Econometrica, Vol. 52, No. 4 (July, 1984)

LOCAL ASYMPTOTIC SPECIFICATION ERROR ANALYSIS

BY NICHOLAS M. KIEFER AND GARY R. SKOOG[1]

An approximation to the inconsistency introduced by imposing an incorrect restriction on a parametric model is given. The approximation can be applied to estimators generated by optimizing any objective function satisfying certain regularity conditions. Examples given include analysis of misspecification in discrete choice and time-series models estimated by maximum likelihood, and in a nonlinear regression model.

SPECIFICATION ERROR ANALYSIS in the linear regression model has been studied by Theil [1], who gives formulas for, e.g., the effect of leaving out relevant variables on the expected values of the estimators of the coefficients of the included variables. In this paper we suggest analogous formulas for estimators obtained by optimizing an objective function subject to restrictions. We have in mind maximizing $(1/n)\times$ loglikelihood and will usually use this terminology. We consider the effect on the limit of the restricted estimator of a small violation of the restrictions. In the linear regression case our formula coincides with that given by Theil.

In order to keep our results widely applicable and to avoid a mass of unnecessary detail we make assumptions on the asymptotic behavior of the loglikelihood function itself, rather than on the data-generating process *per se*. Many alternative sets of assumptions on the data densities can lead to the behavior we require of the loglikelihood functions. These will not be pursued here. The interested reader is referred to, e.g., White [12] for the case of independent observations and Kohn [8] for the time-series case.

1. GENERAL FORMULAS

The general approach we take is based on a linear approximation to the likelihood function at the maximum likelihood estimator. It is in this sense that our analysis is local. For some models the local and global specification error results coincide; a well known case is the effect of omitted regressors in the linear regression model. Essentially the only cases involve linearity, although often there is agreement regarding the signs of the inconsistency. We show below that the local and global results even fail to coincide in the case of misspecified AR processes. Generally however, the global results are unknown.[2]

Taylor expansions are typically used together with assumptions on the data generating process to obtain the asymptotic distribution of the maximum likelihood estimator (Cramer [3]). In this paper we will not concern ourselves with asymptotic distributions of $\sqrt{n}$-normed MLE's since these have been worked

[1] We are indebted to the referees for helpful comments. This research was partly supported by the National Science Foundation.

[2] Precisely, the relationship between the parameters of the misspecified and the correct model is not known. It can be shown that the almost sure limits of parameter estimates of the misspecified model are given so as to minimize the Kullback–Leibler information loss relative to the true model. See White [12].

out (under regularity conditions which imply ours) for both correctly and incorrectly specified models (see White [12] and references listed there). Instead, we concentrate on obtaining an approximation to the effect of misspecification on the limits of the parameter estimates.

Let θ be the parameter vector being estimated, and denote the true value θ^0. We consider the linear restrictions $R\theta - r = 0$. These may be thought of as linear approximations to nonlinear restrictions. Estimates $\hat{\theta}_n$ are obtained by maximizing the objective function $L_n(\theta)$. It is natural to think of $L_n(\theta)$ as the normed loglikelihood function, but our results of course apply to methods based on any criterion function satisfying our assumptions, in particular to many least squares and M-estimators.

Assumption A1: The parameter space Θ is a compact subset of R^K, and θ^0 is interior to Θ.

Assumption A2: The $G \times K$ matrix of restrictions R has full row rank and the true value θ^0 satisfies $R\theta^0 - r = c^0$.

Assumption A3: Assumptions on the objective function are: (a) $L_n(\theta)$ has continuous second derivatives for $\theta \in \Theta$. (b) There exist nonstochastic functions $L(\theta)$, $D(\theta)$, and $H(\theta)$ such that

$$L_n(\theta) \to L(\theta),$$

$$\frac{\partial L_n(\theta)}{\partial \theta} \to D(\theta),$$

$$\frac{\partial^2 L_n(\theta)}{\partial \theta \partial \theta'} \to H(\theta),$$

almost surely uniformly for θ in Θ. (c) Identification: $L(\theta^0) > L(\theta)$ for all $\theta \in \Theta$, $\theta \neq \theta^0$.

