

NOTE ON MAXIMUM-LIKELIHOOD ESTIMATION OF
MISSPECIFIED MODELS

Gregory C. Chow

Econometric Research Program
Research Memorandum No. 300

June 1982

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

NOTE ON MAXIMUM-LIKELIHOOD ESTIMATION OF MISSPECIFIED MODELSBy Gregory C. Chow¹

In a recent paper, White [5] studies the asymptotic distribution of maximum-likelihood estimators when the model is misspecified. This note points out that the covariance matrix of the asymptotic distribution stated therein is incorrect.

Let $f(\cdot; \theta)$ be the specified density function for n independent (vector) observations $(y_1, \dots, y_n) = Y$. The log-likelihood function is

$$(1) \quad \log L(Y; \theta) = \sum_{i=1}^n \log f(y_i; \theta) .$$

Under the assumption of differentiability of f , the maximum-likelihood (ML) estimator $\hat{\theta}$ is obtained by solving

$$(2) \quad n^{-1} \frac{\partial \log L(Y; \hat{\theta})}{\partial \theta} = 0 .$$

Let $g(\cdot)$ be the true density function of y_i ($i = 1, \dots, n$). For simplicity, assume

$$(3) \quad z(\theta) \equiv E[n^{-1} \log L(Y; \theta)] = \int \dots \int [n^{-1} \sum_i \log f(y_i; \theta)] g(y_1) \dots g(y_n) dy_1 \dots dy_n$$

achieves a unique maximum at $\theta = \theta_*$. The vector θ_* can be regarded as the "true" parameter vector of the misspecified model $f(\cdot; \theta)$. Under appropriate regularity conditions, $\hat{\theta}$ is a consistent estimator of θ_* , or more precisely the sequence $\{\hat{\theta}_n\}$ is a consistent estimator of θ_* . (We need not distinguish between strong consistency and weak consistency for the purpose of pointing out White's error.)

As usual, one can derive the asymptotic distribution of $\hat{\theta}$ by expanding (2) about θ_* ,

$$(4) \quad n^{-1} \frac{\partial \log L(Y; \theta_*)}{\partial \theta} + \left[n^{-1} \frac{\partial^2 \log L(Y; \theta_*)}{\partial \theta \partial \theta'} + o(1) \right] (\hat{\theta} - \theta_*) = 0 .$$

The solution of (4) is

$$(5) \quad n^{1/2} (\hat{\theta} - \theta_*) = \left[-n^{-1} \frac{\partial^2 \log L(Y; \theta_*)}{\partial \theta \partial \theta'} + o(1) \right]^{-1} n^{-1/2} \frac{\partial \log L(Y; \theta_*)}{\partial \theta} .$$

Since $n^{-1/2} \partial \log L(Y; \theta_*) / \partial \theta = n^{-1/2} \sum_{i=1}^n \partial \log f(y_i; \theta_*) / \partial \theta$ is a sum of independent vectors, it is asymptotically normal. Its expectation is zero, or

$$(6) \quad E \left[\frac{\partial \log L(Y; \theta_*)}{\partial \theta} \right] = 0$$

which results from differentiating (3) to obtain the maximum at θ_* . Let

$$\text{plim}_{n \rightarrow \infty} \left[-n^{-1} \frac{\partial^2 \log L(Y; \theta_*)}{\partial \theta \partial \theta'} \right] = A(\theta_*) .$$

Then by (5), (6), and the central limit theorem, the asymptotic distribution of $n^{-1/2} (\hat{\theta} - \theta_*)$ is normal with mean zero and covariance matrix

$$(7) \quad A(\theta_*)^{-1} \cdot n^{-1} \text{Cov} \left[\frac{\partial \log L(Y; \theta_*)}{\partial \theta} \right] \cdot A(\theta_*)^{-1} .$$

The formula (7) has been used by Chow [1], [2] and [3], and was implied by Silvey [4].

