Annals of Economic and Social Measurement, 3/3, 1974

RESEARCH METHODOLOGY NOTES

A SIMPLE SUGGESTION TO IMPROVE THE MINCER-ZARNOWITZ CRITERION FOR THE EVALUATION OF FORECASTS

BY MICHIO HATANAKA

The wide use of the Mincer–Zarnowitz criterion[1] to evaluate economic forecasts seems to call for clarification of its logical implication. Further consideration of the criterion leads to a simple suggestion to improve the criterion.[2] This note concludes with such a suggestion.

Let y be the variable to be predicted, and $\hat{y}$ be a prediction of y. Mincer and Zarnowitz proposed to judge $\hat{y}$ in terms of whether or not the condition,

$$E(\hat{y}) = E(y), \tag{1a}$$

$$\operatorname{cov}(y, \hat{y})/\operatorname{var}(\hat{y}) = 1, \tag{1b}$$

is met. I shall call (1a)–(1b) the original M–Z condition. Provided that for a $\hat{y}$

$$E(y|\hat{y}) = E(y) + \beta(\hat{y} - E(\hat{y})),\ \beta = \operatorname{cov}(y, \hat{y})/\operatorname{var}(\hat{y}) \tag{2}$$

holds, we have

$$E(y - \hat{y})^2 = [E(y) - E(\hat{y})]^2 + (1 - \beta)^2 \operatorname{var}(\hat{y}) + E(v^2) \tag{3}$$

where $v = y - E(y|\hat{y})$. The sum of the first two terms on the right hand side measures the extent to which the $\hat{y}$ fails to satisfy the original M–Z condition.

Notice that the original M–Z condition involves only the actual and the predicted values of y. It does not specify the model in which y is determined, and, it does not recognise the variables on the basis of which $\hat{y}$ is obtained. However, it should also be noted that (2) does not hold unless one presupposes some kind of linearity of the model behind y.

Let us temporarily take the standpoint just opposite of the Mincer and Zarnowitz's. The reason for it will become clear as the discussion proceeds. Let X be a vector and y a scalar. Suppose that there is a model which specifies the joint distribution of (y, X), and also that this distribution is known completely.[3] Also assume that the prediction is judged in terms of its mean square error. Let $E(y|X)$ be the mean of y conditional upon X. Among all functions $g(X)$ the minimum of $E(y - g(X))^2$ is attained when $E(y|X)$ is chosen as $g(X)$.[4] $E(y|X)$ is an optimum

[1] Jacob Mincer and Victor Zarnowitz, "The Evaluation of Economic Forecasts," in J. Mincer (ed.), *Economic Forecasts and Expectations* (National Bureau of Economic Research, New York, 1969).

[2] I have been led to consider this problem by C. W. J. Granger and P. Newbold, "Some Comments on the Evaluation of Economic Forecasts," *Applied Economics*, 5 (March, 1973), pp. 35–47, especially p. 40.

[3] I assume that $E(y)$, $E(X)$, and $E(y - E(y|X))^2$ exist. I shall comment on the other second-order moments later.

[4] See Harald Cramer, *Mathematical Methods of Statistics* (Princeton University Press, 1946), p. 272.

prediction. If in addition the distribution is known to be normal, $E(y|X)$ is obtained through the ordinary regression of y upon X.

The relationship between the original M–Z condition and $E(y|X)$ is summarized in the following two statements.

Statement 1. When $E(y|X)$ is chosen for $\hat{y}$, it satisfies the original M–Z condition.

Statement 2. Suppose that $E(y|X)$ is a linear function of X. Then the $\hat{y} = a + B'X$ which satisfies the original M–Z condition is optimal over the class of predictions that can be written as $c + d\hat{y}$ with some constants c and d, but, the $\hat{y}$ is not generally optimal over the unrestricted class of predictions.

A few words of explanations are in order. In both Statements I assume that (1b) holds whenever the nominator and the denominator take identical forms even though the cov. and the var. may involve the integrals that diverge to ∞. As for Statement 1 it is easy to construct a suboptimal prediction which satisfies the original M–Z condition. The failure to meet the original M–Z condition proves $\hat{y}$ suboptimal, but, the success does not prove $\hat{y}$ optimal. As for the Statement 2 a, c, and d are scalars and B is a vector. Statement 2 means that the original M–Z condition ensures the optimality only *over the one-dimensional space* of predictions in the linear model. A corollary of Statement 2 is the following. If X involves a single variable x, and $E(y|x) = \alpha + \beta x$, then an arbitrary linear predictor, $\hat{y} = a + bx$, satisfies the original M–Z condition if and only if $a = \alpha$ and $b = \beta$.[5] The proof of the Statement 2 may be sketched. The assumptions of the Statement imply that $E(y|\hat{y})$ is also linear in $\hat{y}$. Then $E(y|\hat{y}) = E(y) + \beta(\hat{y} - E(\hat{y}))$, where $\beta = \text{cov}(y, \hat{y})/\text{var}(\hat{y})$.[6] By the assumption $\beta = 1$. Thus $E(y|\hat{y}) = \hat{y}$. $E(y - (c + d\hat{y}))^2 = E(y - \hat{y})^2 + E(c + d\hat{y} - \hat{y})^2$, which is minimized by choosing $c = 0$ and $d = 1$.

This note takes the same approach as Mincer and Zarnowitz. Given the current state of econometrics, most models have possibly a very serious specification error. In evaluating economic forecasts one should have a high regard on a judgment which does not rely upon the validity of a specific model. Mincer and Zarnowitz provided a criterion for just such a judgment.

