Econometric Theory, **23**, 2007, 1022–1032. Printed in the United States of America.
DOI: 10.1017/S0266466607070417

THE ASYMPTOTIC VARIANCE OF THE PSEUDO MAXIMUM LIKELIHOOD ESTIMATOR

Jan R. Magnus
Tilburg University
and
University of Tokyo

We present an analytical closed-form expression for the asymptotic variance matrix in the misspecified multivariate regression model.

1. INTRODUCTION

Since the classic papers of Akaike (1973), White (1982), and Vuong (1989), there has existed a growing literature devoted to the study of misspecified models. Furthermore, during the last decade, the "sandwich" variance matrix (also known as the "robust" variance matrix) has been shown to be the proper variance matrix in misspecified models and has been widely used. The sandwich variance matrix estimation procedure was introduced by Huber (1967) and White (1982), and it yields consistent variance matrix estimators, also (and in particular) when the assumed model is misspecified.

The objective of this paper is to derive the analytical closed-form expression of the sandwich variance matrix within the context of the misspecified multivariate regression model. We also derive scalar measures of the asymptotic variance, in particular the trace, determinant, and norm, that play a role in the construction of information criteria. An example of such an application is provided in Bozdogan (2007), where the information complexity (ICOMP) criterion is used to extend the results of Bozdogan and Haughton (1998) from the univariate misspecified regression model to the multivariate case.

2. MULTIVARIATE NORMAL REGRESSION

Consider a set of n vectors $y_1, \dots, y_n$, each of order $p \times 1$, whose first two moments are given by

$$\mathrm{E}(y_i) = B'x_i, \qquad \mathrm{var}(y_i) = \Sigma,$$

I am grateful to Hamparsum Bozdogan of the University of Tennessee for bringing the idea of the sandwich variance matrix within the context of the misspecified multivariate regression model to my attention and to two referees for their constructive and useful comments. Address correspondence to Jan R. Magnus, Department of Econometrics & Operations Research, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands; e-mail: magnus@uvt.nl.

  0266-4666/07 $15.00

where B is a $k \times p$ matrix of unknown coefficients, $X := (x_1, \ldots, x_n)'$ is a nonrandom $n \times k$ matrix of full column rank k, and $\Sigma = (\sigma_{ij})$ is a positive definite unknown $p \times p$ matrix. The full set of coefficients is thus $\theta := ((\operatorname{vec} B)', (\operatorname{vech}(\Sigma))')'$, of order $(kp + \frac{1}{2}p(p+1)) \times 1$, where $\operatorname{vech}(\cdot)$ denotes the half-vec operator defined in the Appendix. Assume that y_i and y_j are uncorrelated for all $i \neq j$ and let $Y := (y_1, \ldots, y_n)'$, of order $n \times p$. Finally, let $n \geq p + k$; this is a necessary condition without which the estimator $\hat{\Sigma}$ in (3), which follows, would be singular. These assumptions imply that

$$\mathrm{E}(Y) = XB, \qquad \operatorname{var}(\operatorname{vec} Y) = \Sigma \otimes I_n.$$

If, in addition, we assume normality, then the log-likelihood function of the sample $y_1, \ldots, y_n$ is given by

$$\ell(\theta) = -\frac{np}{2}\log 2\pi - \frac{n}{2}\log|\Sigma| - \frac{1}{2}\operatorname{tr}(Y - XB)\Sigma^{-1}(Y - XB)'; \tag{1}$$

see, for example, Magnus and Neudecker (1988, p. 321). The first differential of the log-likelihood is

$$\begin{aligned} \mathrm{d}\ell &= -\frac{n}{2}\operatorname{tr}\Sigma^{-1}\,\mathrm{d}\Sigma + \frac{1}{2}\operatorname{tr}(Y - XB)\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}(Y - XB)' \\ &\quad + \operatorname{tr} X(\mathrm{d}B)\Sigma^{-1}(Y - XB)' \\ &= \frac{1}{2}\operatorname{tr}(\Sigma^{-1}(Y - XB)'(Y - XB)\Sigma^{-1} - n\Sigma^{-1})\mathrm{d}\Sigma \\ &\quad + \operatorname{tr}\Sigma^{-1}(Y - XB)'X\,\mathrm{d}B, \end{aligned} \tag{2}$$

leading to the first-order conditions

$$\Sigma^{-1}(Y - XB)'(Y - XB)\Sigma^{-1} = n\Sigma^{-1}, \qquad X'(Y - XB)\Sigma^{-1} = 0,$$

and hence to the maximum likelihood estimators

$$\hat{B} = (X'X)^{-1}X'Y, \qquad \hat{\Sigma} = \frac{(Y - X\hat{B})'(Y - X\hat{B})}{n} = \frac{Y'MY}{n}, \tag{3}$$

where $M := I_n - X(X'X)^{-1}X'$ is the usual idempotent matrix.

