forecast employed in connection with single equation regression models. For review and reference purposes, the analysis leading to such a standard error of forecast is presented in Section II.

II. The Single Equation Error of Forecast.*

* An excellent presentation of this analysis is presented in [5, 280-284]; note, however, that Stone permits X to be either non-stochastic or stochastic in his treatment of least equares. Some complications arising in the case in which X is regarded as stochastic are briefly noted in Section V below.

The single equation regression model is:

$$(2.1) \qquad y_t = \sum_{i=1}^{K} \beta_i X_{it} + v_t \qquad (t=1, 2, \ldots, N)$$

or in matrix notation,

$$(2.2) \qquad y = \beta X + v$$

where y is a row vector of N observations on the dependent variable, X a $K\mathrm{x}N$ matrix of nonstochastic values (or fixed) values taken by K independent variables, β a row vector of unknown parameters, the regression coefficients, and v a random disturbance vector consisting of N independently distributed disturbances,** each with zero mean and a common variance $\sigma_v^2 < \infty$. Upon

** For constructing confidence regions and testing, the additional assumption of normality is needed.

estimating β by the method of least squares, we have

$$(2.3) \qquad y = bX + \bar{v}$$

where $b = yX'(XX')^{-1}$ and $\bar{v}$ is the vector whose components are the N calculated residuals.

Given a predicted value of y, say y_F, obtained using b and a set of values for the exogenous or independent variables, X_F, an expression for the standard error of y_F and a means of making probability statements concerning the difference between y_F and the true value of y_F, say y_F^*, is desired. We have that

$$y_F = bX_F \tag{2.4}$$

and

$$y_F^* = \beta X_F + v_F , \tag{2.5}$$

where X_F is a column vector of the values of the independent variables chosen as a basis for forecasting and v_F is the scalar value of the disturbance in the forecast period. The error of forecast is then,

$$y_F - y_F^* = (b-\beta)X_F - v_F . \tag{2.6}$$

The mean error is

$$E(y_F - y_F^*) = E[(b-\beta)X_F - v_F] = 0 \tag{2.7}$$

since b is an unbiased estimator of β, $Ev_F = 0$, and X_F is assumed to be nonstochastic.*

---

* If X_F is stochastic, as would be the case if a forecast of the exogenous variables for the prediction period were employed, it would be necessary to assume that X_F and $b-\beta$ are distributed independently (or, more weakly, are uncorrelated) for $E(y_F-y_F^*)$ to be equal to zero.

It is to be noted that the error of forecast, shown in (2.6), is a function of two random variables, b and v_F which are independently distributed.

On the further assumption that the disturbance term is normally distributed, it is the case that the forecast error is also normally distributed since it is a linear combination of normally distributed variables. The variance of the forecast error is given by

$$\sigma_F^2 = E\left\{(y_F - y_F^*) \quad (y_F - y_F^*)'\right\} \tag{2.8}$$

or,

$$\sigma_F^2 = \sigma_v^2 \; [1 + X_F' \; (XX')^{-1} \; X_F] \; . \tag{2.9}$$

The square root of (2.9) is the true standard error of forecast. As an estimate of σ_F^2 we use the unbiased estimator

$$\hat{\sigma}_F^2 = \hat{\sigma}_v^2 \; [1 + X_F' \; (XX')^{-1} \; X_F] \tag{2.10}$$

where

$$\hat{\sigma}_v^2 = \frac{yy' - b(XX')b'}{N-K} \; . \tag{2.11}$$

We can now consider the statistic

$$t = \left(\frac{y_F - y_F^*}{\sigma_F}\right) \Big/ (\hat{\sigma}_F / \sigma_F) \tag{2.12}$$

which is the familiar t statistic since, under the assumption that the disturbance term is normally distributed, $y_F - y_F^*$ is $N(0, \sigma_F^2)$, $(N-K)\ \hat{\sigma}_F^2 / \sigma_F^2$ is distributed as χ^2 with $N-K$ degrees of freedom, and $\hat{\sigma}_F$ and $y_F - y_F^*$ are independently distributed.* We can, making use of (2.12), define the forecast

---

* cf. [1, 206]

---

interval as,*

---

* See Section 5 for discussion of the interpretation of this forecast interval.

