COWLES FOUNDATION FOR RESEARCH IN ECONOMICS

AT YALE UNIVERSITY

Box 2125, Yale Station
New Haven, Connecticut

COWLES FOUNDATION DISCUSSION PAPER NO. 269

COMPARISON OF k-CLASS ESTIMATORS WHEN
THE DISTURBANCES ARE SMALL

Joseph B. Kadane

March 18, 1969

Page 4, 4th line from the bottom: delete "this."

Page 5, line 13, should read

...; Γ is a K x G matrix ...

Page 6, 3rd line from the bottom should read

"and for two-stage least squares,"

Page 7a, line 17 should read

$$\geq \sigma^4 \{(-L^2 + 6L)\ QC_1Q + L\ (2 + \frac{L+2}{T-K+2})\ QC_2Q\}$$

Page 13, line 6: a bar is omitted over P_{z_1}

Page 18, second line from the bottom, should read

$+ EX'uu'N*uu'XQC_1$

Page 20, line 9, should read

$... + E(qu'XQX'uq'QX'uu'X)$

COMPARISON OF k-CLASS ESTIMATORS WHEN
THE DISTURBANCES ARE SMALL

by

Joseph B. Kadane*

1. Introduction

The study of simultaneous equation econometric models has led to
many estimators alternative to ordinary least squares: single-equation
limited information maximum likelihood and two-stage least squares, for
example. However, the behavior of these estimators has been difficult
to describe, and it has been difficult to choose among these estimators.
The work described in this paper explores this problem in the case in
which lagged dependent variables are not permitted.

To be most useful for normative purposes, a description must be
detailed enough to give a good approximation and expose differences
between estimators, and yet be simple enough to strengthen intuition and
yield easily-described comparisons. Since detail and simplicity are in
conflict, approaches may differ in this respect.

One important approach used in the past is large-sample asymptotic
theory. This reveals a persistent bias in ordinary least squares, and a
large-sample asymptotic equivalence between two-stage least squares and
single-equation limited information maximum likelihood. Additionally,
Nagar [12] found the $\frac{1}{T}$ term in the large-sample asymptotic bias and the

$\frac{1}{T}$ and $\frac{1}{T^2}$ terms of the moment matrix of two-stage least squares.

Economists have been uneasy, however, about application of large sample theory to samples which may not be "large" in the relevant sense. Additionally large sample asymptotic results often depend on an assumption about the asymptotic behavior of the moment matrix of exogenous variables which is difficult to justify.

Monte Carlo experiments [17, 13, 14, 7, 8], a second approach to the problem, have only rarely been used to explore the domain of validity of large sample approximations to the mean and variance of an estimator. They have provided some insights into the behavior of estimators under a variety of circumstances, but hypotheses generated by Monte Carlo experiments are difficult to place in a general theory unless they have analytic confirmation of some kind. The parameter space is so large that Monte Carlo results often fail to provide a reasonably comprehensive picture. Corollary 1 below gives an example of a result which is true for small models, like those used in most Monte Carlo work, but which appears to be reversed in large models, which are often encountered in practice.

A third approach, introduced by Basmann [2], finds fixed sample exact densities and moments. More recent work on this line has been done by Basmann [3,4], Bergstrom [5], Kabe [9, 10], Richardson [15], Takeuchi [18] and Sawa [16]. All of these papers have been limited to the case of two endogenous variables in the equation being estimated. An important result of this work has been the finding (in a special case by Basmann [2], in a more general form by Takeuchi [18] and Sawa [16]) that for two-stage least squares, moments of order less than K exist, and those larger do not, where K is the number of exogenous variables in the system. However, the method is difficult, as it involves integrating a noncentral Wishart distribution, and the results for the exact moments and densities have been so complicated as not be be very illuminating.

The paper introduces a fourth approach, based on asymptotic series in a scalar multiple, σ, of the variance of the disturbance in the model.

As $\sigma \to 0$ the regression function is an increasingly good description of the random variables generated. Intuitively this is suggested by Gauss's "Theory of Errors"--the errors were never intended to be so large as to swamp the regression function.

The major results of this paper are the computation of the bias (to order σ^2) and matrices of second moments about the true values (to order σ^4) for all k-class estimators (for fixed k) and for single-equation limited information maximum likelihood. These results are given in section 2, and proved in the appendix. They provide the basis for a number of interesting corollaries.

One corollary is that for equations in which the degree of over-identification is less than or equal to six, two-stage least squares uniformly dominates the limited information maximum likelihood estimators (in the sense that the difference between the moment matrices of these estimators is asymptotically (as $\sigma \to 0$) positive semi-definite regardless of the values taken by parameters or exogenous variables). This is unexpected on the basis of the considerations introduced by Chow [6]. Interpreted in his context, this result means that allowing the data to choose the direction of minimization (limited information maximum likelihood) introduces too much variability into the estimator compared to the benefit gained by fixing an arbitrary direction (two-stage least squares). There are some indications that this preference for two-stage least squares is reversed as the degree of overidentification gets large.