Taking the differential of (2), we obtain the second differential of the log-likelihood as

$$\begin{aligned} \mathrm{d}^2\ell &= \operatorname{tr}(\mathrm{d}\Sigma^{-1})(Y - XB)'(Y - XB)\Sigma^{-1}\,\mathrm{d}\Sigma - \frac{n}{2}\operatorname{tr}(\mathrm{d}\Sigma^{-1})\,\mathrm{d}\Sigma \\ &\quad + 2\operatorname{tr}(\mathrm{d}\Sigma^{-1})(Y - XB)'X\,\mathrm{d}B - \operatorname{tr}\Sigma^{-1}(\mathrm{d}B)'X'X\,\mathrm{d}B. \end{aligned}$$

Then, using the facts that $\mathrm{E}(Y - XB) = 0$ and $\mathrm{E}(Y - XB)'(Y - XB) = n\Sigma$, we find

$$
\begin{aligned}
-\mathrm{E}\,\mathrm{d}^2\ell &= \frac{n}{2}\,\mathrm{tr}\,\Sigma^{-1}(\mathrm{d}\Sigma)\Sigma^{-1}\,\mathrm{d}\Sigma + \mathrm{tr}\,\Sigma^{-1}(\mathrm{d}B)'X'X\,\mathrm{d}B \\
&= \frac{n}{2}\,(\mathrm{d}\,\mathrm{vech}(\Sigma))'D_p'(\Sigma^{-1} \otimes \Sigma^{-1})D_p\,\mathrm{d}\,\mathrm{vech}(\Sigma) \\
&\quad + (\mathrm{d}\,\mathrm{vec}\,B)'(\Sigma^{-1} \otimes X'X)\,\mathrm{d}\,\mathrm{vec}\,B, \qquad \textbf{(4)}
\end{aligned}
$$

where D_p denotes the $p^2 \times \frac{1}{2}p(p+1)$ duplication matrix, defined in the Appendix. Hence we obtain the following result.

THEOREM 1. *In the correctly specified case, the information matrix is given by*

$$
\mathcal{I} = \begin{pmatrix} \Sigma^{-1} \otimes X'X & 0 \\ 0 & \frac{n}{2}\,D_p'(\Sigma^{-1} \otimes \Sigma^{-1})D_p \end{pmatrix}
$$

and its inverse by

$$
\mathcal{I}^{-1} = \begin{pmatrix} \Sigma \otimes (X'X)^{-1} & 0 \\ 0 & \frac{2}{n}\,D_p^{+}(\Sigma \otimes \Sigma)D_p^{+\prime} \end{pmatrix},
$$

where $D_p^+ = (D_p'D_p)^{-1}D_p'$. *Furthermore,*

$$
\mathrm{tr}\,\mathcal{I}^{-1} = (\mathrm{tr}\,\Sigma)(\mathrm{tr}(X'X)^{-1}) + \frac{1}{2n}\left(\mathrm{tr}\,\Sigma^2 + (\mathrm{tr}\,\Sigma)^2 + 2\sum_{j=1}^{p}\sigma_{jj}^2\right)
$$

and

$$
|\mathcal{I}^{-1}| = 2^p n^{-(1/2)p(p+1)}|\Sigma|^{p+k+1}|X'X|^{-p}.
$$

Proof. The information matrix $\mathcal{I}$ follows from the fact that we can write (4) as $-\mathrm{E}\,\mathrm{d}^2\ell = (\mathrm{d}\theta)'\mathcal{I}(\mathrm{d}\theta)$. Its inverse follows from Magnus and Neudecker (1988, Thm. 3.13(d), p. 50), and the trace and determinant follow from Lemma A1 in the Appendix. ■

The inverse $\mathcal{I}^{-1}$ of the information matrix provides the asymptotic variance of the maximum likelihood estimator in the correctly specified case. Its trace and determinant provide scalar measures of the asymptotic variance, and they play a role, inter alia, in the construction of information criteria.

3. MULTIVARIATE REGRESSION UNDER MISSPECIFICATION

We next assume the same model as in Section 2, except that we do not assume normality. The first two moments of Y are still given by $\mathrm{E}(Y) = XB$ and $\mathrm{var}(\mathrm{vec}\, Y) = \Sigma \otimes I_n$, but the third and fourth moments of Y are not necessarily equal to the moments that would have been implied by normality.