(2.13) $$\Pr\left\{y_F - \hat{\sigma}_F t_\alpha < y_F^* < y_F + \hat{\sigma}_F t_\alpha\right\} = 1-\alpha$$

where t_α is the value of (2.12) at the α level of significance. This probability statement may also be employed to test hypotheses about y_F^* .

III. Generalized Standard Error of Forecast

We consider a model completely analogous to what is referred to as the reduced form equations in econometric work, namely,

(3.1) $$y_{it} = \sum_{j=1}^{K} \Pi_{ij} X_{jt} + v_{it} \qquad (i=1, 2, \ldots, G;\ t=1, 2, \ldots, N)$$

or in matrix notation

(3.2) $$Y = \Pi X + V$$

where Y is a $G\mathrm{x}N$ matrix of endogenous variables, Π a $G\mathrm{x}K$ matrix of reduced form coefficients, X a $K\mathrm{x}N$ matrix of nonstochastic values taken by the exogenous variables, and V a $G\mathrm{x}N$ matrix of reduced form disturbances. It is further assumed about V that the N column vectors of V are independent random drawings from a G-dimensional population such that each column has zero mean and a common variance-covariance matrix, Σ_{vv} . These conditions may be summarized as

(3.3) $$E\, v_{it} = 0 \qquad i=1, 2, \ldots, G\ ;\ t=1, 2, \ldots, N$$

and

$$E\, v_{it}\, v_{jt'} = \begin{cases} 0 & t \neq t' \\ \sigma_{ik} & t = t' \end{cases} \quad \text{for all } i \text{ and } j\,. \tag{3.4}$$

Thus, we allow interdependence among the disturbances in different equations in the same time period but not among disturbances of different time periods.

By applying least squares to (3.2), we obtain

$$Y = PX + \bar{V} \tag{3.5}$$

where $P = YX'(XX')^{-1}$ is the matrix of estimated coefficients and $\bar{V}$ is the matrix of calculated residuals.

If we now desire to predict the values of the G endogenous variables using P and a set of values for the exogenous variables, X_F, the forecast would be:

$$Y_F = PX_F \tag{3.6}$$

where Y_F is a column vector of forecast values and X_F is a column vector of nonstochastic known values for the exogenous variables. The true value of Y_F, say Y_F^*, is given by

$$Y_F^* = \Pi X_F + V_F \tag{3.7}$$

where V_F is the column vector of values assumed by the disturbances in the forecast period. The error of forecast is

$$Y_F - Y_F^* = (P - \Pi)\, X_F - V_F\,. \tag{3.8}$$

The mean error is

$$E\,(Y_F - Y_F^*) = E[(P-\Pi)X_F - V_F] = 0 \tag{3.9}$$

since P is an unbiased estimator of Π, $EV_F = 0$, and X_F is assumed to be nonstochastic. The variance of the forecast error is given by

$$
\begin{aligned}
\Sigma_{FF} &= E\left\{(Y_F - Y_F^*)\ (Y_F - Y_F^*)'\right\} \\
&= E\left\{(P-\Pi)\ X_F X_F'\ (P-\Pi)'\right\} - E(P-\Pi)\ X_F V_F' - EV_F (P-\Pi)' X_F' + EV_F V_F'
\end{aligned}
\tag{3.10}
$$

The two middle terms in (3.10) are equal to zero since V_F and P are independently distributed.