A second corollary shows that for sufficiently small sample sizes and degree of overidentification, ordinary least squares dominate two-stage least squares in the same sense. This had been suspected by econometricians for some time, I believe. Finally, and a surprise, a third corollary shows that the k-class estimator with smallest asymptotic variance occurs when k is negative (of course, this takes no account of the bias, which can be considerable). These are examples of the usefulness of reasonably simple approximations.

The results obtained bear an interesting relationship to each of the three other approaches. Since the sample size, T, is a parameter in small-σ asymptotics, a natural way to compare large-sample and small-σ results is to allow $T \rightarrow \infty$ in the latter. Remarkably, in each case in which large-sample results are available, the limit of the small-σ expression (as $T \rightarrow \infty$) is the large sample asymptotic expression. Thus the results of Nagar [12] are obtained for the special case $k = 1 + \alpha/T$ (for constant α) in the computation of the bias and moment-matrix (theorems 2 and 3). Also the results of Anderson and Rubin [1] are obtained on the distribution of the root, λ, of the determinental equation appearing in the theory of limited information maximum likelihood. (See also [11].) Thus small-σ asymptotics can be thought of as a reasonably good approximation to the behavior of k-class estimators whenever some combination of large sample (i.e. large T) and low phenomenon variability (small σ) ought to lead to reasonably good estimation. Small-σ asymptotics have the important advantage over large-sample theory of being able to "correct" for sample size. Therefore whenever an econometrician is prepared to trust large-sample theory, he should be willing to trust small-σ theory more.

In conclusion, small-σ asymptotics have the following advantages:
(i) They are as simple as, and a generalization of, large-sample theory,
(ii) They can provide definite answers to normative choice of estimator questions.

The ability of small-σ asymptotic expansions to explain Monte-Carlo studies, and the light this sheds on various past Monte-Carlo studies, is to be considered in a separate paper. Whether ultimately this small-σ asymptotics prove to be the best compromise between simplicity and detail remains to be seen. However, it is interesting to note that this approach can be applied to many other econometric and statistical questions [11].

2. Statement of Results

Let the complete system

(1) $$YB + Z\Gamma + \sigma U = 0$$

have a first equation

(2) $$y = Y_1\beta + Z_1\gamma + \sigma u$$

where Y is a $T \times G$ matrix of endogenous variables, partitioned $Y = (y, Y_1, Y_2)$ where y is $T \times 1$, Y_1 is $T \times G_1$ and Y_2 is $T \times G_2$ $(G = G_1 + G_2 + 1)$; Z is a $T \times K$ matrix of exogenous variables, partitioned $Z = (Z_1, Z_2)$ where Z_1 is $T \times K_1$ and Z_2 is $T \times K_2$ $(K = K_1 + K_2)$, Z is assumed to have rank K ; B is a non-singular $G \times G$ matrix of parameters with first column $(-1, \beta', 0')'$ where -1 is a scalar, β is $G_1 \times 1$ and 0 is a $G_2 \times 1$ vector of zeros; Γ is a $k \times G$ matrix of parameters with first column $(\gamma', 0')'$, where γ is $K_1 \times 1$ and 0 is a $K_2 \times 1$ vector of zeros; U is a $T \times G$ matrix of jointly normal residuals with zero means and covariances $Eu_{ti}u_{t'j} = \sigma_{ij}\delta_{tt'}$ and with first column u ; $\sigma_{11} = 1$ and σ is a (small) positive number. The general k- class estimator of $\begin{pmatrix} \beta \\ \gamma \end{pmatrix}$ is

(3) $$\begin{pmatrix} \beta \\ \gamma \end{pmatrix}_k = \begin{bmatrix} Y_1'Y_1 - kV^{*'}V^* & Y_1'Z_1 \\ Z_1'Y_1 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - kV^*)1 \\ Z_1' \end{bmatrix} y$$

where $V^* = \bar{P}_Z Y_1$ and, where for any matrix X, $\bar{P}_X = I - X(X'X)^{-1}X'$ is the projection onto the space orthogonal to the columns of X . As is well-known, the two stage least squares estimate corresponds to $k = 1$, ordinary least squares corresponds to $k = 0$, and limited information (single equation) maximum likelihood corresponds to $k = \lambda$, where

$$(4) \qquad \lambda = \underset{\beta_*}{\text{Min}} \quad \frac{\beta_*' Y_*' \tilde{P}_{Z_1} Y_* \beta_*}{\beta_*' Y_*' \tilde{P}_Z Y_* \beta_*} = \frac{\hat{\beta}_*' Y_*' \tilde{P}_{Z_1} Y_* \hat{\beta}_*}{\hat{\beta}_*' Y_*' \tilde{P}_Z Y_*' \hat{\beta}_*}$$

and $Y_* = (y, Y_1)$.

