COWLES COMMISSION DISCUSSION PAPER: ~~ECONOMICS~~ Statistics NO. ~~285~~ 348

NOTE: Cowles Commission Discussion Papers are preliminary materials circulated privately to stimulate private discussion and are not ready for critical comment or appraisal in publications. References in publications to Discussion Papers (other than mere acknowledgment by a writer that he has had access to such unpublished material) should be cleared with the author, to protect the tentative character of these papers.

On Specification Bias

By Leonid Hurwicz

0. In making statistical decisions we usually start by assuming certain properties of the universe (or model) from which the observations are being drawn. It may, of course, happen that the statistician will fail to make use of the assumptions he has stated, i.e., he may not be using an optimal decision function relative to his assumptions.* (However, the fact of not using an optimal decision function does not necessarily imply that the assumptions made are not being used.) Furthermore, it sometimes happens that widely varying assumptions will lead to the same optimal decision function. (E.g., in a univariate normal sample the maximum likelihood estimate of the mean is the sample mean regardless of what is assumed (correctly or not) with regard to the variance. But note that the assumptions with regard to the mean would affect the estimates of the variance.)

The general problem of specification bias may be formulated as that of finding out what happens if the assumptions made are incorrect. (In a narrower sense, we are often only paying attention to the first moment of the statistic used.)

---

* An example would be that of estimating the population median in a normal univariate sample through the sample median.

1. Thus let $\mathcal{S}$ be the class of all structures and $\mathcal{S}^A$ ($\mathcal{S}^A \subsetneq \mathcal{S}$) the statistician's model, i.e., the class of structures he considers admissible. (Later on we shall consider cases where the statistician associates probabilities with his models.) Suppose the problem is one of estimation (the generalization to other types of statistical decisions is straight-forward) and let $d(x, \mathcal{S}^A)$ be the estimator function given the observation(s) $\underline{x}$ and the model $\mathcal{S}^A$. (The function $\underline{d}$ need not be optimal.) We shall write $T = d(x, \mathcal{S}^A)$ where T is the estimate of the unknown parameter(s) θ. The object of our interest is the distribution $F(T/S^o)$ of T given S^o where S^o is the true structure. More specifically we may be interested in seeing how F depends on $\mathcal{S}^A$ and on S^o. The general problem of specification bias is that of finding how F depends on S^o when S^o is not in the model $\mathcal{S}^A (S^o \notin \mathcal{S}^A)$, i.e., when the model is "false."

One might raise the question as to the purpose of investigating the nature of dependence ψ of F on S^o and $\mathcal{S}^A$. (We write $F = \psi(\mathcal{S}^A, S^o)$.) One case where the knowledge of ψ is useful is as follows: Suppose Statistician I has computed an estimate $T = d(x, \mathcal{S}^A)$. Statistician II wishes to use the estimate T, but he believes that $\mathcal{S}^A$ is a false model, i.e., $S^o \notin \mathcal{S}^A$. Under such circumstances Statistician II, if he has no access to the observation $\underline{x}$ on which T is based can be guided in his use of T by the knowledge of ψ and $\mathcal{S}^A$. (Clearly, ψ will also depend on $\underline{d}$, but it seems more convenient at this stage not to indicate it in the notation; we may think of it as being fixed.)

2. In what follows we shall present an extremely simple example illustrative of the concepts introduced. The decision function $\underline{d}$ will be the maximum likelihood method of estimation. (This method has the following advantage from the viewpoint of the present investigation. If we think of $\mathcal{S}^A$ as a parameter space, the maximum likelihood method will never yield an estimate which is outside $\mathcal{S}^A$.) Later on examples of more direct interest to the applied econometrician will be presented.

Example 1.

Let a sample of size one be drawn from a bivariate normal universe with a known covariance matrix Σ. The likelihood function is given by

(1) $$P = \frac{1}{2\pi |\Sigma|^{1/2}} \exp\left\{ -\frac{1}{2} \sum_{i,j=1}^{2} \sigma^{ij}(X_j - \theta_j) \right\}$$

where the two means are denoted by θ_i. We shall assume that the statistican's model S^A consists of assuming equation (1) and

(2) $\theta_2 = \theta_2^*$

where θ_2^* is a constant known to the statistician.