To evaluate the first term in (3.10), we let $(P-\Pi) = A$ and $X_F X_F' = M_F$. Then we have

$$
A = \begin{bmatrix} a_{11} & \cdots & a_{1K} \\ \vdots & & \vdots \\ a_{G1} & \cdots & a_{GK} \end{bmatrix} = \begin{bmatrix} a_{1k} \\ a_{2k} \\ \vdots \\ a_{Gk} \end{bmatrix} \qquad (k=1,\ 2,\ \ldots,\ K)
\tag{3.11}
$$

and

$$
M_F = \begin{bmatrix} m_{11} & \cdots & m_{1K} \\ \vdots & & \vdots \\ m_{K1} & \cdots & m_{KK} \end{bmatrix} = (m_{k1},\ m_{k2},\ \ldots, m_{kK}) \qquad (k=1,\ 2,\ \ldots,\ K)
$$

so that $[a_{gk}]$ is the g'th row of A and $[m_{gk}]$ is the g'th column of M_F. We find that

$$
E[(P-\Pi)X_F X_F'(P-\Pi)'] = E\ AM_F A' = \begin{bmatrix} \sum_{k'=1}^{K} \sum_{k=1}^{K} \sigma_{k'k}^{(1)(1)} m_{kk'} & \cdots & \sum_{k'=1}^{K} \sum_{k=1}^{K} \sigma_{k'k}^{(1)(G)} m_{kk'} \\ \vdots & \ddots & \vdots \\ \sum_{k'=1}^{K} \sum_{k=1}^{K} \sigma_{k'k}^{(G)(1)} m_{kk'} & \cdots & \sum_{k'=1}^{K} \sum_{k=1}^{K} \sigma_{k'k}^{(G)(G)} m_{kk'} \end{bmatrix}
\tag{3.12}
$$

where $\sigma_{k'k}^{(i)(j)}$ is simply the k',k element in the covariance matrix of the i'th and j'th rows of P. The covariance between P_i and P_j, two rows of P, is

$$E(P_i-\Pi_i)(P_j-\Pi_j)' = (XX')^{-1} E\left\{\sum_{t=1}^{N}(y_{it}-Ey_{it})X_{it} \sum_{t'=1}^{N}(y_{jt'}-Ey_{jt'})X_{jt}\right\}(XX')^{-1}$$

$$= (XX')^{-1} \sum_{t,t'}^{N} E(y_{it}-Ey_{it})(y_{jt'}-Ey_{jt'})XX'(XX')^{-1}$$

(3.13)

$$=(XX')^{-1}\left[\sum_{t=t'}^{N} E(y_{it}-Ey_{it})(y_{jt}-Ey_{jt}) + \sum_{t\neq t'}^{N} E(y_{it}-Ey_{it})(y_{jt'}-Ey_{jt'})\right]$$

$$= \sigma_{ij}(XX')^{-1}$$

since the last term is zero in view of (3.4). This means, in summary, that the row vector of GK components, $(P_1, P_2, \ldots, P_G)$ is normally distributed (provided the disturbances are normally distributed)* with mean $(\Pi_1, \Pi_2, \ldots, \Pi_G)$,

* cf. [1 , 182] .

and with variance-covariance matrix, Σ_{GG}, where

$$(3.14) \qquad \Sigma_{GG} = \begin{bmatrix} \sigma_{11}(XX')^{-1} & \cdots & \sigma_{1G}(XX')^{-1} \\ \vdots & & \vdots \\ \sigma_{G1}(XX')^{-1} & \cdots & \sigma_{GG}(XX')^{-1} \end{bmatrix}$$

and Σ_{GG} is of size $(GK \times GK)$.

From (3.12) and (3.14), letting $[X_{kk'}^{-1}]$ be a typical element of $(XX')^{-1}$, we have for a typical element $[\sum_{k,k'} \sigma_{gg'} X_{kk'}^{-1} m_{kk'}]$ of $E(AM_FA')$ that

$$(3.15)\qquad [\sum_{k,k'} \sigma_{gg'} X^{-1}_{kk'} m_{k'k}] = \sigma_{gg'} \sum_{k,k'} X^{-1}_{kk'} m_{k'k} = \sigma_{gg'}\, q$$

where $q = X_F'(XX')^{-1}X_F$ is a scalar. So we find that

$$(3.16)\qquad E\,(AM_F A') = q\,\Sigma_{vv}\,.$$

From (3.10) and (3.16), we have for the variance covariance matrix of $Y_F - Y_F^*$