We estimate the unknown parameters by pseudo maximum likelihood (PML); that is, we take the normal log-likelihood function (1) as our starting point. The PML estimators are given by (3). The expectation of the first differential is still zero (first-order regularity), but it is no longer true that $\mathrm{E}(\mathrm{d}\ell)^2 = -\mathrm{E}\,\mathrm{d}^2\ell$ (second-order regularity). This is because the evaluation of $\mathrm{E}(\mathrm{d}\ell)^2$ involves third and fourth moments.

Let us standardize Y by defining $V := (Y - XB)\Sigma^{-1/2}$, so that

$$\mathrm{E}(V) = 0, \qquad \mathrm{var}(\mathrm{vec}\, V) = I_{pn}.$$

Let us also introduce matrix generalizations of the usual skewness and kurtosis measures by defining

$$\Gamma_1 := \mathrm{E}(\mathrm{vec}\, V)(\mathrm{vec}(V'V - nI_p))', \qquad \Gamma_2 := \mathrm{E}(\mathrm{vec}\, V'V)(\mathrm{vec}\, V'V)'.$$

In the special case of correct specification, this specializes to

$$\Gamma_1 = 0, \qquad \Gamma_2 = 2nN_p + n^2(\mathrm{vec}\, I_p)(\mathrm{vec}\, I_p)',$$

where N_p denotes the $p^2 \times p^2$ symmetrizer matrix defined in the Appendix. If $n = p = 1$, the kurtosis further specializes to $\Gamma_2 = 3$, as expected.

We now evaluate $\mathrm{E}(\mathrm{d}\ell)^2$. Squaring equation (2) gives

$$(\mathrm{d}\ell)^2 = \left(\frac{1}{2}\,\mathrm{tr}(\Sigma^{-1/2}V'V\Sigma^{-1/2} - n\Sigma^{-1})\mathrm{d}\Sigma + \mathrm{tr}\,\Sigma^{-1/2}V'X\,\mathrm{d}B\right)^2.$$

Then, letting $\Delta := D_p'(\Sigma^{-1/2} \otimes \Sigma^{-1/2})D_p$, we find

$$\begin{aligned}
\mathrm{E}(\mathrm{d}\ell)^2 &= \frac{1}{4}\,\mathrm{E}(\mathrm{tr}(\Sigma^{-1/2}V'V\Sigma^{-1/2} - n\Sigma^{-1})\,\mathrm{d}\Sigma)^2 + \mathrm{E}(\mathrm{tr}\,\Sigma^{-1/2}V'X\,\mathrm{d}B)^2 \\
&\quad + \mathrm{E}(\mathrm{tr}(\Sigma^{-1/2}V'V\Sigma^{-1/2} - n\Sigma^{-1})\,\mathrm{d}\Sigma)(\mathrm{tr}\,\Sigma^{-1/2}V'X\,\mathrm{d}B) \\
&= \frac{1}{4}\,(\mathrm{d}\,\mathrm{vec}\,\Sigma)'(\Sigma^{-1/2} \otimes \Sigma^{-1/2})\mathrm{var}(\mathrm{vec}\, V'V)(\Sigma^{-1/2} \otimes \Sigma^{-1/2})\mathrm{d}\,\mathrm{vec}\,\Sigma \\
&\quad + (\mathrm{d}\,\mathrm{vec}\, B)'(\Sigma^{-1/2} \otimes X')\mathrm{var}(\mathrm{vec}\, V)(\Sigma^{-1/2} \otimes X)\,\mathrm{d}\,\mathrm{vec}\, B \\
&\quad + (\mathrm{d}\,\mathrm{vec}\,\Sigma)'(\Sigma^{-1/2} \otimes \Sigma^{-1/2})\Gamma_1'(\Sigma^{-1/2} \otimes X)\,\mathrm{d}\,\mathrm{vec}\, B
\end{aligned}$$

$$= \frac{1}{4}(\mathrm{d}\,\mathrm{vech}(\Sigma))'\Delta D_p^+(\Gamma_2 - n^2(\mathrm{vec}\, I_p)(\mathrm{vec}\, I_p)')D_p^{+\prime}\Delta\, \mathrm{d}\,\mathrm{vech}(\Sigma)$$
$$+ (\mathrm{d}\,\mathrm{vec}\, B)'(\Sigma^{-1} \otimes X'X)\mathrm{d}\,\mathrm{vec}\, B$$
$$+ (\mathrm{d}\,\mathrm{vech}(\Sigma))'\Delta D_p^+ \Gamma_1'(\Sigma^{-1/2} \otimes X)\mathrm{d}\,\mathrm{vec}\, B. \quad \textbf{(5)}$$

Thus we obtain the following result.