Theorem 3.2 of White [5] states that the covariance matrix of

$$(8) \quad \frac{\partial \log L(Y; \theta_*)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \log f(y_i; \theta_*)}{\partial \theta}$$

is

$$(9) \quad \sum_{i=1}^n E \left[\frac{\partial \log f(y_i; \theta_*)}{\partial \theta} \cdot \frac{\partial \log f(y_i; \theta_*)}{\partial \theta'} \right]$$

The mistake lies in equating the covariance matrix of $\partial \log f(y_i; \theta_*) / \partial \theta$ with

$$E \left[\frac{\partial \log f(y_i; \theta_*)}{\partial \theta} \cdot \frac{\partial \log f(y_i; \theta_*)}{\partial \theta'} \right]$$

because

$$(10) \quad E \left[\frac{\partial \log f(y_i; \theta_*)}{\partial \theta} \right] \neq 0.$$

The fact that (6) or the expectation of (8) is zero does not imply that each individual term $\partial \log f(y_i; \theta_*) / \partial \theta$ has zero expectation. This point was stressed in Chow [2, pp. 205-206], where the covariance matrix (7) was exhibited and its estimation was discussed.²

To illustrate this point, let the true model be a normal linear regression model with two sets of explanatory variables

$$(11) \quad y = X_1 \beta_{01} + X_2 \beta_{02} + \varepsilon \quad (E\varepsilon\varepsilon' = I\sigma_0^2)$$

and let the misspecified model result from omitting the second set

$$(12) \quad y = X_1 \beta_{*1} + \varepsilon^*.$$

The log-likelihood of the latter model is

$$(13) \quad -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2} (y - X_1 \beta_1)' (y - X_1 \beta_1) / \sigma^2 .$$

The expectation of (13) evaluated by the true model (11) is

$$(14) \quad -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2} E[X_1(\beta_{01} - \beta_1) + X_2 \beta_{02} + \epsilon]' [X_1(\beta_{01} - \beta_1) + X_2 \beta_{02} + \epsilon] / \sigma^2$$

$$= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} [X_1(\beta_{01} - \beta_1) + X_2 \beta_{02}]' [X_1(\beta_{01} - \beta_1) + X_2 \beta_{02}]$$

$$- \frac{n\sigma_0^2}{2\sigma^2} .$$

Maximizing (14) with respect to β_1 and σ^2 yields

$$(15a) \quad \beta_{*1} = \beta_{01} + (X_1' X_1)^{-1} X_1' X_2 \beta_{02}$$

$$(15b) \quad \sigma_*^2 = \sigma_0^2 + n^{-1} \beta_{02}' X_2' [I - X_1 (X_1' X_1)^{-1} X_1'] X_2 \beta_{02} .$$

For this model, with x_{i1}' denoting the i -th row of X_1 ,

$$\log f(y_i; \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 - \frac{1}{2} (y_i - x_{i1}' \beta_1)^2 / \sigma^2$$

and the derivative of $\log f(y_i; \theta)$ with respect to β_1 evaluated at β_{*1} and σ_*^2 is

$$(16) \quad x_{i1}' (y_i - x_{i1}' \beta_{*1}) / \sigma_*^2 .$$

The expectation of the column vector (16) is not zero because $(y_i - x_{i1}' \beta_{*1})$ does not necessarily have zero expectation for each i . Note that the sum of the terms (16) over i has zero expectation, or the derivative of (13) with respect to β_1 evaluated at β_{*1} and σ_*^2 has zero expectation since

$$\begin{aligned}
 (17) \quad E \sum_i x_{i1} (y_i - x_{i1}' \beta_{*1}) &= E(x_{11} \dots x_{n1}) (y - X_1 \beta_{*1}) = EX_1' (y - X_1 \beta_{*1}) \\
 &= X_1' [X_1 \beta_{01} + X_2 \beta_{02} - X_1 \beta_{01} - X_1 (X_1' X_1)^{-1} X_1' X_2 \beta_{02}] \\
 &= X_1' [I - X_1 (X_1' X_1)^{-1} X_1'] X_2 \beta_{02} = 0.
 \end{aligned}$$

For this example with $\theta' = (\beta_1' \ \sigma^2)$, Chow [1, p. 26] has evaluated

$$\text{Cov} \left[\frac{\partial \log L(Y; \theta_*)}{\partial \beta_1} \right] = E \left[\frac{\partial \log L(Y; \theta_*)}{\partial \beta_1} \cdot \frac{\partial \log L(Y; \theta_*)}{\partial \beta_1'} \right] = \frac{\sigma_0^2}{\sigma_*^4} X_1' X_1$$

which the reader can easily check. Using White's formula and (16), one would obtain

$$\text{Cov} \left[\frac{\partial \log L(Y; \theta_*)}{\partial \beta_1} \right] = \sum_{i=1}^n E \left[\frac{\partial \log f(y_i; \theta_*)}{\partial \beta_1} \cdot \frac{\partial \log f(y_i; \theta_*)}{\partial \beta_1'} \right] = \sigma_*^{-4} \sum_{i=1}^n x_{i1}' x_{i1} E(y_i - x_{i1}' \beta_{*1})^2$$

which is incorrect since $E(y_i - x_{i1}' \beta_{*1})^2$ ($i = 1, \dots, n$) do not all equal σ_0^2 , and not even on the average.