$$(3.17)\qquad \Sigma_{(Y_F-Y_F^*)} = q\,\Sigma_{vv} + \Sigma_{v_F v_F} = (1+q)\,\Sigma_{vv}$$

since

$$\Sigma_{v_F v_F} = \Sigma_{vv}\,.$$

In general we do not know Σ_{vv}; therefore we estimate Σ_{vv} employing the unbiased estimator $S_{\hat{v}\hat{v}}$ given by

$$(3.18)\qquad S_{\hat{v}\hat{v}} = (YY' - P(XX')P')/(N-K)$$

Thus, from (3.17) and (3.18), we finally have

$$(3.19)\qquad S_{(Y_F-Y_F^*)} = (1+q)S_{\hat{v}\hat{v}}\,.$$

In order to construct a forecast region (or to test hypotheses), we use the statistic T^2 where

$$(3.20)\qquad T^2 = (Y_F-Y_F^*)'\, S^{-1}_{\hat{v}\hat{v}}\, (Y_F-Y_F^*)\;/(1+q)$$

This statistic will have the Fisher F distribution if the following two conditions are met:*

---

* cf. [1, 105-106]

(i) $Y_F - Y_F^*$ is normally distributed with mean zero and variance-covariance matrix Σ_{vv}, and

(ii) $(N-K)\ S_{\hat{v}\hat{v}}^{-1}$ is distributed independently of $Y_F - Y_F^*$ as $\sum_{t=1}^{N-K} z_t z_t'$ with the z_t independent, each distributed as $N(0, \Sigma_{vv})$.

The first condition follows from (3.9), (3.17), and from the fact that $Y_F - Y_F^*$ is but a linear combination of disturbances which here are each assumed to be normally distributed. The second condition follows from the theorem* that

---

* cf. [1, 83-85]

P is distributed independently of $\Sigma z_t z_t'$, with the z_t independently distributed as $N(0, \Sigma_{vv})$. Since $Y_F - Y_F^* = (P - \Pi) X_F - V_F$, it is clear that $Y_F - Y_F^*$ is independent of $\Sigma z_t z_t'$. We have then that

$$\frac{(N-K-G+1)T^2}{(N-K)G} \tag{3.21}$$

is distributed as F with G and N-K-G+1 degrees of freedom.

We can now construct **forecast regions** (or test hypotheses) making use of (3.21). The procedure is to choose a confidence level (or level of significance), say α, and then to find the corresponding value of F in the F-tables. Then from (3.21) we can find the corresponding value of T^2, say T^2_α. The set of points for which the inequality

$$\left(\frac{N-K-G+1}{(N-K)G}\right) T^2 \leq F_\alpha \tag{3.22}$$

holds form the area of a region which is the forecast region. This region may be interpreted in the same way as the forecast interval for a single equation; i.e., if many samples are taken, holding X and X_F fixed, then $1-\alpha$ per cent of the time a region constructed as described above will cover the true value of Y_F, namely, Y_F^* .

It is interesting to determine those values of the exogenous variables selected for the forecast, i.e., X_F, which minimize the variance of the error of forecast given in (3.17). It is clear that (3.17) will be minimized when q is at a minimum value. For the present purpose, we write q as

$$(3.23) \qquad q = \sum_k \sum_{k'} X^{-1}_{kk'} \, m_{k'k} = X_F'(XX')X_F$$

To take account of the constant term in each equation, we let X_{1t} be a dummy variable, that is $X_{1t} \equiv 1$ for $t=1, 2, \ldots, N$, and this is also the value that X_1 assumes in the forecast period (i.e., $X_{1F} \equiv 1$). Then we can partition as follows:

$$(3.24) \qquad X_F = \begin{bmatrix} 1 \\ \cdots\cdots \\ X_{2F} \\ \vdots \\ X_{KF} \end{bmatrix} = \begin{bmatrix} 1 \\ \cdots\cdots \\ \tilde{X}_F \end{bmatrix}$$

$$(XX') = \begin{bmatrix} N & \sum_{t=1}^{N} X_{it} \\ \sum_{t=1}^{N} X_{it} & \sum_{t=1}^{N} X_{it}X_{jt} \end{bmatrix} = \begin{bmatrix} N & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, \quad (j, i=2, \ldots, K) .$$