THEOREM 2. *In the misspecified case, we have*

$$-\mathrm{E}(\mathrm{d}^2\ell) = (\mathrm{d}\theta)'\mathcal{I}\,\mathrm{d}\theta, \qquad \mathrm{E}(\mathrm{d}\ell)^2 = (\mathrm{d}\theta)'\mathcal{R}\,\mathrm{d}\theta,$$

where $\mathcal{I}$ is given in Theorem 1,

$$\mathcal{R} := \begin{pmatrix} \Sigma^{-1} \otimes X'X & \frac{1}{2}(\Sigma^{-1/2} \otimes X')\Gamma_1 D_p^{+\prime}\Delta \\ \frac{1}{2}\Delta D_p^+ \Gamma_1'(\Sigma^{-1/2} \otimes X) & \frac{1}{4}\Delta D_p^+ \Gamma_2^* D_p^{+\prime}\Delta \end{pmatrix},$$

and $\Gamma_2^* := \Gamma_2 - n^2(\mathrm{vec}\, I_p)(\mathrm{vec}\, I_p)'$.

Proof. The expression $-\mathrm{E}(\mathrm{d}^2\ell)$ is not affected by the misspecification because it uses the first two moments only. Hence the matrix $\mathcal{I}$ is the same as in Theorem 1. In contrast, equation (5) implies the expression for $\mathcal{R}$. ∎

The matrix $\mathcal{R}$ is sometimes called the "outer-product form" of the information matrix because it is based on $\mathrm{E}(\mathrm{d}\ell)^2$. The "Hessian form" $\mathcal{I}$ is based on $-\mathrm{E}(\mathrm{d}^2\ell)$. In the correctly specified case where $\Gamma_1 = 0$ and $\Gamma_2^* = 2nN_p$, one verifies that $\mathcal{R} = \mathcal{I}$.

4. ASYMPTOTIC VARIANCE OF THE PML ESTIMATOR

We have seen that, in the presence of misspecification, second-order regularity does not hold and that therefore $\mathcal{I}$ and $\mathcal{R}$ are not the same. The asymptotic variance of the PML estimator $\hat{\theta}$ is therefore not given by either $\mathcal{I}^{-1}$ or $\mathcal{R}^{-1}$ but rather by $\mathcal{V} := \mathcal{I}^{-1}\mathcal{R}\mathcal{I}^{-1}$. This important result was implied or proved in papers by Huber (1967), Jennrich (1969), Malinvaud (1970), Gallant and Holly (1980), Burguete, Gallant, and Souza (1982), White (1982), and Gouriéroux, Monfort, and Trognon (1984) and more recently by Gouriéroux and Monfort (1995a, p. 237), Gouriéroux and Monfort (1995b, p. 170), Hendry (1995, p. 391), and White (1996).

Although the sandwich matrix $\hat{\mathcal{V}} := \mathcal{V}(\hat{\theta})$ evaluated at the maximum likelihood estimator $\hat{\theta}$ provides a consistent estimator of the variance of $\hat{\theta}$, it is not the only consistent estimator. An alternative would be to evaluate minus the

Hessian matrix H (instead of $\mathcal{I}$) and the sample variance of the score contributions R (instead of $\mathcal{R}$) and to use these in constructing $V := H^{-1}RH^{-1}$, as in White (1982). The estimator $\hat{V} := V(\hat{\theta})$ is also consistent and hence an alternative to $\hat{\mathcal{V}}$. It is difficult to judge, in general, which estimator is to be preferred. In our case, the alternative estimator $\hat{V}$ allows for heteroskedasticity and "heteroskewness" which is excluded by our model assumptions, and this might be one reason to prefer $\hat{\mathcal{V}}$ over $\hat{V}$. Our main result is as follows.

THEOREM 3. *The sandwich matrix $\mathcal{V}$ is given by*

$$\mathcal{V} = \begin{pmatrix} \Sigma \otimes (X'X)^{-1} & \frac{1}{n}(\Sigma^{1/2} \otimes (X'X)^{-1}X')\Gamma_1 D_p \Delta^{-1} \\ \frac{1}{n}\Delta^{-1}D_p'\Gamma_1'(\Sigma^{1/2} \otimes X(X'X)^{-1}) & \frac{1}{n^2}\Delta^{-1}D_p'\Gamma_2^* D_p \Delta^{-1} \end{pmatrix}.$$