These assumptions insure that $D(\theta)$ and $H(\theta)$ are the $K \times 1$ vector of first derivatives and $K \times K$ matrix of second derivatives of $L(\theta)$, respectively, and that $H(\theta)$ is negative definite in a neighborhood of θ^0. Define the constrained maximum likelihood estimator

$$\hat{\theta}_n = \arg \max_{\theta \in \Theta^*} \{ L_n(\theta) \}$$

where $\Theta^* = \{\theta \in \Theta \mid R\theta = r\}$. Kohn [9, Lemma 2] shows in a more general framework that $\hat{\theta}_n \to \theta^0$ almost surely when $c^0 = 0$, i.e., when the constraints are satisfied by the true parameter values, and that $\lim L_n(\hat{\theta}_n) < L(\theta^0)$ almost surely when the constraints are not satisfied by the true parameter value. White [12] gives conditions under which $\hat{\theta}_n$ will converge to a limit when the constraints are not satisfied by the true value; the primary additional assumption is that of the existence of a unique constrained maximum. We assume the following.

Assumption A4: $L(\theta)$ has a unique maximizer $\theta^*(c)$ in the set Θ^* for all c in a neighborhood of c^0.

Given Assumptions A1–A4, a standard argument, following Jennrich [7, Theorem 6] shows that $\hat{\theta}_n \to \theta^*(c^0)$ almost surely. White also provides an interpretation of the limiting parameter vector under misspecification as well as a test statistic for misspecification. We do not pursue this approach, and ask instead the following question: Suppose c^0 is small, so the constraints $R\theta = r$ almost hold at the true parameter value. Can we obtain a useful approximation to $\theta^*(c^0) - \theta^0$, i.e. an estimate of the effect of a hypothesized misspecification?

This question can be answered using conventional comparative statics techniques. Define the inconsistency function $B(c^0) = \theta^*(c^0) - \theta^0$. Note that $B(0) = 0$. We propose $\Delta = \partial B(0) c^0 / \partial c$ for small c^0 as a measure of local inconsistency. To estimate this quantity we consider the maximization problem

$$\max_{\theta \in \Theta} L(\theta) \qquad \text{subject to} \quad R\theta - r = c$$

and the necessary conditions

$$R\theta - r - c = 0,$$

$$D(\theta) + R'\lambda = 0,$$

where λ is the lagrange multiplier associated with the constraint $R\theta - r = c$. Solving the necessary conditions gives the functions $\theta(c)$ and $\lambda(c)$ (dependence on θ^0 is supressed for notational convenience). Our inconsistency function $B(c)$ can be written $\theta(0)$–$\theta(c)$. Standard manipulations (total differentiation of the necessary condition and substitution) lead to

$$\Delta = \frac{\partial B(0)}{\partial c} c^0$$

$$= -H^{-1} R' (R H^{-1} R')^{-1} c^0$$

for small c^0 where the function $H(\theta)$ is evaluated at $\theta(0)$. The consistent estimate $\hat{\theta}_n$ can be substituted for $\theta(0)$ to estimate Δ consistently, due to the uniform continuity and definiteness of the function H.

It is useful to consider an important special case in order to fix ideas. Partition θ into θ_1 and θ_2, $K_1 \times 1$ and $K_2 \times 1$, let $R = [0 \ I]$ where 0 is a $K_2 \times K_1$ matrix of zeros and I is the K_2 identity matrix, and set $r = 0$. The restriction considered is thus $\theta_2 = 0$. We wish to consider the effects of a potential violation of this restriction. We consider violations of the form $\theta_2 = c$, indicating the direction in which we suspect the restriction may fail. Partitioning

$$H = \begin{bmatrix} H_{11} & H_{12} \\ H_{21} & H_{22} \end{bmatrix}, \qquad H^{-1} = \begin{bmatrix} H^{11} & H^{12} \\ H^{21} & H^{22} \end{bmatrix}, \qquad \Delta = \begin{bmatrix} \Delta_1 \\ \Delta_2 \end{bmatrix},$$

it is clear that Δ_2, the inconsistency associated with imposing the constraint $\theta_2 = 0$, is simply $-c$. Applying our general formula yields

$$\Delta_1 = -H^{12}(H^{22})^{-1}c = H_{11}^{-1}H_{12}c.$$

This is a generalization of the Theil specification analysis result for the effect of leaving out regressors in the linear model. Clearly, our local result is global when H does not depend on unknown parameters, as in the linear model. We expect that the usefulness of the estimate Δ in analyzing misspecification will depend in general on the variability of H with respect to θ and on the size of c.