Designating $(XX')^{-1}$ as $\begin{bmatrix} A & B \\ B' & D \end{bmatrix}$, we can thus determine this inverse matrix by solving the following equation system:

$$(3.25) \qquad \begin{bmatrix} A & B \\ B' & D \end{bmatrix} \begin{bmatrix} N & X_{12} \\ X_{21} & X_{22} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & I \end{bmatrix}$$

and we obtain

$$(3.26) \qquad (XX')^{-1} = \begin{bmatrix} A & B \\ B' & D \end{bmatrix} = \begin{bmatrix} (N + X_{12}DX_{21})/N^2 & -X_{12}D'/N \\ -DX_{21}/N & N(NX_{22} - X_{21}X_{12})^{-1} \end{bmatrix}$$

where $D=N(NX_{22}-X_{21}X_{12})^{-1}$. Making use of (3.24) and (3.26), we find that

$$(3.27) \qquad q = (N + X_{12}DX_{21})/N^2 - (2\tilde{X}_F' D X_{21})/N + \tilde{X}_F' D\tilde{X}_F .$$

Differentiating with respect to $\tilde{X}_F$ and setting the results equal to zero, we have

$$(3.28) \qquad -(DX_{21})/N + D\tilde{X}_F = 0 .$$

Thus we have for the minimizing value of $\tilde{X}_F$, say $\tilde{X}_F^*$

$$(3.29) \qquad \tilde{X}_F^* = X_{21}/N = (1/N)$$

which shows that the variance of the error of forecast is minimized when the values of the exogenous variables used in the forecast are set equal to the mean values of the exogenous variables used in estimating the parameters. The minimum value of q obtained by substituting (3.29) in (3.27) is,

$$(3.30) \qquad q_{Min} = 1/N + X_{12}DX_{21}/N^2 - (2X_{12}DX_{21})/N^2 + (X_{12}DX_{21})/N^2 = 1/N .$$

The minimum value of the variance of the error of forecast is then,

$$(3.31) \qquad \Sigma^{Min}_{Y_F - Y_F^*} = (1 + \frac{1}{N})\, \Sigma_{vv} .$$

IV. Construction of Forecast Regions: An Example.

For the purpose of illustrating the construction of the forecast region for a set of simultaneous equations we shall use Haavelmo model.* This model

---

* cf. [4, 83-91] for a complete description of this model.

---

may be written in the reduced form as

$$(4.1) \qquad \begin{aligned} c_t &= \Pi_{11}X_{1t} + \Pi_{12}X_{2t} + V_{1t} \\ y_t &= \Pi_{21}X_{1t} + \Pi_{22}X_{2t} + V_{2t} \end{aligned}$$

where the endogenous variables are consumers expenditures (c_t) and disposable income y_t . The exogenous variables are $X_{1t} \equiv 1$ and 'gross investment' (X_{2t}) . The error terms are V_{1t} and V_{2t} . The coefficients estimated by least-squares are

$$(4.2) \qquad \begin{aligned} P_{11} &= 298.554 \qquad & P_{12} &= 1.499 \\ P_{21} &= 285.787 \qquad & P_{22} &= 2.105 \end{aligned}$$

and the inverse of the estimated variance-covariance matrix of the error terms is

$$(4.3) \qquad S_{\hat{V}\hat{V}}^{-1} = \begin{bmatrix} .02507 & -.02570 \\ -.02570 & -.03125 \end{bmatrix} .$$

The inverse of the moment matrix of the exogenous variables is

$$(4.4) \qquad (XX')^{-1} = \begin{bmatrix} .7329396 & -.00701333 \\ -.00701333 & .00007498 \end{bmatrix}$$