The trace and determinant of $\mathcal{V}$ are

$$\operatorname{tr}(\mathcal{V}) = (\operatorname{tr}\Sigma)(\operatorname{tr}(X'X)^{-1}) + \frac{1}{n^2}\operatorname{tr} D_p^+(\Sigma^{1/2} \otimes \Sigma^{1/2})\Gamma_2^*(\Sigma^{1/2} \otimes \Sigma^{1/2})D_p^{+\prime}$$

and

$$|\mathcal{V}| = 2^{-p(p-1)} n^{-p(p+1)} |\Sigma|^{p+k+1} |X'X|^{-p} \times |D_p'(\Gamma_2^* - \Gamma_1'(I_p \otimes X(X'X)^{-1}X')\Gamma_1)D_p|,$$

and the norm of $\mathcal{V}$, defined as $\|\mathcal{V}\| := \sqrt{\operatorname{tr}(\mathcal{V}^2)}$, is the square root of

$$\operatorname{tr}(\mathcal{V}^2) = \operatorname{tr}(\Sigma^2)\operatorname{tr}((X'X)^{-2}) + \frac{1}{n^4}\operatorname{tr}(\Gamma_2^* Q)^2 + \frac{2}{n^2}\operatorname{tr}[(\Sigma \otimes X(X'X)^{-2}X')(\Gamma_1 Q \Gamma_1')],$$

where

$$Q := \frac{1}{2} N_p(\Sigma \otimes \Sigma)N_p + \frac{1}{2}(\Sigma^{1/2} \otimes \Sigma^{1/2})\Xi_p(\Sigma^{1/2} \otimes \Sigma^{1/2})$$

and $\Xi_p := \sum_{i=1}^{p}(e_i e_i' \otimes e_i e_i')$. The vectors e_i are unit vectors, so that e_i denotes the ith column of the identity matrix I_p.

Proof. From Theorems 1 and 2 it follows that the matrix $\mathcal{I}^{-1}\mathcal{R}$ is equal to

$$\begin{pmatrix} I_{pk} & \frac{1}{2}(\Sigma^{1/2} \otimes (X'X)^{-1}X')\Gamma_1 D_p^{+\prime}\Delta \\ \frac{1}{n} D_p^+(\Sigma \otimes \Sigma)D_p^{+\prime}\Delta D_p^+ \Gamma_1'(\Sigma^{-1/2} \otimes X) & \frac{1}{2n} D_p^+(\Sigma \otimes \Sigma)D_p^{+\prime}\Delta D_p^+ \Gamma_2^* D_p^{+\prime}\Delta \end{pmatrix},$$

so that the expression for $\mathcal{V}$ follows from the properties of N_p and D_p and the fact that $D_p^{+\prime}\Delta D_p^+(\Sigma \otimes \Sigma)D_p^{+\prime} = D_p\Delta^{-1}$. Furthermore,

$$\begin{aligned} \operatorname{tr}(\mathcal{V}) &= \operatorname{tr}\Sigma \otimes (X'X)^{-1} + \frac{1}{n^2}\operatorname{tr}\Delta^{-1}D_p'\Gamma_2^* D_p\Delta^{-1} \\ &= (\operatorname{tr}\Sigma)(\operatorname{tr}(X'X)^{-1}) \\ &\quad + \frac{1}{n^2}\operatorname{tr} D_p^+(\Sigma^{1/2} \otimes \Sigma^{1/2})\Gamma_2^*(\Sigma^{1/2} \otimes \Sigma^{1/2})D_p^{+\prime}, \end{aligned}$$

and

$$\begin{aligned} |\mathcal{V}| &= |\Sigma \otimes (X'X)^{-1}| \times \left|\frac{1}{n^2}\Delta^{-1}D_p'(\Gamma_2^* - \Gamma_1'(I_p \otimes X(X'X)^{-1}X')\Gamma_1)D_p\Delta^{-1}\right| \\ &= 2^{-p(p-1)}n^{-p(p+1)}|\Sigma|^{p+k+1}|X'X|^{-p} \\ &\quad \times |D_p'(\Gamma_2^* - \Gamma_1'(I_p \otimes X(X'X)^{-1}X')\Gamma_1)D_p|. \end{aligned}$$

Next we compute $\operatorname{tr}(\mathcal{V}^2)$. Denote the four blocks of $\mathcal{V}$ by $\mathcal{V}_{ij}$ $(i,j = 1,2)$. Then,

$$\operatorname{tr}(\mathcal{V}^2) = \operatorname{tr}(\mathcal{V}_{11}^2) + \operatorname{tr}(\mathcal{V}_{22}^2) + 2\operatorname{tr}(\mathcal{V}_{12}\mathcal{V}_{21}).$$