$\hat{\beta}_*'$ in (4), when normalized, can be written as $(-1, \hat{\beta}_\lambda')$ where $\hat{\beta}_\lambda$ is

limited information maximum likelihood estimator of β.

To write the reduced form of the system, partition B^{-1} con-

formably with $Y = (y, Y_1, Y_2)$ as $B^{-1} = (b, B_1, B_2)$ where b is

$G \times 1$, B_1 is $G \times G_1$ and B_2 is $G \times G_2$. Then

$$y = -Z\Gamma b - \sigma U b$$
$$Y_1 = -Z\Gamma B_1 - \sigma U B_1$$
$$Y_2 = -Z\Gamma B_2 - \sigma U B_2$$

From (5), write

$$[Y_1, Z_1] = [-Z\Gamma B_1, Z_1] + \sigma[-U B_1, 0] = \chi + \sigma V,$$

and let $Q = (\chi'\chi)^{-1}$. Also let the first column of Σ be σ_1. Then,

following Nagar [12], define

$$q = \text{Cov}(V, u) = E(V'u) = \begin{bmatrix} -B_1' \sigma_1 \\ \cdots 0 \cdots \end{bmatrix}, \text{ a } G \times 1 \text{ vector,}$$

$$C_1 = qq'$$

$$C_2 = \begin{pmatrix} B_1' (\Sigma - \sigma_1 \sigma_1') B_1 & 0 \\ 0 & 0 \end{pmatrix}, \text{ a } G \times G \text{ matrix.}$$

Finally, let $r_k = (1-k)T + kK - G_1 - K_1 - 1$, so that,

for ordinary least squares,

$$r_0 = T - G_1 - K_1 - 1$$

for the two-stage least squares

$$r_1 = K - G_1 - K_1 - 1 = K_2 - G_1 - 1 = L - 1$$

where $L = K_2 - G_1$ is the degree of overidentification.

Theorem 1 (asymptotic bias)

(6) $E(e_k) = \sigma^2 r_k Q q + O(\sigma^4)$ for fixed k, and

$E(e_\lambda) = -\sigma^2 Q q + O(\sigma^4)$ for limited-information maximum likelihood,

where

$$e_k = \begin{pmatrix} \beta \\ \gamma \end{pmatrix}_k - \begin{pmatrix} \beta \\ \gamma \end{pmatrix} .$$

Substituting $k = 1 + \alpha/T$ in (6) yields Nagar's [12] bias result as $T \to \infty$. Also notice that

$$k = \frac{T - G_1 - K_1 - 1}{T - K} = 1 + \frac{L-1}{T-K}$$

yields a small-σ asymptotically unbiased estimator. However lack of bias is not a very attractive property for econometric problems, so this is not recommended as an estimator. (The small-σ asymptotic bias of Nagar's "large-sample unbiased" estimator can be evaluated using (6).).

Theorem 2. Let $s_k = k(k-1)(T-K)$. Then

(7) $E(e_k e_k') = \sigma^2 Q + \sigma^4 [\{(1-2r_k) \text{ tr } (C_1 Q) + \text{tr } (C_2 Q)\}Q$

$+ ((r_k-1)^2 + 2s_k + 2) QC_1 Q + (s_k - r_k + 1) QC_2 Q] + O(\sigma^6)$

for fixed k, and, provided $T > K + 2$,

$$E(e_\lambda e_\lambda') = \sigma^2 Q + \sigma^4 \left[\{3 \text{ tr } (C_1 Q) + \text{tr } (C_2 Q)\} Q \right.$$
$$+ 6 QC_1 Q + \left\{ \frac{(L+2)(T-K+L-2)}{T-K-2} \right\} QC_2 Q$$
$$\left. \vphantom{\frac{(L+2)(T-K+L-2)}{T-K-2}} \right]$$
$$+ O(\sigma^5).$$

Again notice that substituting $k = 1 + \alpha/T$ in (7) yields the same expression as Nagar [12] found as $T \to \infty$.

In order to compare moment matrices, a strong criterion is adopted: $A \geq B$ if and only if $A - B$ is non-negative definite (since A and B are loss functions, this implies that B is preferred). The following lemma is useful:

Lemma: If $A \geq 0$, then

$$(\text{tr } A) \, I \geq A .$$

Proof:

Let H be an orthogonal matrix which diagonalizes A. Then

$$e_i \, H\{(\text{tr } A) \, I - A\} \, H'e^0 = \sum_{j=1}^{n} \lambda_j - \lambda_i = \sum_{j \neq i} \lambda_j \geq 0$$

where $\{\lambda_i\}$ are the eigenvalues of A and e_i is a unit vector.