The covariance matrix of the ML estimator for θ_* of a misspecified model depends on the true model. It would be nice if the covariance matrix could be consistently estimated without specifying the true model, as White claims. If White were correct, (9) could be replaced by its sample analogue, with E omitted and θ_* replaced by $\hat{\theta}$, to construct a consistent estimate of (7), with $A(\theta_*)$ also replaced by its sample analogue

$$- n^{-1} \frac{\log L(Y; \hat{\theta})}{\partial \theta \partial \theta'}$$

which is consistent. Unfortunately, the error of using (9) invalidates White's

method of estimating the covariance matrix of the ML estimator for misspecified models as well as the other claims of his paper [5] which depend on this covariance matrix. The latter include his proposed modifications of the Lagrangian multiplier and Wald test statistics [5, Theorems 3.4 and 3.5].

There may be situations in which the equality sign instead of the inequality sign holds in equation (10) for a misspecified model, so that White's estimate of the covariance matrix is consistent. However, as our regression example illustrates, most misspecified models in econometrics do not fall into these situations. The ML estimation of a misspecified simultaneous-equations model has been treated in Chow [2] and [3] where a correctly specified model is postulated.

REFERENCES

- [1] Chow, G. C.: "A Comparison of the Information and Posterior Probability Criteria for Model Selection," Journal of Econometrics, 16 (1981), 21-33.
- [2] Chow, G. C.: "Selection of Econometric Models by the Information Criterion," in E. G. Charatsis, ed., Proceedings of the Econometric Society European Meeting 1979, 199-214. Amsterdam: North-Holland Publishing Company, 1981.
- [3] Chow, G. C.: "Evaluation of Econometric Models by Decomposition and Aggregation," in J. Kmenta and J. B. Ramsey, eds., Large-Scale Macro-Econometric Models, 423-444. Amsterdam: North-Holland Publishing Company, 1981.
- [4] Silvey, S. D.: "The Lagrangian Multiplier Test," Annals of Mathematical Statistics, 30 (1959), 389-407.
- [5] White, H.: "Maximum Likelihood Estimation of Misspecified Models," Econometrica, 50 (1982), 1-25.

FOOTNOTES

¹I would like to thank Adrian Pagan for helpful comments and acknowledge financial support from the National Science Foundation through Grant No. SES80-12582.

²I quote from the above reference:

"However the estimation of $\text{Cov}[\partial \log L^* / \partial \theta_1]$ is not so simple.

One might propose the estimate

$$\sum_{i=1}^n \left[\frac{\partial \log f(y_i | \hat{\theta}^*)}{\partial \theta_1} \cdot \frac{\log f(y_i | \hat{\theta}^*)}{\partial \theta_1'} \right] \quad (2.12)$$

in view of the relation

$$\text{Cov} \left[\frac{\partial \log L(Y, \theta^*)}{\partial \theta_1} \right] = \sum_{i=1}^n \text{Cov} \left[\frac{\partial \log f(y_i | \theta^*)}{\partial \theta_1} \right]. \quad (2.13)$$

The estimate (2.12) is inappropriate because, while the expectation of $\partial \log L(Y, \theta^*) / \partial \theta_1$ is zero, the expectations of its components $\partial \log f(y_i | \theta^*) / \partial \theta_1$ ($i=1, \dots, n$) are not zero when the correct model is $f(\cdot | \theta^0)$. Eq. (3.11) of section 3, for example, shows that $\partial \log f(y_i | \theta^*) / \partial \theta_1$ is not zero where θ_1 consists of coefficients of the exogenous variables in a simultaneous-equation model, to be denoted by the elements of the matrix B in section 3."