Now,

$$\operatorname{tr}(\mathcal{V}_{11}^2) = \operatorname{tr}(\Sigma \otimes (X'X)^{-1})^2 = \operatorname{tr}(\Sigma^2)\operatorname{tr}((X'X)^{-2}),$$

and, using Lemma A2 in the Appendix,

$$\begin{aligned} \operatorname{tr}(\mathcal{V}_{22}^2) &= \frac{1}{n^4}\operatorname{tr}(\Delta^{-1}D_p'\Gamma_2^* D_p\Delta^{-1})^2 \\ &= \frac{1}{n^4}\operatorname{tr}[\Gamma_2^*(\Sigma^{1/2} \otimes \Sigma^{1/2})(D_pD_p')^+(\Sigma^{1/2} \otimes \Sigma^{1/2})]^2 = \frac{1}{n^4}\operatorname{tr}(\Gamma_2^* Q)^2, \end{aligned}$$

and

$$\operatorname{tr}(\mathcal{V}_{12}\mathcal{V}_{21}) = \frac{1}{n^2}\operatorname{tr}[(\Sigma^{1/2}\otimes(X'X)^{-1}X')\Gamma_1 D_p\Delta^{-2}D_p'\Gamma_1'(\Sigma^{1/2}\otimes X(X'X)^{-1})]$$

$$= \frac{1}{n^2}\operatorname{tr}[(\Sigma\otimes X(X'X)^{-2}X')(\Gamma_1 Q\Gamma_1')].$$

This completes the proof. ■

The sandwich matrix $\mathcal{V}$ thus provides the asymptotic variance of the PML estimator in the misspecified case. As in Theorem 1, its trace, determinant, and norm provide scalar measures of the asymptotic variance. These measures, together with other scalars such as $\operatorname{tr}(\mathcal{I}^{-1}\mathcal{R})$, play a crucial role in the construction of information criteria.

An interesting special case is obtained when the true joint distribution belongs to the linear exponential family, giving rise to the well-known quasi-generalized PML estimators; see Gouriéroux et al. (1984, Sect. 5). We do not, however, investigate this avenue in this paper.

We notice, after a little algebra, that

$$\operatorname{tr}(\mathcal{I}^{-1}\mathcal{R}) = \operatorname{tr}(I_{pk}) + \frac{1}{2n}\operatorname{tr}(D_p^+(\Sigma\otimes\Sigma)D_p^{+\prime}\Delta D_p^+\Gamma_2^* D_p^{+\prime}\Delta)$$

$$= pk + \frac{1}{2n}\operatorname{tr} N_p\Gamma_2^* = pk + \frac{1}{2n}\operatorname{tr}\Gamma_2^*, \tag{6}$$

which simplifies to

$$\operatorname{tr}(\mathcal{I}^{-1}\mathcal{R}) = pk + \frac{1}{2}p(p+1) \tag{7}$$

in the special case of correct specification where $\Gamma_1 = 0$ and $\Gamma_2^* = 2nN_p$.

We also notice that, in the case of correct specification,

$$\operatorname{tr}(\mathcal{V}^2) = \operatorname{tr}(\Sigma^2)\operatorname{tr}((X'X)^{-2}) + \frac{4}{n^2}\operatorname{tr}(Q^2)$$

$$= \operatorname{tr}(\Sigma^2)\operatorname{tr}((X'X)^{-2}) + \frac{1}{2n^2}(\operatorname{tr}\Sigma^2)^2 + \frac{1}{2n^2}\operatorname{tr}(\Sigma^4)$$

$$+ \frac{1}{n^2}\sum_{ij}\sigma_{ij}^4 + \frac{2}{n^2}\sum_i\Big(\sum_j\sigma_{ij}^2\Big)^2. \tag{8}$$

REFERENCES

Akaike, H. (1973) Information theory as an extension of the maximum likelihood principle. In B.N. Petrov & F. Csaki (eds.), *Second International Symposium on Information Theory*, pp. 267–281. Akademiai Kiado.

Bozdogan, H. (2007) Misspecified Multivariate Linear Regression Models Using Genetic Algorithm and Information Complexity as the Fitness Function. Working paper, Department of Statistics, University of Tennessee.

Bozdogan, H. & D.M.A. Haughton (1998) Informational complexity criteria for regression models. *Computational Statistics and Data Analysis* 28, 51–76.

Burguete, J., R. Gallant, & G. Souza (1982) On unification of the asymptotic theory of nonlinear econometric models. *Econometric Reviews* 1, 151–190.