Corollary 1.

$$E(e_\lambda e_\lambda') \geq E(e_1 e_1') \quad \text{if } L \leq 6 .$$

Proof:

$$E(e_\lambda e_\lambda') - E(e_1 e_1') = \sigma^4 \, \{ 2L(\text{tr}(C_1 Q)Q$$

$$+ (4-L)LQ \, C_1 Q + L \, (2 + \frac{L+2}{T-K+2}) \, QC_2 Q \}$$

$$+ 0(\sigma^5)$$

$$\geq \sigma^4 \, (-L^2 + 6L) \, QC_1 Q + L \, (2 + \frac{L+2}{T-K+2}) \, QC_2 Q$$

$$+ 0(\sigma^5)$$

$$\geq o + 0(\sigma^5) \quad \text{if } L \leq 6$$

since

$$QC_1 Q = Q^{1/2}(Q^{1/2} \, C_1 Q^{1/2})Q^{1/2} \leq Q^{1/2}(\text{tr } (Q^{1/2}C_1 Q^{1/2}))Q^{1/2}$$

$$= (\text{tr } (C_1 Q))Q \quad \text{using the Lemma.}$$

This means that the expected mean squared error of any linear combination of parameters is smaller (for $L \leq 6$) estimated by two stage least squares than by limited information maximum likelihood. Also it

suggests that for really large systems the reverse might be true;
Monte Carlo results from small systems, where two-stage least squares
is better, may mislead when applied to large systems.

Corollary 2.

$$E(e_0 e_0') \leq E(e_1 e_1') \quad \text{if} \quad 0 \leq T-K \leq 2(3-L)$$

Proof:

$$
\begin{aligned}
E(e_1 e_1') - E(e_0 e_0') &= \sigma^4 (r_0 - r_1) \{2 \text{ tr } (C_1 Q)Q + (2 - r_0 - r_1)QC_1 Q \\
&\quad + QC_2 Q\} + 0(\sigma^5) \\
&\geq \sigma^4 (r_0 - r_1) \{(4 - r_0 - r_1)QC_1 Q + QC_2 Q\} + 0(\sigma^5) \\
&\geq 0 + 0(\sigma^5)
\end{aligned}
$$

if the above condition is satisfied   QED

Corollary 3 .

$$\frac{d}{dk}\{E(e_k e_k') - (Ee_k)(Ee_k')\} = 2(T-K)\sigma^4 \{(\text{tr } C_1 Q)Q + k(2QC_1 Q + QC_2 Q)\} + 0(\sigma^5)$$

Therefore the minimum variance k-class estimator occurs when k is
negative.

REFERENCES

1. Anderson, T.W. and H. Rubin (1950), "The Asymptotic Properties of Estimation of the Parameters of a Single Equation in a Complete System of Stochastic Equations", Annals of Mathematical Statistics, 21, 570-582.

2. Basmann, R.L. (1961), "A Note on the Exact Finite Sample Frequency Functions of Generalized Classical Linear Estimators in Two Leading Overidentified Cases", Journal of the American Statistical Association, 56, 619-636.

3. Basmann, R.L. (1963a), "A Note on the Exact Finite Sample Frequency Functions of Gereralized Classical Linear Estimators in a Leading Three Equation Case", Journal of the American Statistical Association, 58, 161-171.

4. Basmann, R.L. (1963b), "Remarks Concerning the Application of Exact Finite Sample Distribution Functions of GCL Estimators in Econometric Statistical Inference", Journal of the American Statistical Association, 58, 943-976.

5. Bergstrom, A.R. (1962), "The Exact Sampling Distributions of Least Squares and Maximum Likelihood Estimators of the Marginal Propensity to Consume", Econometrica, 30, 480-490.

6. Chow, G.C. (1964), "A Comparison of Alternative Estimators for Simultaneous Equations", Econometrica, 32, 532-553.

7. Cragg, J.G. (1964), "Small-Sample Properties of Various Simultaneous Equation Estimators: The Results of Some Monte-Carlo Experiments", (Princeton University, Econometric Research Program, Research Memorandum No. 68).

8. Cragg, J.G. (1967), "On the Relative Small-Sample Properties of Several Structural-Equation Estimators, Econometrica, 35, 89-110.

9. Kabe, D.G. (1963), "A Note on the Exact Distributions of the GCL Estimators in Two Leading Over Identified Cases", Journal of the American Statistical Association, 58, 535-537.