NOTE ON MAXIMUM-LIKELIHOOD ESTIMATION OF
MISSPECIFIED MODELS

Gregory C. Chow

Econometric Research Program
Research Memorandum No. 300

June 1982

Econometric Research Program
PRINCETON UNIVERSITY
207 Dickinson Hall
Princeton, New Jersey

NOTE ON MAXIMUM-LIKELIHOOD ESTIMATION OF MISSPECIFIED MODELSBy Gregory C. Chow¹

In a recent paper, White [5] studies the asymptotic distribution of maximum-likelihood estimators when the model is misspecified. This note points out that the covariance matrix of the asymptotic distribution stated therein is incorrect.

Let $f(\cdot; \theta)$ be the specified density function for n independent (vector) observations $(y_1, \dots, y_n) = Y$. The log-likelihood function is

$$(1) \quad \log L(Y; \theta) = \sum_{i=1}^n \log f(y_i; \theta) .$$

Under the assumption of differentiability of f , the maximum-likelihood (ML) estimator $\hat{\theta}$ is obtained by solving

$$(2) \quad n^{-1} \frac{\partial \log L(Y; \hat{\theta})}{\partial \theta} = 0 .$$

Let $g(\cdot)$ be the true density function of y_i ($i = 1, \dots, n$). For simplicity, assume

$$(3) \quad z(\theta) \equiv E[n^{-1} \log L(Y; \theta)] = \int \dots \int [n^{-1} \sum_i \log f(y_i; \theta)] g(y_1) \dots g(y_n) dy_1 \dots dy_n$$

achieves a unique maximum at $\theta = \theta_*$. The vector θ_* can be regarded as the "true" parameter vector of the misspecified model $f(\cdot; \theta)$. Under appropriate regularity conditions, $\hat{\theta}$ is a consistent estimator of θ_* , or more precisely the sequence $\{\hat{\theta}_n\}$ is a consistent estimator of θ_* . (We need not distinguish between strong consistency and weak consistency for the purpose of pointing out White's error.)

As usual, one can derive the asymptotic distribution of $\hat{\theta}$ by expanding (2) about θ_* ,

$$(4) \quad n^{-1} \frac{\partial \log L(Y; \theta_*)}{\partial \theta} + \left[n^{-1} \frac{\partial^2 \log L(Y; \theta_*)}{\partial \theta \partial \theta'} + o(1) \right] (\hat{\theta} - \theta_*) = 0 .$$

The solution of (4) is

$$(5) \quad n^{1/2} (\hat{\theta} - \theta_*) = \left[-n^{-1} \frac{\partial^2 \log L(Y; \theta_*)}{\partial \theta \partial \theta'} + o(1) \right]^{-1} n^{-1/2} \frac{\partial \log L(Y; \theta_*)}{\partial \theta} .$$

Since $n^{-1/2} \frac{\partial \log L(Y; \theta_*)}{\partial \theta} = n^{-1/2} \sum_{i=1}^n \frac{\partial \log f(y_i; \theta_*)}{\partial \theta}$ is a sum of independent vectors, it is asymptotically normal. Its expectation is zero, or

$$(6) \quad E \left[\frac{\partial \log L(Y; \theta_*)}{\partial \theta} \right] = 0$$

which results from differentiating (3) to obtain the maximum at θ_* . Let

$$\text{plim}_{n \rightarrow \infty} \left[-n^{-1} \frac{\partial^2 \log L(Y; \theta_*)}{\partial \theta \partial \theta'} \right] = A(\theta_*) .$$

Then by (5), (6), and the central limit theorem, the asymptotic distribution of $n^{-1/2} (\hat{\theta} - \theta_*)$ is normal with mean zero and covariance matrix

$$(7) \quad A(\theta_*)^{-1} \cdot n^{-1} \text{Cov} \left[\frac{\partial \log L(Y; \theta_*)}{\partial \theta} \right] \cdot A(\theta_*)^{-1} .$$

The formula (7) has been used by Chow [1], [2] and [3], and was implied by Silvey [4].