We shall choose as X_F the vector $\begin{bmatrix} 1 \\ 100 \end{bmatrix}$ whose components are the values of the exogenous variables used in the forecast. The value of $X_{2F}=100$ is close to the mean of this variable in the sample period ($\bar{X}_2 = 93.5385$). We then obtain that,

$$(4.5) \qquad (1 + q) = 1.08007$$

Now from (3.21) we have that $\frac{T^2 \, (N-K-G+1)}{(N-K)G}$ has the "F" distribution, $F_{G,N-K-G+1}$. In our case N=13, K=2, G=2. So choosing the 5 per cent level of significance we find that F=4.10. The corresponding value of T^2 is:

$$(4.6) \qquad T^2_{5\%} = \frac{(4.10)(1.08007)(11)(2)}{10} = 9.742$$

As the forecast values of the variables we have

$$(4.7) \qquad \begin{aligned} c_F &= 448.454 \\ y_F &= 496.287 \end{aligned}$$

Using this and (4.3), (4.4) and (4.6) we obtain

$$(4.8) \qquad 9.742 = 0.02507\ (c_F-c_F^*)^2 - .05139\ (c_F-c_F^*)(y_F-y_F^*) + 0.03125(y_F-y_F^*)^2 .$$

This represents an ellipse in the parameter space of c_F^* and y_F^* where the center of the ellipse is c_F, y_F and the area covered by the ellipse is the forecast region for c_F^* and y_F^*, the latter being the true values of the endogenous variables in the forecast period. This is ellipse A in Fig. 1. Ellipse B is the forecast region corresponding to an $X_{2F}=200$ for the value of gross investment in the forecast period.