10. Kabe, D.G. (1964), "A Note on the Exact Distributions of the GCL Estimators in a Leading Three-Equation Case", Journal of the American Statistical Association, 59, 881-894.

11. Kadane, J.B. (1968), "Testing Overidentifying Restrictions When the Disturbances are Small", submitted for publication.

12. Nagar, A.L. (1959), "The Bias and Moment Matrix of the General k-Class Estimators of the Parameters in Simultaneous Equations", Econometrica, 27, 575-595

13. Quandt, R. (1962), "Some Small Sample Properties of Certain Sturctural Equation Estimators", (Princeton University, Econometric Research Program, Research Memorandum No. 48).

14. Quandt, R. (1965), "On Certain Small Sample Properties of k-Class Estimators", International Economic Review, 6, 92-104.

15. Richardson, D.H. (1966), "An Investigation of the Sampling Distribution of Estimators and Test Statistics Associated with a Simultaneous Equations Model", (unpublished dissertation, Purdue University).

16. Sawa, T. (1968), "The Exact Sampling Distribution of Ordinary Least Squares and Two Stage Least Squares", mimeographed.

17. Summers, R.L. (1965), "A Capital Intensive Approach to the Samll Sample Properties of Various Simultaneous Equation Estimators, Econometrica, 33, 1-41.

18. Takeuchi, K. (1967), "On the Exact Moment of the OLS, the TSLS, and the Instrumental Variables Method Estimators", presented at the second Far Eastern Meeting of the Econometric Society, mimeographed.

Appendix A: Proof of Theorems

Throughout these proofs it is important to distinguish the cases when k is fixed from those when it may be random. When k can be either, it is denoted k^* ; when fixed, it is denoted k .

Some additional notation is useful here: for any matrix X , $P_X = X(X'X)^{-1}X'$ is the projection onto the space spanned by X .

$\bar{P}_X = I - P_X$, which was introduced in Section 2 just under equation (3), is the projection onto the space orthogonal to the space spanned by X. Finally, let

$$S = V'X + X'V .$$

Lemma 1. If $k^* = 0_p(1)$, then

$$e_{k*} = \sigma QX'u + \sigma^2 Q(V'N*u - SQ_X'u)$$

$$+ \sigma^3 Q(SQSQX'u - V'N*VQX'u - SQV'N*u) + 0_p(\sigma^4)$$

where $N^* = I - k^*\bar{P}_Z$.

Proof:

$$e_{k*} = \begin{bmatrix} Y_1'Y_1 - k^*V^{*'}V^* & Y_1'Z_1 \\ Z_1'Y_1 & Z_1'Z_1 \end{bmatrix}^{-1} \begin{bmatrix} (Y_1 - k^*V^*)' \\ Z_1' \end{bmatrix} \sigma u$$

$$= \sigma\{(X'X) + \sigma S + \sigma^2 V'N*V\}^{-1}(X' + \sigma V'N*)u$$

$$= \sigma\{I + Q(\sigma S + \sigma^2 V'N*V)\}^{-1}Q(X' + \sigma V'N*)u.$$

Use of the standard geometric expansion for the inverse of a matrix:

$$(I + h\Delta)^{-1} = I - h\Delta + h^2\Delta^2 - h^3\Delta^3 + \ldots$$

yields

$$e_{k*} = \sigma Q(X' + \sigma V'N*)u - \sigma Q[\sigma S + \sigma^2 V'N*V]Q(X' + \sigma V'N*)u$$

$$+ \sigma Q(\sigma S + \sigma^2 V'N*V)Q(\sigma S + \sigma^2 V'N*V)Q(X' + \sigma V'N*)u + \ldots$$

$$= \sigma QX'u + \sigma^2 Q(V'N*u - SQX'u) + \sigma^3 Q(SQSQX'u$$

$$- V'N*VQX'u - SQV'N*u) + 0_p(\sigma^4) \qquad\qquad \text{QED}$$

Lemma 2. $\lambda = 0_p(1)$

Proof: See Lemma 1 of Kadane [11].

Lemma 3. $\lambda = \dfrac{u'\bar{\bar{P}}_\chi u}{u'P_Z u} + 2\sigma \left[\dfrac{(u'\chi QV'\bar{P}_Z u)(u'\bar{P}_\chi u) - (u'\chi QV'\bar{\bar{P}}_\chi u)(u'\bar{P}_Z u)}{(u'\bar{P}_Z u)^2} \right]$

$\qquad\qquad + 0_p(\sigma^2)$.

(The first order term of Lemma 3 is derived for models possibly including lagged endogenous variables, and is discussed in relation to the previous literature in [11].)

Proof.