2. MISSPECIFICATION IN PROBIT-LOGIT AND OTHER DISCRETE MODELS[3]

A. *Left-out Variables*

A number of models for analysis of discrete data are based on an underlying latent variable

$$y_i = x_i\beta + \epsilon_i,$$

only the sign of which is observed. As an example y_i might be the difference between an individual's wage and his reservation wage, but the only data available may be whether he is employed or not. More generally y_i is the difference in utility between two choices, and observations are made on choices. The random variable ϵ_i has distribution function $F(\cdot)$ so that the probability that y_i is greater than zero is $F_i = 1 - F(-x_i\beta)$. Let $d_i = 0$ if $y_i < 0$, $d_i = 1$ if $y_i > 0$, $\bar{d}_i = 1 - d_i$ and $\bar{F}_i = 1 - F_i$. Then the normed loglikelihood function is

$$L_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} d_i \ln F_i + \frac{1}{n}\sum_{i=1}^{n} \bar{d}_i \ln \bar{F}_i$$

and

$$\frac{\partial^2 L_n}{\partial\beta\partial\beta'} = \frac{1}{n}\sum_{i=1}^{n} d_i \frac{\partial^2 \ln F_i}{\partial(x_i\beta)^2} x_i'x_i + \frac{1}{n}\sum_{i=1}^{n} \bar{d}_i \frac{\partial^2 \ln \bar{F}_i}{\partial(x_i\beta)^2} x_i'x_i = \frac{1}{n} X'\Lambda X$$

with Λ diagonal with ith element

$$\lambda_i = d_i \frac{\partial^2 \ln F_i}{\partial(x_i\beta)^2} + \bar{d}_i \frac{\partial^2 \ln \bar{F}_i}{\partial(x_i\beta)^2}$$

and X the x_i stacked up.

[3] A more detailed and general treatment of the results of this section and closely related results is given by Yatchew and Griliches [13].

Partition X into $[X_1 : X_2]$ and β correspondingly into

$$\begin{pmatrix}\beta_1\\ \beta_2\end{pmatrix}.$$

Suppose X_2 is omitted from the model, i.e., $\hat{\beta}_2$ is constrained to be zero. Then Δ_1, the local inconsistency in $\hat{\beta}_1$, is given by

$$\Delta_1 = H_{11}^{-1}H_{12}\beta_2^0$$

which is consistently estimated by

$$\hat{\Delta}_1 = (X_1'\hat{\Lambda}X_1)^{-1}X_1'\hat{\Lambda}X_2\beta_2^0$$

with β_2^0 the true value of β_2. These formulas are analogous to the corresponding result in linear regression; the extent of the inconsistency depends on the correlation between X_1 and X_2 and on the value of β_2^0. In many cases it will be possible to take expectations and to use Δ_1 rather than $\hat{\Delta}_1$; whether this will lead to a better measure of inconsistency is an open question.

B. *Heteroscedasticity*

Heteroscedasticity in logit and probit models, and discrete models generally, can lead to inconsistent parameter estimates if ignored. This result is in contrast with results in the linear model, in which ignoring heteroscedasticity does not introduce inconsistency. In this section we consider a simple form of heteroscedasticity in a probit model and show that its effects on the parameter estimate may not be too serious according to local misspecification analysis.

Suppose the latent variables y_{1i} and y_{2i} are distributed as

$$y_{1i} \sim N(x_{1i}\beta^*, \sigma_1^2),$$

$$y_{2i} \sim N(x_{2i}\beta^*, \sigma_2^2),$$

where $N(\mu, \sigma^2)$ denotes a normal distribution with mean μ and variance σ^2. Let

$$y = \begin{pmatrix}y_1\\ y_2\end{pmatrix} \quad \text{and} \quad X = \begin{pmatrix}X_1\\ X_2\end{pmatrix}.$$

As usual, suppose only the signs of the elements of y are observed and we wish to estimate coefficients of x by probit.

We wish to evaluate Δ, the inconsistency of the MLE, when heteroscedasticity is ignored. Here there are two possible normalizations: (1) $\beta_1 = \beta^*/\sigma_1$ and (2) $\beta_2 = \beta^*/\sigma_2$. Consider the first normalization, so regard the MLE as an estimator for β_1. We can write

$$\Pr(y_{1i} > 0) = \Phi(x_{1i}\beta_1),$$

$$\Pr(y_{2i} > 0) = \Phi\left(x_{2i}\beta_1 + x_{2i}\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\right)\beta^*\right).$$