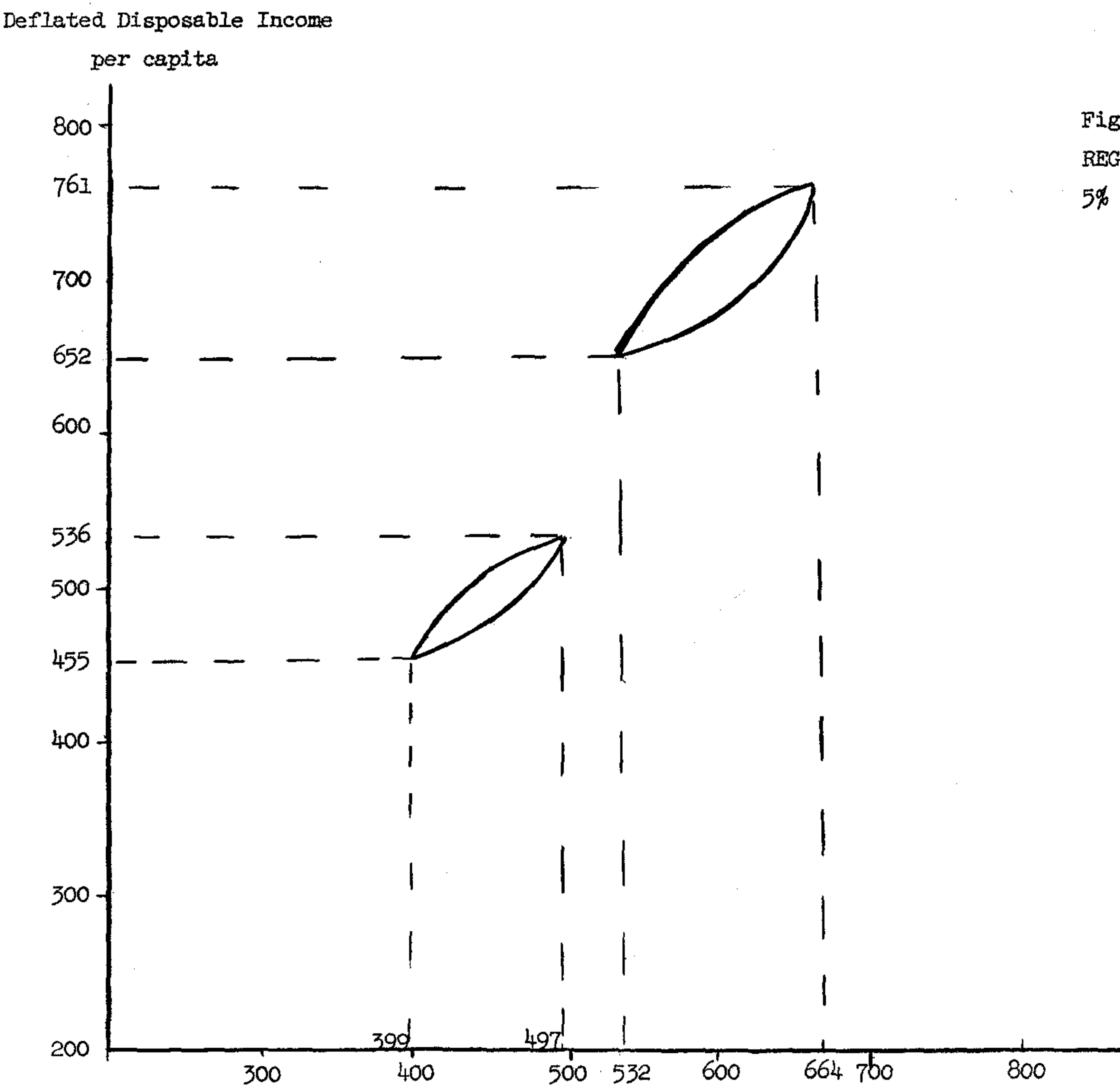

Figure 1: JOINT FORECAST REGIONS FOR C_F^* AND Y_F^*; 5% LEVEL OF SIGNIFICANCE.

V. Conclusions

Our analysis has yielded an expression for the standard error of forecast for interdependent reduced form equations. In an illustrative calculation this result has been employed to construct forecast regions. Calculation of such regions is extremely important in appraising the forecasts provided by reduced form systems. Since point forecasts, unaccompanied by a forecast region, may on occasion be seriously misleading, we recommend that those who forecast take the additional trouble needed to construct forecast regions.

Further, our work provides what is necessary to test hypotheses about forecasts. Thus, for example, it is possible to test the hypothesis that a forecast from a reduced form system is not significantly different from a "judgment forecast." Perhaps the performance of such tests would indicate that in many cases prolonged discussion of which of the two forecasts is better is not necessary.

Finally we wish to point out several limitations associated with our result. First, probability statements made about forecast regions constructed as we indicated above are valid providing the vector of exogenous variables employed in making the forecast is fixed, that is non-stochastic. Given this condition, one can state, as is usually done, that the constructed region will cover the true values of the endogenous variables in the forecast period a certain proportion of the time in repeated trials. However, in an actual set of repeated forecasts it may be impossible to hold fixed all the exogenous variables and in this situation the interpretation given above for the forecast region does not apply.* A similar situation prevails when some of the

---

* R. A. Fisher discusses this problem in a different context. cf.[3, pp. 83-85] .

exogenous variables have to be forecasted; that is, the vector X_F can no longer be regarded as non-stochastic. Also, it is interesting to note that a similar situation is encountered if one tries to forecast from the equations provided in the second stage of the "two-stage least squares" method of estimation.*

---

* See [2] and [6].

REFERENCES

[1] ANDERSON, T. W., An Introduction to Multivariate Statistical Analysis, New York: John Wiley and Sons, Inc., 1958.

[2] BASMANN, R. L., "A Generalized Classical Method of Linear Estimation of Coefficients in a Structural Equation," Econometrica, Vol. 25, January, 1957, pp. 77-83.

[3] FISHER, R. A., Statistical Methods and Scientific Inference, Edinburgh and London: Oliver and Boyd, 1956.

[4] HAAVELMO, T., "Methods of Measuring the Marginal Propensity to Consume," Chapter IV in Wm. C. Hood and T. C. Koopmans, Studies in Econometric Method, New York: John Wiley and Sons, Inc., 1953, pp. 75-91, reprinted from Journal of the American Statistical Association, Vol. 42, March, 1947, pp. 105-122.

[5] STONE, R., Measurement of Consumers' Expenditure and Behavior in the United Kingdom, 1920-1938, Cambridge: The University Press, 1954.

[6] THEIL, H., Economic Forecasts and Policy, Amsterdam: North-Holland Publishing Company, 1958.