Let χ^* be the first G_1 rows of Q and χ^{**} the last K_1 rows, so that

$$Q = \begin{bmatrix} \chi^* \\ \chi^{**} \end{bmatrix}.$$

Then

$$\bar{P}_{Z_1} Y_* \hat{\beta}_* = \bar{P}_{Z_1} (y, Y_1) \left[\left(\begin{pmatrix} -1 \\ \beta \end{pmatrix} + \sigma \begin{pmatrix} 0 \\ \chi^* \end{pmatrix} \right) \chi'u + \sigma^2 \begin{pmatrix} 0 \\ \chi^* \end{pmatrix} (V'N*u - SQ\chi'u) + 0_p(\sigma^3) \right]$$

$$= \bar{P}_{Z_1} [-y + Y_1\beta + \sigma Y_1\chi^*\chi'u + \sigma^2 Y_1\chi^* (V'N*u - SQ\chi'u) + 0_p(\sigma^3)]$$

$$= \bar{P}_{Z_1} [-Z_1\gamma - \sigma u + \sigma Z_1\chi^{**}\chi'u + \sigma Y_1\chi^*\chi'u + \sigma^2 Y_1\chi^*(V'N*u - SQ\chi'u) + \sigma^2 Z_1\chi^{**}$$

$$(V'N*u - SQ\chi'u) + 0_p(\sigma^3)]$$

$$= \bar{\bar{P}}_{Z_1}\left[-\sigma\bar{\bar{P}}_\chi u + \sigma^2 \ \bar{\bar{P}}_\chi VQ\chi'u + \sigma^2 \ \chi QV' \ (N^*-P_\chi)u + 0_p(\sigma^3)\right]$$

$$= -\sigma\bar{\bar{P}}_\chi u + \sigma^2 \ \bar{\bar{P}}_\chi VQ\chi'u + \sigma^2\bar{\bar{P}}_{Z_1} \chi QV'(N^*-P_\chi)u + 0_p(\sigma^3)$$

Writing

$$\lambda = \frac{N}{D} = \frac{\sigma^2 N_2 + \sigma^3 N_3 + \ldots}{\sigma^2 D_2 + \sigma^3 D_3 + \ldots}$$

$$= \frac{N_2}{D_2} + \sigma\left[\frac{N_3 D_2 - D_3 N_2}{D_2^2}\right] + 0_p(\sigma^2) \ ,$$

$$N = \hat{\beta}_*' \ Y_*' \ \bar{\bar{P}}_{Z_1} Y_* \hat{\beta}_* = (\hat{\beta}_*' \ Y_*' P_{Z_1})(\bar{\bar{P}}_{Z_1} Y_* \hat{\beta}_*)$$

$$= \{-\sigma u'\bar{\bar{P}}_\chi + \sigma^2 u'\chi QV'\bar{\bar{P}}_\chi + \sigma^2 u'(N^*-P_\chi) \ VQ\chi' + 0_p(\sigma^3)\}$$

$$\{-\sigma\bar{\bar{P}}_\chi u + \sigma^2 \ \bar{\bar{P}}_\chi VQ\chi'u + \sigma^2\bar{\bar{P}}_{Z_1} \chi QV'(N^*-P_\chi) \ u + 0_p(\sigma^3)\}$$

$$= \sigma^2 \ u'\bar{\bar{P}}_\chi u - 2\sigma^3 \ u'\bar{\bar{P}}_\chi VQ\chi'u + 0_p(\sigma^4)$$

Therefore $N_2 = u'\bar{\bar{P}}_\chi u$ and $N_3 = -2u' \ \bar{\bar{P}}_\chi VQ\chi'u$

Similarly $D = \hat{\beta}_*' \ Y_*' \ \bar{\bar{P}}_Z Y_* \hat{\beta}_* = (\hat{\beta}_*'Y_*'\bar{\bar{P}}_{Z_1}) \ \bar{\bar{P}}_Z(\bar{\bar{P}}_{Z_1} Y_* \hat{\beta}_*)$

$$= \{-\sigma u'\bar{\bar{P}}_\chi + \sigma^2 u'\chi QV' \ \bar{\bar{P}}_\chi + \sigma^2 u' \ (N^*-P_\chi) \ VQ\chi'\bar{\bar{P}}_{Z_1} + 0_p(\sigma^3)\}\bar{\bar{P}}_Z$$

$$\{-\sigma\bar{\bar{P}}_\chi u + \sigma^2\bar{\bar{P}}_\chi VQ\chi'u + \sigma^2\bar{\bar{P}}_{Z_1} \chi QV'(N^*-P_\chi) \ u + 0_p(\sigma^3)\}$$

$$= \sigma^2 u'\bar{\bar{P}}_Z u - 2\sigma^3 \ u' \ \bar{\bar{P}}_Z VQ\chi'u + 0_p(\sigma^4)$$