We shall now find the maximum likelihood estimate of θ_1 computed on the assumption that both (1) and (2) hold. (This estimate will be denoted by $\hat{\theta}_1^*$, as distinct from the maximum likelihood estimate $\hat{\theta}_1 = X_1$ obtained without assuming (2).) Then we shall study the distribution of $\hat{\theta}_1^*$ in the class of cases where (2) is false, i.e., where $\theta_2^0 \neq \theta_2^*$. ($\theta_i^0$ denotes the true value of θ_i.)

To facilitate computation we rewrite (1) in the factored form and incorporating assumption given in equation (2) as

(3) $$P^* = \frac{1}{\sqrt{2\pi}\,\sigma_{1.2}} \exp\left\{ -\frac{1}{2\sigma_{1.2}^2} [(X_1 - \theta_1) - \gamma(X_2 - \theta_2^*)]^2 \right\} \frac{1}{\sqrt{2\pi}\,\sigma_{22}} \left\{ -\frac{1}{2\sigma_{22}} (X_2 - \theta_2^*)^2 \right\}$$

where $\sigma_{1.2}^2 = \sigma_{11}(1 - \rho^2)$, $\rho = \frac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$, $\gamma = \frac{\sigma_{12}}{\sigma_{22}}$.

The maximization of P^* with regard to θ_1, θ_2^* being considered fixed yields the desired estimate $\hat{\theta}_1^*$

(4) $\hat{\theta}_1^* = X_1 + \gamma(\theta_2^* - X_2)$

The distribution of $\hat{\theta}_1^*$ is normal with

$$(5)\begin{cases} E\,\hat{\theta}_1^{*} = \theta_1^{o} + \gamma(\theta_2^{*} - \theta_2^{o}) \\ E[\hat{\theta}_1^{*} - E\hat{\theta}_1^{*}]^2 = \sigma_{11} - \gamma^2\sigma_{22} \end{cases}$$

so that

$$(6)\quad E[\hat{\theta}_1^{*} - \theta_1^{o}]^2 = (\sigma_{11} - \gamma^2\sigma_{22}) + \gamma^2(\theta_2^{*} - \theta_2^{o})^2 = \sigma_{11} - \gamma^2[\sigma_{22} - (\theta_2^{*} - \theta_2^{o})^2]$$

Clearly, when $\gamma = 0$, we have $\hat{\theta}_1^{*} = \hat{\theta}_1$, and no use is made of θ_2^{*}. Hence it does not matter whether $\theta_2^{*} \neq \theta_2^{o}$. However, when $\gamma \neq 0$, $\hat{\theta}_1^{*}$ is a better or worse estimate (as measured by the expression in equation (6)) depending on the absolute size of the difference between θ_2^{*} and θ_2^{o}. In particular, since

$$(7)\quad \begin{aligned} &E\,\hat{\theta}_1 = \theta_1^{o} \\ &E[\hat{\theta}_1 - \theta_1^{o}]^2 = \sigma_{11}, \end{aligned}$$

it follows that $\hat{\theta}_1^{*}$ is superior (equally good, inferior) to $\hat{\theta}_1$ when the expression $\sigma_{22} - (\theta_2^{*} - \theta_2^{o})^2$ is positive (zero, negative).