The problem is now in the left-out regressor form: we restrict $\sigma_1 = \sigma_2$, thus leaving the second x_2 term out of the model. Here

$$X\beta = \begin{bmatrix} X_1 & 0 \\ X_2 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta^*\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\right) \end{bmatrix}$$

and we restrict

$$\beta^*\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\right)$$

to zero. Now,

$$X'\Lambda X = \begin{bmatrix} X_1'\Lambda X_1 + X_2'\Lambda X_2 & X_2'\Lambda X_2 \\ X_2'\Lambda X_2 & X_2'\Lambda X_2 \end{bmatrix}$$

and if the x_i are each assumed to be generated from the same process, and if there are equal numbers of observations from models 1 and 2, then it is a reasonable approximation to set $X_1'\Lambda X_1 = X_2'\Lambda X_2$. Then

$$\Delta_1 = \frac{1}{2}(X_2'\Lambda X_2)^{-1}(X_2'\Lambda X_2)\beta^*\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\right) = \frac{1}{2}\left(\frac{1}{\sigma_2} - \frac{1}{\sigma_1}\right)\beta^*.$$

Here Δ_1 is the local inconsistency of $\hat{\beta}$ as an estimator of β_1.

Under normalization 2 we are trying to estimate the parameter $\beta_2 = \beta^*/\sigma_2$. Using the symmetry of the problem, we note that the local inconsistency in $\hat{\beta}$ as an estimate of β_2 is

$$\Delta_2 = \frac{1}{2}\left(\frac{1}{\sigma_1} - \frac{1}{\sigma_2}\right)\beta^*.$$

Therefore when $\sigma_1 > \sigma_2$, $\beta_1 < \hat{\beta} < \beta_2$, and when $\sigma_2 > \sigma_1$ then $\beta_2 < \hat{\beta} < \beta_1$. Consequently, the effect of the heteroscedasticity does not alter our ability to estimate the parameters up to a scale factor. For purposes of prediction,

$$\Phi\left(x\left(\frac{1}{2\sigma_1} + \frac{1}{2\sigma_2}\beta^*\right)\right)$$

lies between $\Phi(x\beta_1)$ and $\Phi(x\beta_2)$.

Note that the analysis of the effect of heteroscedasticity did not require that the sample separation be known. Of course it would be useful but not necessary to know the sample separation in order to correct for the heteroscedasticity.

3. UNIVARIATE TIME SERIES SPECIFICATION ANALYSIS

To motivate the general result below, we consider the second order stationary autoregressive process

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + u_t.$$

When we fit a first order autoregression to data generated by this process by maximum likelihood (or any other asymptotically equivalent method) the resulting constrained estimate of ϕ_1, $\hat{\phi}_1^c$, will be chosen so as to minimize $\sum(x_t - \phi_1 x_{t-1})^2$. Now the almost sure limit of $\sum x_t x_{t-1}/\sum x_{t-1}^2 = \hat{\phi}_1^c$ is $\rho(1)$, the first autocorrelation coefficient of the x_t process, which standard calculations (Box and Jenkins [2, 3.2.27, p. 60]) show to be $\phi_1/(1 - \phi_2)$. This is a *global* inconsistency result, unlike the results of previous sections and some results to follow. It is the nature of these time series models that the effects of misspecification may be characterized as the result of variational problems; the misspecified model and inconsistencies depend only upon the theoretical parameters of the true model (and the (p,q) choice for misspecified model). Of course closed form expressions for these a.s. limits will become unwieldly, and attention may shift to the analysis of another essentially global quantity, $\sigma_{u_0}^2$, the forecasting error variance attained when the misspecified model is used to forecast one period ahead.[4]

It is of interest generally to compare our *local* inconsistency measure, for any case with the global inconsistency measure, whenever the latter is available. In the notation of Section 1, $\theta = (\phi_1, \phi_2, \sigma^2)$ and

$$L_n(\theta) = -\frac{1}{2}\log 2\pi - \frac{1}{2}\log\sigma^2 - \frac{1}{2n\sigma^2}\sum_{t=1}^{n}(x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2})^2$$

conditional on x_0 and x_{-1}. Thus

$$D_n(\theta) = \begin{bmatrix} \frac{\partial L_n}{\partial \phi_1} \\ \frac{\partial L_n}{\partial \phi_2} \\ \frac{\partial L_n}{\partial \sigma^2} \end{bmatrix} = \frac{1}{n}\begin{bmatrix} \frac{1}{\sigma^2}\sum_{t=1}^{n} u_t x_{t-1} \\ \frac{1}{\sigma^2}\sum_{t=1}^{n} u_t x_{t-2} \\ -\frac{n}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum u_t^2 \end{bmatrix}$$