Gallant, R. & A. Holly (1980) Statistical inference in an implicit, nonlinear, simultaneous equation model in the context of maximum likelihood estimation. *Econometrica* 48, 697–720.

Gouriéroux, C. & A. Monfort (1995a) *Statistics and Econometric Models*, vol. 1. Cambridge University Press.

Gouriéroux, C. & A. Monfort (1995b) *Statistics and Econometric Models*, vol. 2. Cambridge University Press.

Gouriéroux, C., A. Monfort, & A. Trognon (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52, 681–700.

Hendry, D.F. (1995) *Dynamic Econometrics*. Oxford University Press.

Huber, P.J. (1967) The behavior of maximum likelihood estimates under non-standard conditions. In L.M. LeCam & J. Neyman (eds.), *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 221–233. University of California Press.

Jennrich, R. (1969) Asymptotic properties of nonlinear least squares estimators. *Annals of Mathematical Statistics* 40, 633–643.

Magnus, J.R. (1988) *Linear Structures*. Charles Griffin & Company and Oxford University Press.

Magnus, J.R. & H. Neudecker (1980) The elimination matrix: Some lemmas and applications. *SIAM Journal on Algebraic and Discrete Methods* 1, 422–449.

Magnus, J.R. & H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley. Revised edition, 1999.

Malinvaud, E. (1970) The consistency of nonlinear regressions. *Annals of Mathematical Statistics* 41, 956–969.

Vuong, Q.H. (1989) Likelihood ratio tests for model selection and non-nested hypotheses. *Econometrica* 57, 307–333.

White, H. (1982) Maximum likelihood estimation of misspecified models. *Econometrica* 50, 1–26.

White, H. (1996) *Estimation, Inference and Specification Analysis*. Cambridge University Press.

APPENDIX: The Duplication Matrix— Some New Results

Let A be a square matrix of order $p \times p$. The two vectors $\operatorname{vec} A$ and $\operatorname{vec} A'$ contain the same p^2 components but in a different order. Hence there exists a unique permutation matrix that transforms $\operatorname{vec} A$ into $\operatorname{vec} A'$. This $p^2 \times p^2$ matrix is (a special case of) the *commutation matrix* and is denoted by K_p; it is implicitly defined by the operation $K_p \operatorname{vec} A = \operatorname{vec} A'$.

Closely related to the commutation matrix is the $p^2 \times p^2$ *symmetrizer matrix* N_p with the property $N_p \operatorname{vec} A = \frac{1}{2}\operatorname{vec}(A + A')$ for every square $p \times p$ matrix A. It is easy to see that $N_p = \frac{1}{2}(I_{p^2} + K_p)$.

We now introduce the half-vec operator vech$(\cdot)$. For any $p \times p$ matrix A, the vector vech(A) denotes the $\frac{1}{2}p(p+1) \times 1$ vector that is obtained from vec A by eliminating all supradiagonal elements of A. For example, for $p = 2$,

$$\operatorname{vec} A = (a_{11}, a_{21}, a_{12}, a_{22})' \quad \text{and} \quad \operatorname{vech}(A) = (a_{11}, a_{21}, a_{22})',$$

where the supradiagonal element a_{12} has been removed. Thus, for symmetric A, vech(A) only contains the distinct elements of A. Now, if A is symmetric, the elements of vec A are those of vech(A) with some repetitions. Hence, there exists a unique $p^2 \times \frac{1}{2}p(p+1)$ matrix D_p, called the *duplication matrix*, that transforms, for symmetric A, vech(A) into vec A, that is,

$$D_p \operatorname{vech}(A) = \operatorname{vec} A \qquad (A = A').$$

The matrices D_p and N_p are connected through $D_p D_p^+ = N_p$. The duplication matrix was introduced by Magnus and Neudecker (1980). A systematic treatment of K_p, N_p, and D_p, among others, is given in Magnus (1988).

We now present two new properties, both of which are used in this note.

LEMMA A1. *Let $A = (a_{ij})$ be a square matrix of order $p \times p$. The determinant and trace of the matrix $D_p^+(A \otimes A)D_p^{+\prime}$ are given by*

$$|D_p^+(A \otimes A)D_p^{+\prime}| = 2^{-(1/2)p(p-1)}|A|^{p+1}$$

and

$$\operatorname{tr}(D_p^+(A \otimes A)D_p^{+\prime}) = \frac{1}{4}\operatorname{tr}(A'A) + \frac{1}{4}(\operatorname{tr} A)^2 + \frac{1}{2}\sum_{j=1}^{p} a_{jj}^2.$$