3.1 Now consider the problem faced by the statistician whose model is that expressed by equations (1) and (2) with the following modification: instead of being certain that (2) holds, the statistician has a (subjective) probability α that (2) holds. Furthermore, he admits only one alternative value θ_2^{**} with probability $1 - \alpha$. In this case three possible courses of action suggest themselves (others might also be considered). These are to use $\hat{\theta}_1$, $\hat{\theta}_1^{*}$, or $\hat{\theta}_1^{**}$; these estimates being respectively obtained by the maximization of (1) with regard to θ_1 subject to no restrictions, the restriction $\theta_2 = \theta_2^{*}$, and the restriction $\theta_2 = \theta_2^{**}$. Suppose the expression the statistician wishes to minimize is $v_{T_1} \equiv E[T_1 - \theta_1^{o}]^2$ where T_1 is the estimate of θ_1 used. Clearly

$$(8)\begin{cases} v_{\hat{\theta}_1} = \sigma_{11} \\ v_{\hat{\theta}_1^{*}} = \alpha(\sigma_{11} - \gamma^2\sigma_{22}) + (1-\alpha)\left\{\sigma_{11} - \gamma^2[\sigma_{22} - (\theta_2^{*} - \theta_2^{**})^2\right\} \\ v_{\hat{\theta}_1^{**}} = (1-\alpha)(\sigma_{11} - \gamma^2\sigma_{22}) + \alpha\left\{\sigma_{11} - \gamma^2[\sigma_{22} - (\theta_2^{**} - \theta_2^{*})^2]\right\}. \end{cases}$$

It can be seen that $\hat{\theta}_1^*$ is preferable to $\hat{\theta}_1$ if the expression $\sigma_{11} - \gamma^2[\sigma_{22} - (1 - \alpha)(\theta_2^* - \theta_2^{**})^2]$ is positive, i.e., if $\alpha > 1 - \frac{\sigma_{22}}{(\theta_2^* - \theta_2^{**})^2}$. Similarly, $\hat{\theta}_1^{**}$ is preferable to $\hat{\theta}_1$ if $1 - \alpha > 1 - \frac{\sigma_{22}}{(\theta_2^* - \theta_2^{**})^2}$.

Depending on the value of α in relation to the right-hand member of the last inequality (and its complement to 1), it can be shown that each of the three estimates $\hat{\theta}_1$, $\hat{\theta}_1^*$, $\hat{\theta}_1^{**}$ may be inferior to the other two.

3.2 The case where no a priori probabilities (of θ_2^* versus θ_2^{**}) are involved (α unknown) is also of interest. Here the maximum likelihood estimate (denoted by $\hat{\hat{\theta}}_1$) is given by

(9) $\hat{\hat{\theta}}_1 = X_1 + \gamma(\bar{\theta}_2 - X_2)$

where $|\theta_2^* - X_2| < |\theta_2^{**} - X_2|$, then $\bar{\theta}_2 = \theta^*$ (and vice versa). (This is quite instructive since it might be regarded as a two-stage procedure consisting of a "preliminary" test as to whether $\theta_2 = \theta_2^*$.) With the help of equation (8) one may set up the "pay-off" matrix for this case and consider it from the viewpoint of minimising the maximum risk or an alternative principle.

3.3 A fact of practical importance is connected with internal estimation of θ_1. Such an interval is obtained on the assumption that $\theta_2 = \theta_2^*$ (while $\theta_2^* \neq \theta_2^0$) would not only be improperly "centered" (say around $\hat{\theta}_1^*$) but would be excessively short since the "apparent" variance $\sigma^{*2}_{\hat{\theta}_1}$ of $\hat{\theta}_1^*$ computed on the assumption that $\theta_2 = \theta_2^*$ would be $\sigma_{11} - \gamma^2 \sigma_{22}$ instead of the true expression in (6). (In extreme cases the "apparent" variance might even be zero!)

4.0 From the viewpoint of econometric applications two cases of specification bias are of special interest.

4.1 "Partial system" bias.