[4] In Skoog [10], for various misspecifications, the quantity $\sigma_{u_0}^2$ is studied to determine when the effects of model misspecification will be benign if the end is forecasting accuracy and not parameter estimation.

and

$$H_n(\theta) = \frac{\partial^2 l_n}{\partial\theta\partial\theta'}$$

$$= \frac{-1}{n\sigma^2}\begin{bmatrix} \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} & \frac{1}{\sigma^2}\sum u_t x_{t-1} \\ & \sum x_{t-2}^2 & \frac{1}{\sigma^2}\sum u_t x_{t-2} \\ \text{symmetric} & & \frac{-n}{2\sigma^2} + \frac{\sum u_t^2}{(\sigma^2)^2} \end{bmatrix}.$$

Using $D(\theta^0) = 0$ and expressions for the autocovariance function of an AR(2) process,

$$H(\phi_1, \phi_2, \sigma^2)$$

$$= -\begin{bmatrix} \frac{(1-\phi_2)}{(1+\phi_2)}\left(\frac{1}{(1-\phi_2)^2 - \phi_1^2}\right) & \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]} & 0 \\ \frac{\phi_1}{(1+\phi_2)[(1-\phi_2)^2 - \phi_1^2]} & \frac{(1-\phi_2)}{1+\phi_2}\left(\frac{1}{(1-\phi_2)^2 - \phi_1^2}\right) & 0 \\ 0 & 0 & \frac{1}{2(\sigma^2)^2} \end{bmatrix}.$$

Because of the block of zeros there is no need to concentrate the likelihood with respect to σ^2, and $H_{11}^{-1}H_{12}\phi_2$ becomes $\phi_1\phi_2/(1-\phi_2)$ with limiting slope ϕ_1 as ϕ_2 goes to zero. The global inconsistency was

$$\frac{\phi_1}{1-\phi_2} - \phi_1 = \frac{\phi_1}{1-\phi_2}\phi_2,$$

differing from the local result. Thus the global inconsistency function

$$B(\phi_2) = \frac{\phi_1\phi_2}{1-\phi_2}, \quad \text{and} \quad \frac{\partial B(0)}{\partial\phi_2} = \phi_1$$

is the limiting slope. Note that our local measure correctly signs the inconsistency throughout the parameter space consistent with stationarity as sign $(\phi_1\phi_2/(1-\phi_2))$ agrees with the sign $(\phi_1\phi_2)$ since $1-\phi_2 > 0$. General theory only guaranteed agreement in a neighborhood of $\phi_2 = 0$.

We may now exploit similarity between MA and AR H matrices. We see immediately that, omitting θ_2 in estimating the MA(2) process $z_t = u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2}$ will result in the local slope θ_1 since the H matrix of (θ_1, θ_2) is of the same form as that for (ϕ_1, ϕ_2) (Box–Jenkins [2, p. 283]). Unlike the AR case, $\lim \hat{\theta}_1^c$, while characterizable as the solution of a variational problem, does not appear amenable to a closed form solution.

To complete the analyses of the parsimonious ARMA models, we record the results of misspecifying an ARMA (1, 1) as an AR(1) and an MA(1) respectively. Using the expression on page 284 in Box–Jenkins, we find the H matrix of (ϕ, θ) and obtain the local inconsistencies as $-(1-\phi^2)\theta$ and $-(1-\theta^2)\phi$. In fact, using the result that

$$\rho(1) = \frac{(1-\phi\theta)(\phi-\theta)}{1+\theta^2-2\phi\theta},$$

the global inconsistency in the former case is $\rho(1) - \phi$, or $-(1-\phi^2)\theta/((1+\theta^2) - 2\phi\theta)$. This will always agree in sign with the local inconsistency, since stationarity and the innovations representation of the ARMA process ($|\phi| < 1$, $|\theta| < 1$) imply $1 + \theta^2 - 2\phi\theta > 0$, and $1 - \phi^2 > 0$.