Proof. Because

$$D_p^+(A \otimes A)D_p^{+\prime} = (D_p'D_p)^{-1}D_p'(A \otimes A)D_p(D_p'D_p)^{-1},$$

we obtain, from Magnus (1988, Thm. 4.11(i)),

$$\begin{aligned}|D_p^+(A \otimes A)D_p^{+\prime}| &= |D_p'D_p|^{-1}|D_p'(A \otimes A)D_p|\,|D_p'D_p|^{-1}\\ &= 2^{-(1/2)p(p-1)}2^{(1/2)p(p-1)}|A|^{p+1}2^{-(1/2)p(p-1)} = 2^{-(1/2)p(p-1)}|A|^{p+1}.\end{aligned}$$

This proves the first result. To prove the second result, let δ_{st} denote the Kronecker delta and write $u_{ij} = \operatorname{vech}(e_i e_j')$, where e_i denotes the ith column of the identity matrix I_p. Then,

$$\begin{aligned}
\operatorname{tr}(D_p^+(A\otimes A)D_p^{+\prime}) &= \operatorname{tr}(D_p^+(A\otimes A)D_p)(D_p'D_p)^{-1}\\
&= \frac{1}{2}\operatorname{tr}\Bigg(\sum_{i\geq j}\sum_{s\geq t}(a_{it}a_{js}+a_{is}a_{jt}-\delta_{st}a_{is}a_{js})u_{ij}u_{st}'\Bigg)\\
&\quad\times\Bigg(I_{(1/2)p(p+1)}+\sum_{k=1}^{p}u_{kk}u_{kk}'\Bigg)\\
&= \frac{1}{2}\sum_{i\geq j}(a_{ij}a_{ji}+a_{ii}a_{jj}-\delta_{ij}a_{ii}a_{jj})+\frac{1}{2}\sum_{j=1}^{p}a_{jj}^2\\
&= \frac{1}{4}\sum_{ij}a_{ij}a_{ji}+\frac{1}{4}\sum_{ij}a_{ii}a_{jj}+\frac{1}{2}\sum_{j=1}^{p}a_{jj}^2\\
&= \frac{1}{4}\operatorname{tr}(A'A)+\frac{1}{4}(\operatorname{tr}A)^2+\frac{1}{2}\sum_{j=1}^{p}a_{jj}^2,
\end{aligned}$$

where the second equality follows from the proof of Theorem 4.9 and Theorem 4.4(ii) in Magnus (1988). ∎

LEMMA A2. *Letting* $\Xi_p := \sum_{i=1}^{p}(e_ie_i'\otimes e_ie_i')$ *and* $\alpha_k := \frac{1}{2}^k$, *we have*

$$[(D_pD_p')^+]^k = \alpha_k N_p + (1-\alpha_k)\Xi_p \qquad (k=1,2,\ldots),$$

a weighted average of two idempotent matrices.

Proof. We prove the result first for $k=1$. Let $S_{ij} := (e_ie_j' + e_je_i')/2$. Then, using Theorem 4.6(ii) of Magnus (1988),

$$\begin{aligned}
(D_pD_p')^+ &= \sum_{i\geq j}(\operatorname{vec}S_{ij})(\operatorname{vec}S_{ij})'\\
&= \frac{1}{2}\sum_{i=1}^{p}(\operatorname{vec}S_{ii})(\operatorname{vec}S_{ii})' + \frac{1}{2}\sum_{i=1}^{p}\sum_{j=1}^{p}(\operatorname{vec}S_{ij})(\operatorname{vec}S_{ij})'\\
&= \frac{1}{2}\sum_{i=1}^{p}(\operatorname{vec}e_ie_i')(\operatorname{vec}e_ie_i')' + \frac{1}{8}\sum_{i,j}(\operatorname{vec}(e_ie_j'+e_je_i'))(\operatorname{vec}(e_ie_j'+e_je_i'))'\\
&= \frac{1}{2}\sum_{i=1}^{p}(e_ie_i'\otimes e_ie_i') + \frac{1}{4}\sum_{i,j}(e_ie_i'\otimes e_je_j + e_ie_j'\otimes e_je_i')\\
&= \frac{1}{2}\Xi_p + \frac{1}{4}(I_{p^2}+K_p) = \frac{1}{2}(\Xi_p + N_p),
\end{aligned}$$

because $K_p = \sum_{i=1}^{p}\sum_{j=1}^{p}(e_ie_j'\otimes e_je_i')$ by Theorem 3.2 in Magnus (1988). This proves the result for $k=1$. The general result follows by induction, using the facts that both N_p and Ξ_p are idempotent and that $N_p\Xi_p = \Xi_pN_p = \Xi_p$. ∎