The well-known special case of this is the use of the "single-equation" approach. In general, it consists in considering a part of the system as if it were complete. Let the system be written (in matrix form) as

$$(10)\begin{cases} (10.1) & \beta_{11} y_1 + \beta_{12} y_2 + \Gamma_1 z = u_1 \\ (10.2) & \beta_{21} y_1 + \beta_{22} y_2 + \Gamma_2 z = u_2 \end{cases}$$

where y_1, y_2, z, u_1, u_2 are all column vectors. Now suppose the objective is to estimate the parameters of (10.1) and that this is done on the (possibly false) assumption that

$$(11)\begin{cases} \beta_{21} = 0 \\ E\, u_1 u_2 = 0 \end{cases}$$

Assuming (11) implies that (10.1) is a complete system. If, in fact, (11) does not hold, what would be the bias in the estimates of β_{11}, etc? I shall here only indicate how the asymptotic bias can be obtained. (Reference: Paper by Jean Bronfenbrenner.) The (biased) maximum likelihood estimates converge stochastically to the same values as the "instrumental variable" estimates (or the "just identified" case they are identical). Hence we consider the latter, thus obtaining the asymptotic relations

$$(12.1)\quad \beta^*_{11} E\, y_1 y_2' + \beta^*_{12} E\, y_2 y_2' + \Gamma^*_1 E\, z\, y_2' = 0$$

where θ^* is the asymptotic value of θ estimated on the assumption that (11) holds. The other relation is

$$(12.2)\quad \beta^*_{11} E\, y_1 z' + \beta^*_{12} E\, y_2 z' + \Gamma^*_1 E\, z\, z' = 0.$$

(Of course, additional identifying restrictions are also needed.)

Since the true values of $E\, y_1 y_2'$, etc. can be obtained from the "reduced form" of equation (10) in terms of the z's and the true parameters are substituted into equation (12), the bias $\beta^*_{11} - \beta_{11}$, etc. can be evaluated. (A simple example will be supplied later.)

4.2 Another case of econometric interest arises in connection with autoregressive shock-error models. Consider the simple case

$$(13)\quad \begin{cases} \xi_t - \alpha\, \xi_{t-1} = u_t, \quad |\alpha| < 1 \\ x_t = \xi_t + v_t \end{cases}$$

where x is the observed, ξ the true value of the variable; u and v are the stock and error random components respectively.

Most econometric work assumes either

(14.1) $\quad v_t = 0 \qquad$ ("shock-model")

or

(14.2) $\quad u_t = 0 \qquad$ ("error model").

The question of bias arises when either of equations (14) is assumed when it does not hold. For instance, suppose (14.1) is assumed when it is not true. The maximum likelihood estimate $\hat{\alpha}^{\#} = \frac{\Sigma X_t X_{t-1}}{\Sigma X_{t-1}^2}$ of α then obtained converges stochastically to $\alpha^{\#} = E\, X_t X_{t-1} / E\, X_t^2$. Now, in a model specified by (13), we have

$$(15)\quad \begin{cases} E\, X_t X_{t-1} = E\, \xi_t \xi_{t-1} = \alpha \dfrac{\sigma_u^2}{1-\alpha^2} \\ E\, X_t^2 = E\, \xi_t^2 + \sigma_v^2 = \dfrac{\sigma_u^2}{1-\alpha^2} + \sigma_v^2 \end{cases}$$

hence

$$(16)\quad \alpha^{\#} = \alpha \frac{1}{1 + \frac{\sigma_v^2}{\sigma_u^2}(1-\alpha^2)}$$

and is clearly biased when (14.1) does not hold. (Reference: T. W. Anderson and L. Hurwicz.)

4.3. A reference may be made to the system

$$(17)\quad \begin{cases} y_1 + \beta y_2 = u_1 \\ y_2 + \gamma_1 z_1 + \gamma_2 z_2 = u_2. \end{cases}$$

The question arises when z_2 is ignored and the system used is

$$(18)\quad \begin{cases} y_1 + \beta y_2 & = u_1 \\ y_2 + \bar{\gamma}_1 z_1 = \bar{u}_2. \end{cases}$$

It has been shown by Steve Allen that when (18) is used, the estimate $\hat{\hat{\beta}}$ of β is less efficient than that based on (17). However, $\hat{\hat{\beta}}$ is not biased.