To generalize these results we need to establish some notation. For a covariance stationary process x_t with mean 0, $\rho_k = (Ex_t x_{t-k})/Ex_t^2$. $P_{0,p-1}$ will denote the $p \times p$ Toeplitz matrix whose elements on the k sub- or super-diagonal are all equal to ρ_k. The subscripts indicate that the first row is $(\rho_0, \rho_1, \ldots, \rho_{p-1})$. Next we define the vectors $\rho'_{1,p} = (\rho_1, \rho_2, \ldots, \rho_p)$, $\underline{\phi}'_{1,p} = (\phi_1, \phi_2, \ldots, \phi_p)$, $\underline{\phi}'_{p_0+1,p} = (\phi_{p_0+1}, \ldots, \phi_p)$, $\phi_{1,p_0} = (\phi_1^0, \ldots, \phi_{p_0}^0)$ and the lag operators $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ and $\Phi_0(L) = 1 - \phi_1^0 L - \cdots - \phi_{p_0}^0 L^{p_0}$. When $\Phi(L)x_t = u_t$, u_t i.i.d. with mean 0, variance σ^2, and the vector $\phi'_{1,p}$ is characterized by the Yule–Walker equations:

$$P_{0,p-1}\phi_{1,p} = \underline{\rho}_{1,p}$$

which, in more detail, appears as

$$\begin{bmatrix} 1 & \rho_1 \cdots & \rho_{p_0-1} & \rho_{p_0} & \cdots & \rho_{p-1} \\ \rho_1 & 1 & & & \ddots & \\ & \ddots & & & & \\ \rho_{p_0-1} & & 1 & \rho_1 & & \rho_{p-p_0} \\ \rho_{p_0} & \cdots & \rho_1 & 1 & \cdots & \rho_{p-p_0-1} \\ & \ddots & & & \ddots & \\ \rho_{p-1} & & \rho_{p-p_0} & \rho_{p-p_0-1} & & 1 \end{bmatrix} \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_{p_0} \\ \phi_{p_0+1} \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \vdots \\ \rho_{p_0} \\ \rho_{p_0+1} \\ \vdots \\ \rho_p \end{bmatrix}$$

or, more compactly, as

$$\begin{bmatrix} P_{0,p_0-1} & P_{p_0,p-1} \\ P'_{p_0,p-1} & P_{0,p-p_0-1} \end{bmatrix} \begin{bmatrix} \underline{\phi}_{1,p_0} \\ \underline{\phi}_{p_0+1,p} \end{bmatrix} = \begin{bmatrix} \underline{\rho}_{1,p_0} \\ \underline{\rho}_{p_0+1,p} \end{bmatrix}.$$

When the covariance stationary x_t does not follow a p_0th order autoregression, but such an AR(p_0) is fit to the data, the Yule Walker equations tell us the almost sure limit of $\hat{\phi}_{1,p_0}$ as fit by ML will be $P_{0,p-1}^{-1}\underline{\rho}_{1,p}$. This is because ML and OLS are asymptotically equivalent, and the OLS estimator will be $R_{0,p-1}^{-1}r_{1,p}$, where a matrix or vector with sample autocorrelations replacing population autocorrelations is indicated by replacing P by R. Since the $\hat{r}(k)$ are strongly consistent for ρ_k (Hannan and Heyde [5]), the result follows.

We may now give a general result concerning the global and local inconsistency in fitting an AR(p) by an AR(p_0), $p_0 < p$. The fact that the OLS estimators are not linear in the data is enough to ensure that the global and local inconsistency measures differ. Note that, following time-series conventions, p_0 is the misspecified order and p the true order.

THEOREM: *When the misspecified model* $\Phi_0(L)x_t = u_t^0$ *is fit to the* AR(p) *process* $\Phi(L)x_t = u_t$ *the global bias in the coefficients of* $\underline{\phi}_{1,p_0}$ *is* $P_{0,p_0-1}^{-1}P_{p_0,p-1}\underline{\phi}_{p_0+1,p}$ *while the local bias is the same quantity with the first two terms evaluated at* $(\phi_1, \ldots, \phi_{p_0}, 0, \ldots, 0)$.

PROOF: Writing out the first p_0 equations from the compact representation above yields $P_{0,p_0-1}\phi_{1,p_0} + P_{p_0,p-1}\underline{\phi}_{p_0+1,p} = \rho_{1,p_0}$. The previous discussion also showed that the equation $P_{0,p_0-1}\phi_{1,p_0}^0 = \underline{\rho}_{1,p_0}$ characterizes the almost sure limit of the estimator $\hat{\underline{\phi}}_{1,p_0}^0$ which results when OLS is used in fitting the AR(p_0), thereby omitting $x_{t-p_0-1}, \ldots, x_{t-p}$. Subtracting the second equation from the first shows the global inconsistency equal to $\phi_{1,p_0}^0 - \phi_{1,p_0} = P_{0,p_0-1}^{-1}P_{p_0,p-1}\phi_{p_0+1,p}$. To obtain the local inconsistency, concentrating σ^2 from the loglikelihood results in our taking second derivatives of $\sum(x_t - \phi_1 x_{t-1} - \cdots - \phi_p x_{t-p})^2$ with respect to $(\phi_1, \ldots, \phi_p)$ to form $H_n(\theta)$. Thus $H_n = R_{0,p-1}$. When $R_{0,p}$ is partitioned appropriately and the general formula $H_{11}^{-1}H_{12}\theta_2^0$ used, the previous formula $P_{0,p_0-1}^{-1}P_{p_0,p-1}\underline{\phi}_{p_0+1,p}$ results, since $(H_{11})^{-1} \to P_{0,p_0-1}^{-1}$ and $H_{12} \to P_{p_0,p-1}$ almost surely. *Q.E.D.*