Hence $D_2 = u'\ddot{P}_Z u$ and $D_3 = -\sigma u'\ddot{P}_Z VQX'u$, and

$$\lambda = \frac{u'\ddot{P}_X u}{u'\ddot{P}_Z u} + 2\sigma \left[\frac{(u'\chi QV'\ddot{P}_Z u)\,(u'\ddot{P}_X u) - (u'\chi QV'\ddot{P}_X u)\,(u'\ddot{P}_Z u)}{(u'\ddot{P}_Z u)^2} \right] + 0_p(\sigma^2)$$

<div align="center">QED</div>

The following numbers are useful in the proofs of theorems 1 and 2:

$$a = \text{tr } \ddot{P}_Z = T - K$$

$$b = \text{tr } (\ddot{P}_X - \ddot{P}_Z) = \text{tr } (P_Z - P_X) = K_2 - G_1 = L$$

$$c = \text{tr } P_X = K_1 + G_1 = T - a - b \ .$$

Also the notation $[....]^*$ is used to denote a single expression, valid for both $k^* = k$ and $k^* = \lambda$, the upper part referring to k , the lower part to λ . For example

$$N^* = I - k^*\ddot{P}_Z = I - \begin{bmatrix} k \\ \lambda \end{bmatrix}^* \ddot{P}_Z = \begin{bmatrix} (1-k)I + kP_Z \\ \\ I - \lambda\ddot{P}_Z \end{bmatrix}^*$$

The lemmas on expectation in Appendix B are used without special comment in the proofs to follow.

Proof of Theorem 1.

Recalling Lemma 1,

(A1) $E(e_{k^*}) = \sigma EQX'u + \sigma^2 E(QV'N^*u - Q(V'X + X'V)QX'u) + 0(\sigma^3)$

(A2) $E(QX'u) = QX'E(u) = 0 \ ,$

Let $W' = V' - qu'$. W' is obviously independent of u' , by

construction.   Then

(A3) $E(QV'N*u) = E(QV'(I - k*\bar{\bar{P}}_Z)u)$

$= QE(W'(I - k*\bar{\bar{P}}_Z)u) + QqEu'(I - k*\bar{\bar{P}}_Z)u$

$= Qq(T - E(k*u'\bar{\bar{P}}_Z u))$

$$= Qq\left(T - \begin{bmatrix} ka \\ a+b+ 0_{(\sigma)} \end{bmatrix}^*\right) .$$

(A4) $E(QV'XQX'u) = QE(W'XQX'u + qu'XQX'u) = Qq \text{ tr } P_X = cQq$

(A5) $E(QX'VQX'u) = QX'E(WQX'u + uq'QX'u) = QX'Euu'XQq = Qq$

Substitution of (A2), (A3), (A4) and (A5) in (A1) yields Theorem 1 .

To prove Theorem 2, again return to Lemma 1.

(A6) $Q^{-1}e_{k*}e_{k*}'Q^{-1} = B_1 + B_2 + B_2' + B_3 + B_4 + B_4' + 0_p(\sigma^5)$

where

$B_1 = \sigma^2(X'uu'X)$

$B_2 = \sigma^3(X'u(u'N*V - u'XQS))$

(A7) $B_3 = \sigma^4(V'N*u - SQX'u)(u'N*V - u'XQS)$

$B_4 = \sigma^4(SQSQX'u - V'N*VQX'u - SQV'N*u)(u'X)$

The expectation of each of the above terms is evaluated below.

(A8) $E(B_1) = \sigma^2 X'Euu'X = \sigma^2 X'X = \sigma^2 Q^{-1}$.

For B_2, since expectations of products with an odd number of factors of any normal / variables with zero mean are zero, the only possible non-zero contribution may come from $N*$ when $k* = \lambda$.

(A9) $EX'uu'[I - \lambda\bar{\bar{P}}_Z]V = -E\lambda X'uu'\bar{\bar{P}}_Z V$

Because the computation of $e_{k*}e_{k*}'$ is sought in this proof to order σ^4, both σ^0 and σ^1 order terms of λ are relevant here. Hence

$$(A10) \quad -E\lambda X'uu'\bar{P}_Z V = -E\frac{u'\bar{P}_X u}{u'\bar{P}_Z u}X'uu'\bar{P}_Z V$$

$$+ 2\sigma E\left\{\left[\frac{(u'XQV'\bar{P}_X u)(u'\bar{P}_Z u) - (u'XQV'\bar{P}_Z u)(u'\bar{P}_X u)}{(u'\bar{P}_Z u)^2}\right]X'uu'\bar{P}_Z V\right\}$$