However the meaning of the criterion cannot be investigated without some minimum presupposition about the model. ((2) itself assumes some.) Statements 1 and 2 have been the result of such an investigation. I propose a simple method to strengthen the Mincer–Zarnowitz criterion. The new criterion is also stated with no reference to a model, but its justification will be made in terms of its implication in a model.

Let $(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)$ be a given set of linearly independent[7] predictions of y. Consider the condition,

$$E(y) = E(\hat{y}_0) \tag{4a}$$

$$\text{cov}(y; \hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)\,\text{var}(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)^{-1} = (1, 0, \ldots, 0), \tag{4b}$$

[5] I think therefore that Granger and Newbold, op. cit. have made a bit too severe criticism of the original M–Z condition.

[6] See Cramer, op. cit., p. 274.

[7] In the terminology of the matrix algebra.

where cov $(y;\ \hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)$ is the row vector that consists of cov $(y, \hat{y}_0)$, cov $(y, \hat{y}_1), \ldots,$ cov $(y, \hat{y}_k)$, and var $(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)$ is the covariance matrix of $\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k$. I say $\hat{y}_0$ is the best in the set $(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)$ when (4a) and (4b) hold. (4a) and (4b) will be called the strengthened M–Z condition. I assume that (4b) holds whenever the first column of var $(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)$ and the vector cov $(y;\ \hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)$ have the identical forms even though they may involve diverging integrals.

The justification of this criterion is in the following two Statements.

Statement 3. Let $\hat{y}_1, \ldots, \hat{y}_k$ be some functions of X. Then the set $(E(y|X), \hat{y}_1, \ldots, \hat{y}_k)$ satisfies the strengthened M–Z condition, provided that $E(y|X)$, $\hat{y}_1, \ldots, \hat{y}_k$ are linearly independent.

Statement 4. Suppose that $E(y|X)$ is a linear function of X, that $\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k$ are linear functions of X and linearly independent, and that the set $(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)$ satisfies the strengthened M–Z condition. Then $\hat{y}_0$ is optimal over the class of predictions that can be written as $c + d_0\hat{y}_0 + d_1\hat{y}_1 + \ldots + d_k\hat{y}_k$ with some constants, $c, d_0, d_1, \ldots, d_k$, but, $\hat{y}_0$ is not generally optimal over the unrestricted class of predictions.

Thus the failure to meet the strengthened M–Z condition proves $\hat{y}_0$ suboptimal, but, the success still does not prove $\hat{y}_0$ optimal. In the linear model, however, *we have now expanded the dimension* of the space over which the optimality can be claimed.

If the set $(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)$ does not satisfy the strengthened M–Z condition, construct

$$\text{(5)} \qquad \hat{y}^* = E(y) + \text{cov}\,(y;\ \hat{y}_0, \hat{y}_1, \ldots \hat{y}_k)\, \text{var}\,(\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k)^{-1}$$

$$\text{times} \begin{pmatrix} \hat{y}_0 - E(\hat{y}_0) \\ \hat{y}_1 - E(\hat{y}_1) \\ \vdots \\ \hat{y}_k - E(\hat{y}_k) \end{pmatrix}.$$

Statement 5. Suppose that $E(y|X)$ is a linear function of X and that $\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k$ are linear functions of X and linearly independent. Then each of the sets, $(\hat{y}^*, \hat{y}_1, \ldots, \hat{y}_k)$, $(\hat{y}^*, \hat{y}_0, \hat{y}_2, \ldots, \hat{y}_k), \ldots, (\hat{y}^*, \hat{y}_0, \hat{y}_1, \ldots, \hat{y}_{k-1})$ satisfies the strengthened M–Z condition.

Statement 6. Under the same assumptions as in Statement 5

$$\text{(6)} \quad E(y - \hat{y}_i)^2 = (E(y) - E(\hat{y}_i))^2 + \text{var}\,(\hat{y}_i - \hat{y}^*) + E(y - \hat{y}^*)^2, \qquad i = 0, 1, \ldots, k.$$

Suppose that the records of predictions $\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k$ over $t = 1, \ldots, T (T > k + 1)$ are available for the evaluation. Run the least squares, using the data of y over $t = 1, \ldots, T$ as the dependent variable and the records of $\hat{y}_0, \hat{y}_1, \ldots, \hat{y}_k$ as the independent variables. Let the coefficients thus obtained be $c, d_0, d_1, \ldots, d_k$ where c is the constant term. Construct a new prediction, $\hat{y}^* = c + \sum_{i=0}^{k} d_i\hat{y}_i$. The logic of the least squares assures the *identity*:

$$\text{(7)} \qquad \text{M.S.E. of } \hat{y}_i = (\text{bias of } \hat{y}_i)^2 + \text{var of } (\hat{y}_i - \hat{y}^*) + \text{M.S.E. of } \hat{y}^*,$$

$$i = 0, 1, \ldots, k$$

which is a sample counterpart of (6). The last term on the right hand side of (7) may be regarded as the mean square error of the prediction that is optimal over the class of predictions represented in the form $\gamma + \delta_1 \hat{y}_1 + \ldots + \delta_k \hat{y}_k$ with some constants $\gamma, \delta_1, \ldots, \delta_k$. The sum of the other two terms measures the failure of $\hat{y}_i$ to be identical to this optimal prediction.

Osaka University, Japan

Submitted March 1974