It is useful to compare this result with the usual fixed regressor case in which the variates X_2 are omitted in the regression $y = X_1\beta_1 + X_2\beta_2 + u$. The bias in β_1 is $(X_1'X_1)^{-1}X_1'X_2\beta_2$; to analyze the bias, information is most likely available on β_2, but problematic regarding $(X_1'X_1)^{-1}X_1'X_2$. In the time series case, $\phi_{p_0+1,p}$, the analog of β_2 is most likely to be difficult to ascertain, whereas $P_{0,p_0-1}^{-1}P_{p_0,p-1}$ is an easily computable function of readily available information, the sample autocorrelations.

As regards the general problem of misspecifying an ARMA(p,q) by an ARMA (p_0,q_0), we could formally apply our general formula $H_{11}^{-1}H_{12}\theta_2^0$ to appropriately partitioned submatrices of the information matrix of the correct model. While the formulae on p. 283 of Box–Jenkins relate such an information matrix to an associated AR($p+q$) process, it seems we are unlikely to be successful in deducing qualitative information on any but the parsimonious models analyzed earlier.

4. MISSPECIFICATION IN NONLINEAR REGRESSIONS

The analysis of misspecification in nonlinear regression models is straightforward and has wide applicability.[5] Write the model

$$Y = g(x_1, \ldots, x_k, \theta_1, \theta_2) + \epsilon$$

with Y, $g(\cdot)$, and $x_1, \ldots, x_k$ all $n \times 1$ and suppose ϵ is normally distributed with mean zero and variance $\sigma^2 I_n$.

Here

$$H_n(\theta) = -\frac{1}{n}\sigma^{-2}G'G$$

with

$$G = \frac{\partial g}{\partial \theta}; \qquad \theta = \begin{pmatrix} \theta_1 \\ \theta_2 \end{pmatrix}.$$

The local inconsistency in $\hat{\theta}_1$ from setting $\hat{\theta}_2$ equal to zero is

$$\Delta_1 = H_{11}^{-1}H_{12}\theta_2^0.$$

As an example, consider the model

$$y_t = x_t\beta + \alpha y_{t-1} + u_t,$$
$$u_t = \rho u_{t-1} + \epsilon_t,$$

with the ϵ_t normally distributed with mean zero and variance σ^2 and independent. Transforming gives

$$y_t = x_t\beta - \rho x_{t-1}\beta + (\alpha+\rho)y_{t-1} - \alpha\rho y_{t-2} + \epsilon_t.$$

We will find the almost sure limit of

$$\hat{\theta}_1 = \begin{pmatrix} \hat{\alpha} \\ \hat{\beta} \end{pmatrix}$$

when $\hat{\theta}_2 = \rho$ is constrained to zero (with the "true" ρ equal to ρ^0). In this case, it suffices to calculate H under the constraint. We have (at $\rho = 0$) for the tth row

[5] Regularity conditions for nonlinear regressions are given by Jennrich [7].

of G,

$$G_t = \left[y_{t-1} \vdots x_t \vdots (y_{t-1} - x_{t-1}\beta - \alpha y_{t-2}) \right].$$

Drop the first two observations and let

$$X = \begin{bmatrix} x_3 \\ \vdots \\ x_n \end{bmatrix} \quad \text{and} \quad Y_{-1} = \begin{bmatrix} y_2 \\ \vdots \\ y_{n-1} \end{bmatrix}.$$

Then, almost surely

$$-\sigma^2 H_{11} = \lim \frac{1}{n} \begin{bmatrix} Y'_{-1}Y_{-1} & Y'_{-1}X \\ X'Y_{-1} & X'X \end{bmatrix},$$

$$-\sigma^2 H_{12} = \lim \frac{1}{n} \begin{bmatrix} \sum_{t=3}^{n} y_{t-1}(y_{t-1} - x_{t-1}\beta - \alpha y_{t-2}) \\ \sum_{t=3}^{n} x_t(y_{t-1} - x_{t-1}\beta - \alpha y_{t-2}) \end{bmatrix}.$$