The writer was originally inclined to classify this problem as one of specification bias and thought at first that the reason for absence of bias was perhaps analogous to the case $\gamma = 0$ in Example 1 (section 2) above. However, in the light of a remark made by Hildreth it became clear that this case is not one of specification bias; i.e., there exist $\bar{\gamma}_1 \neq \gamma_1$, $\bar{u}_2 \neq u_2$, such that (18) holds and $E\, z_1 u_2 = 0$. Hence if (17) is true, so is (18) and there is no specification bias. [In the notation of section 1 above, we have $S^o \in \mathcal{S}^A$ whether S^o is written as (17) or (18).]

5.0. In connection with Example 1 we considered the following situation. It is known that θ_2 must have one of the two values θ_2^* or θ_2^{**}. There may or may not exist as a priori probability α that $\theta_2 = \theta_2^*$, say. If the probability exists, we have the "Bayes" case. If it does not, we have the case of "ignorance" case. (We shall disregard here the complications due to the fact that there is another unknown parameter -- θ_1 -- in the problem.)

It seems natural to ask whether there can exist an "intermediate" situation, e.g., when it can be said that $\alpha \geq .5$, but its value cannot be specified more precisely.

5.1. The preceding question may be reformulated in the language of the theory of games. Let there be two players and a "payoff" matrix M

1 \ 2	A	B
S	u_{SA}	u_{SB}
T	u_{TA}	u_{TB}

where the u's are (measurable) utilities. Denote by η the probability that Player 2 will choose A rather than B. The "Bayes" case consists in Player 1 knowning that $\eta = \eta_o$. The "ignorance" case consists in Player 1 only knowing that $\eta \in I$ where I is the closed interval [0, 1].

The "intermediate" case consists in Player 1 knowing that $\eta \in I_o$ where $I_o \subset I$ and I_o consists of two or more points. When I_o is a closed convex set, it is possible to set up a game with a new payoff matrix M_o such that in terms of M_o we again have the case of "ignorance" and the strategic solutions are the same as those for M with the additional information that $\eta \in I^o$.

5.2. One may also wish to consider the case where it is considered that $\eta \in I^o$ with a specified probability. This can also be reduced to the case of "ignorance."

5.3. The problem presented in 5.1 is of relevance in the following type of situation, related to that in Example 1, section 2.

Let it be known that $\theta_2 = \theta_2^{*}$ or $\theta_2 = \theta_2^{**}$. Now suppose that on a priori grounds θ_2^{*} is "more probable," though one cannot say by how much. In game-theoretical language used by Wald this may be interpreted as saying: "We have a tip that the strategy selected by Nature is not one of those in which the probability of θ_2^{**} exceeds 1/2." Such a "tip" would enable us to improve our strategy.

6.0. The topic of this paper is closely connected with the problem of choosing proper identifying assumptions.

Of course, the latter problem would not arise if (a) the assumptions did not happen to affect the estimates, or (b) we were 100 percent certain of the correctness of our assumptions. As a rule neither (a) nor (b) hold. However, it may be that we consider the identifying assumptions to be quite plausible, i.e., we are willing to assign to them some fairly high (though unspecified) a priori probability. In this case the question arises whether utilizing the assumption is worth while. The solution must balance the advantages of the cases where the assumption was correct against the disadvantages of the others. The latter are in the nature of specification bias.

It should be noted that in some, though not all cases, the optimal procedure might consist in a "preliminary" test of the identifying assumptions, to be followed by the utilization or non-utilization of the assumption, depending on the outcome of the test.

6.1. A simple example of the identification problem is the following: Let it be known that

(19) $X_1 + \alpha X_2 = u$

where X_1, X_2, u are random variables and α is the parameter to be estimated. Clearly, (19) is not identified (though limits for α can be obtained). Now it may be considered very likely on a priori grounds that

(20) $E\, X_2\, u = 0.$

The econometrician's problem is whether or not to use (20). To answer this question we must know the payoff (weight) function. In particular, we must know the value of that function for the case where the econometrician believed as if (20) were true (i.e., used the single equation approach) while actually (20) was false. This is precisely the problem of specification bias as defined in this paper. We may say that the knowledge of the specification bias is needed for the construction of the payoff function.