Theorem 3.2 of White [5] states that the covariance matrix of

$$(8) \quad \frac{\partial \log L(Y; \theta_*)}{\partial \theta} = \sum_{i=1}^n \frac{\partial \log f(y_i; \theta_*)}{\partial \theta}$$

is

$$(9) \quad \sum_{i=1}^n E \left[\frac{\partial \log f(y_i; \theta_*)}{\partial \theta} \cdot \frac{\partial \log f(y_i; \theta_*)}{\partial \theta'} \right]$$

The mistake lies in equating the covariance matrix of $\partial \log f(y_i; \theta_*) / \partial \theta$ with

$$E \left[\frac{\partial \log f(y_i; \theta_*)}{\partial \theta} \cdot \frac{\partial \log f(y_i; \theta_*)}{\partial \theta'} \right]$$

because

$$(10) \quad E \left[\frac{\partial \log f(y_i; \theta_*)}{\partial \theta} \right] \neq 0.$$

The fact that (6) or the expectation of (8) is zero does not imply that each individual term $\partial \log f(y_i; \theta_*) / \partial \theta$ has zero expectation. This point was stressed in Chow [2, pp. 205-206], where the covariance matrix (7) was exhibited and its estimation was discussed.²

To illustrate this point, let the true model be a normal linear regression model with two sets of explanatory variables

$$(11) \quad Y = X_1 \beta_{01} + X_2 \beta_{02} + \varepsilon \quad (E\varepsilon\varepsilon' = I\sigma_0^2)$$

and let the misspecified model result from omitting the second set

$$(12) \quad Y = X_1 \beta_{*1} + \varepsilon^*.$$

The log-likelihood of the latter model is

$$(13) \quad -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2} (y - X_1 \beta_1)' (y - X_1 \beta_1) / \sigma^2 .$$

The expectation of (13) evaluated by the true model (11) is

$$(14) \quad -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2} E[X_1(\beta_{01} - \beta_1) + X_2 \beta_{02} + \epsilon]' [X_1(\beta_{01} - \beta_1) + X_2 \beta_{02} + \epsilon] / \sigma^2$$

$$= -\frac{n}{2} \log(2\pi) - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} [X_1(\beta_{01} - \beta_1) + X_2 \beta_{02}]' [X_1(\beta_{01} - \beta_1) + X_2 \beta_{02}]$$

$$= -\frac{n\sigma_0^2}{2\sigma^2} .$$

Maximizing (14) with respect to β_1 and σ^2 yields

$$(15a) \quad \beta_{*1} = \beta_{01} + (X_1' X_1)^{-1} X_1' X_2 \beta_{02}$$

$$(15b) \quad \sigma_*^2 = \sigma_0^2 + n^{-1} \beta_{02}' X_2' [I - X_1 (X_1' X_1)^{-1} X_1'] X_2 \beta_{02} .$$

For this model, with x_{i1}' denoting the i -th row of X_1 ,

$$\log f(y_i; \theta) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \sigma^2 - \frac{1}{2} (y_i - x_{i1}' \beta_1)^2 / \sigma^2$$

and the derivative of $\log f(y_i; \theta)$ with respect to β_1 evaluated at β_{*1} and σ_*^2 is

$$(16) \quad x_{i1}' (y_i - x_{i1}' \beta_{*1}) / \sigma_*^2 .$$

The expectation of the column vector (16) is not zero because $(y_i - x_{i1}' \beta_{*1})$ does not necessarily have zero expectation for each i . Note that the sum of the terms (16) over i has zero expectation, or the derivative of (13) with respect to β_1 evaluated at β_{*1} and σ_*^2 has zero expectation since

$$\begin{aligned}
 (17) \quad E \sum_i x_{i1} (y_i - x_{i1}' \beta_{*1}) &= E(x_{11} \dots x_{n1}) (y - x_1' \beta_{*1}) = E x_1' (y - x_1' \beta_{*1}) \\
 &= x_1' [x_1 \beta_{01} + x_2 \beta_{02} - x_1 \beta_{01} - x_1 (x_1' x_1)^{-1} x_1' x_2 \beta_{02}] \\
 &= x_1' [I - x_1 (x_1' x_1)^{-1} x_1'] x_2 \beta_{02} = 0.
 \end{aligned}$$