Now

$$H_{12}(\alpha^0, \beta^0, 0) = \begin{pmatrix} \sigma^2 \\ 0 \end{pmatrix}$$

so we need only calculate the first column of H_{11}^{-1},

$$\frac{1}{\sigma^2} \lim \frac{1}{n} \begin{Bmatrix} \left(Y'_{-1}Y_{-1} - Y'_{-1}X(X'X)^{-1}X'Y_{-1}\right)^{-1} \\ -(X'X)^{-1}X'Y_{-1}\left(Y'_{-1}Y_{-1} - Y'_{-1}X(X'X)^{-1}X'Y_{-1}\right)^{-1} \end{Bmatrix}$$

and we estimated the local inconsistency by

$$\Delta_1 = \begin{pmatrix} S^{-2} \\ -(X'X)^{-1}X'Y_{-1}S^{-2} \end{pmatrix} \rho^0$$

where S^2 is the mean squared error from the regression of lagged (once) Y on X. The local inconsistency in $\hat{\alpha}$ thus has the sign of ρ^0. The local inconsistencies of the elements of β have minus the sign of the corresponding coefficients of the regression of lagged y on X (times ρ^0).

5. CONCLUSION

Local specification error analysis provides a method of assessing the effect on parameter estimates of small departures from the assumptions of the model. The method complements the tools of formal specification testing (Davidson–

MacKinnon [4], Hausman [6], White [12]) since imposing constraints does reduce variance (Aitchison and Silvey [1]). Hence one may wish to assess the tradeoff between inconsistency and variance reduction. We have not studied distribution theory here, but we will note that with the following additional assumption, the ($\sqrt{n}$-normed) restricted MLE minus θ^0 will be asymptotically normally distributed with zero mean and variance $-(H^{-1} - H^{-1}R'(RH^{-1}R')RH^{-1})$ when θ^0 satisfies the restrictions imposed.

ASSUMPTION A5: $\sqrt{n}\, H(\theta^0)^{-1/2} D(\theta^0) \to N(0, I)$ in distribution.

For small c, it may be appropriate to shift the mean to Δ and to retain the variance formula. To make this technique precise, c must depend on the number of observations.

Cornell University
and
University of Chicago

Manuscript received July, 1980; final revision received November, 1983.

REFERENCES

[1] AITCHISON, S. AND S. O. SILVEY: "Maximum-Likelihood Estimation of Parameters Subject to Restraints," *Annals of Mathematical Statistics*, 29(1958), 812–828.
[2] BOX, G. E. P., AND G. M. JENKINS: *Time Series Analysis: Forecasting and Control*. San Francisco: Holden Day, 1976.
[3] CRAMER, H.: *Mathematical Methods of Statistics*. Princeton: Princeton University Press, 1946.
[4] DAVIDSON, R., AND J. G. MACKINNON: "Several Tests for Model Specification in the Presence of Alternative Hypotheses," *Econometrica*, 49(1981), 781–795.
[5] HANNAN, E. J., AND C. C. HEYDE: "On Limit Theorems for Quadratic Functions of Discrete Time Series," *Annals of Mathematical Statistics*, 43(1972), 2058–2066.
[6] HAUSMAN, J. A.: "Specification Tests in Econometrics," *Econometrica*, 46(1978), 1251–1272.
[7] JENNRICH, R. I.: "Asymptotic Properties of Non-Linear Least Squares Estimators," *Annals of Mathematical Statistics*, 40(1969), 633–643.
[8] KOHN, R.: "Asymptotic Estimation and Hypothesis Testing Results for Vector Linear Time Series Models," *Econometrica*, 47(1979), 1005–1030.
[9] ——: "Consistent Estimation of Minimal Model Dimension: I," University of Chicago, 1981.
[10] SKOOG, G.: "Specification Error Analysis for ARMA Models," University of Chicago, 1980.
[11] THEIL, H.: "Specification Errors and the Estimation of Economic Relationships," *Review of the International Statistical Institute*, 25(1957), 41–51.
[12] WHITE, H.: "Maximum Likelihood Estimation of Misspecified Models," *Econometrica*, 50(1982), 1–25.
[13] YATCHEW, A., AND Z. GRILICHES: "Specification Error in Probit and Logit Models," HIER Discussion Paper 717, July, 1979.