For this example with $\theta' = (\beta_1' \ \sigma^2)$, Chow [1, p. 26] has evaluated

$$\text{Cov} \left[\frac{\partial \log L(Y; \theta_*)}{\partial \beta_1} \right] = E \left[\frac{\partial \log L(Y; \theta_*)}{\partial \beta_1} \cdot \frac{\partial \log L(Y; \theta_*)}{\partial \beta_1'} \right] = \frac{\sigma_0^2}{\sigma_*^4} x_1' x_1$$

which the reader can easily check. Using White's formula and (16), one would obtain

$$\text{Cov} \left[\frac{\partial \log L(Y; \theta_*)}{\partial \beta_1} \right] = \sum_{i=1}^n E \left[\frac{\partial \log f(y_i; \theta_*)}{\partial \beta_1} \cdot \frac{\partial \log f(y_i; \theta_*)}{\partial \beta_1'} \right] = \sigma_*^{-4} \sum_{i=1}^n x_{i1} x_{i1}' E(y_i - x_{i1}' \beta_{*1})^2$$

which is incorrect since $E(y_i - x_{i1}' \beta_{*1})^2$ ($i = 1, \dots, n$) do not all equal σ_0^2 , and not even on the average.

The covariance matrix of the ML estimator for θ_* of a misspecified model depends on the true model. It would be nice if the covariance matrix could be consistently estimated without specifying the true model, as White claims. If White were correct, (9) could be replaced by its sample analogue, with E omitted and θ_* replaced by $\hat{\theta}$, to construct a consistent estimate of (7), with $A(\theta_*)$ also replaced by its sample analogue

$$-n^{-1} \frac{\log L(Y; \hat{\theta})}{\partial \theta \partial \theta'}$$

which is consistent. Unfortunately, the error of using (9) invalidates White's

method of estimating the covariance matrix of the ML estimator for misspecified models as well as the other claims of his paper [5] which depend on this covariance matrix. The latter include his proposed modifications of the Lagrangian multiplier and Wald test statistics [5, Theorems 3.4 and 3.5].

There may be situations in which the equality sign instead of the inequality sign holds in equation (10) for a misspecified model, so that White's estimate of the covariance matrix is consistent. However, as our regression example illustrates, most misspecified models in econometrics do not fall into these situations. The ML estimation of a misspecified simultaneous-equations model has been treated in Chow [2] and [3] where a correctly specified model is postulated.

REFERENCES

- [1] Chow, G. C.: "A Comparison of the Information and Posterior Probability Criteria for Model Selection," Journal of Econometrics, 16 (1981), 21-33.
- [2] Chow, G. C.: "Selection of Econometric Models by the Information Criterion," in E. G. Charatsis, ed., Proceedings of the Econometric Society European Meeting 1979, 199-214. Amsterdam: North-Holland Publishing Company, 1981.
- [3] Chow, G. C.: "Evaluation of Econometric Models by Decomposition and Aggregation," in J. Kmenta and J. B. Ramsey, eds., Large-Scale Macro-Econometric Models, 423-444. Amsterdam: North-Holland Publishing Company, 1981.
- [4] Silvey, S. D.: "The Lagrangian Multiplier Test," Annals of Mathematical Statistics, 30 (1959), 389-407.
- [5] White, H.: "Maximum Likelihood Estimation of Misspecified Models," Econometrica, 50 (1982), 1-25.

FOOTNOTES

¹I would like to thank Adrian Pagan for helpful comments and acknowledge financial support from the National Science Foundation through Grant No. SES80-12582.

²I quote from the above reference:

"However the estimation of $\text{Cov}[\partial \log L^* / \partial \theta_1]$ is not so simple.

One might propose the estimate

$$\sum_{i=1}^n \left[\frac{\partial \log f(y_i | \hat{\theta}^*)}{\partial \theta_1} \cdot \frac{\log f(y_i | \hat{\theta}^*)}{\partial \theta_1'} \right] \quad (2.12)$$

in view of the relation

$$\text{Cov} \left[\frac{\partial \log L(Y, \theta^*)}{\partial \theta_1} \right] = \sum_{i=1}^n \text{Cov} \left[\frac{\partial \log f(y_i | \theta^*)}{\partial \theta_1} \right]. \quad (2.13)$$

The estimate (2.12) is inappropriate because, while the expectation of $\partial \log L(Y, \theta^*) / \partial \theta_1$ is zero, the expectations of its components $\partial \log f(y_i | \theta^*) / \partial \theta_1$ ($i=1, \dots, n$) are not zero when the correct model is $f(\cdot | \theta^0)$. Eq. (3.11) of section 3, for example, shows that $\partial \log f(y_i | \theta^*) / \partial \theta_1$ is not zero where θ_1 consists of coefficients of the exogenous variables in a simultaneous-equation model, to be denoted by the elements of the matrix B in section 3."