$$+ 0(\sigma^2)$$

Again using the independence of $W = V - uq'$ and u,

$$(A11) \quad E\left\{\frac{u'\bar{P}_X u}{u'\bar{P}_Z u}X'uu'\bar{P}_Z V\right\} = E\left\{\frac{u'\bar{P}_X u}{u'\bar{P}_Z u}X'uu'\bar{P}_Z uq'\right\} = Eu'\bar{P}_X uX'uq' = 0$$

$$2\sigma X'Eu\left[\frac{(u'XQV'\bar{P}_X u)(u'\bar{P}_Z u) - (u'XQV'\bar{P}_Z u)(u'\bar{P}_X u)}{(u'\bar{P}_Z u)^2}\right]u'\bar{P}_Z V =$$

$$2\sigma X'Eu\left[\frac{(u'XQV'\bar{P}_X u)(u'\bar{P}_Z u) - (u'XQV'\bar{P}_Z u)(u'\bar{P}_X u)}{(u'\bar{P}_Z u)^2}\right]u'\bar{P}_Z uq'$$

$$+ 2\sigma X'Eu\left[\frac{(u'XQV'\bar{P}_X u)(u'\bar{P}_Z u) - (u'XQV'\bar{P}_Z u)(u'\bar{P}_X u)}{(u'\bar{P}_Z u)^2}\right]u'\bar{P}_Z W$$

$$= 2\sigma X'Eu\left[\frac{u'XQqu'\bar{P}_X uu'\bar{P}_Z u - u'XQqu'\bar{P}_Z uu'\bar{P}_X u}{(u'\bar{P}_Z u)}\right]q'$$

$$+ 2\sigma X'Eu\left[\frac{(u'XQW'\bar{P}_X u)(u'\bar{P}_Z u) - (u'XQW'\bar{P}_Z u)(u'\bar{P}_X u)}{(u'\bar{P}_Z u)^2}\right]u'\bar{P}_Z W$$

$$= 2\sigma X'Eu \left[\frac{(u'\bar{P}_Z u)u'XQW'\bar{P}_X u - (u'\bar{P}_X u \ \ u'XQW'\bar{P}_Z u)}{(u'\bar{P}_Z u)^2} \right] u'\bar{P}_Z W$$

$$= 2\sigma X'Eu \left[\frac{(u'\bar{P}_Z u)u'XQ \ tr \ (\bar{P}_X uu'\bar{P}_Z) - (u'\bar{P}_X u)(u'XQ)tr \ (\bar{P}_Z uu'\bar{P}_Z)}{(u'\bar{P}_Z u)^2} \right] C_2$$

$$= 2\sigma X'Eu \left[u'XQ - \frac{(u'\bar{P}_X u)u'XQ}{u'\bar{P}_Z u} \right] C_2$$

$$= -2\sigma X'Eu \left[\frac{u'(\bar{P}_X - \bar{P}_Z)u}{u'\bar{P}_Z u} u'XQC_2 \right]$$

$$= -2\sigma X' \left[\frac{b+2}{a-2}(\bar{P}_X - \bar{P}_Z) + \frac{b}{a}\bar{P}_Z + \frac{b}{a-2}P_X \right] XQC_2 = -2\sigma \frac{b}{a-2} C_2$$

(A12)
$$EB_2 = \sigma^4 \begin{bmatrix} 0 \\ -\frac{2b}{a-2} \end{bmatrix}^* C_2 + 0(\sigma^5) ,$$

using (A9), (A10) and (A11).

Moving now to B_3 ,

(A13)
$$E(V'N*u - SQX'u)(u'N*V - u'XQS =$$

$$E\{V'N*uu'N*V\} - E\{V'N*uu'XQS\} - E\{SQX'uu'N*V\}$$

$$+ E\{SQX'uu'XQS\} .$$

(A14)
$$E\{V'N*uu'N*V\} = E\{W'N*uu'N*W\} + E\{qu'N*uu'N*uq'\}$$

$$= E(u'(N*)^2 u) \ C_2 + E(u'N*u)^2 \ C_1$$

$$u'(N*)^2 u = u'[I - k*\bar{P}_Z][I - k*\bar{P}_Z]u$$

$$= u'[(P_Z) + (k* - 1)^2 \bar{P}_Z]u$$

$$E\{u'N^2u\} = T - a + (k - 1)^2 a$$

$$E\{u'[P_Z + (\lambda - 1)^2 \bar{\bar{P}}_Z]u = T - a + E\frac{(u'(\bar{P}_X - \bar{\bar{P}}_Z)u)^2}{u'\bar{\bar{P}}_Z u} + 0(\sigma)$$

$$= T - a + \frac{b(b + 2)}{a - 2} + 0(\sigma)$$

$$u'N*u = u'[P_Z + (1 - k*)\bar{\bar{P}}_Z]u = u'P_Z u + (1 - k*)u'\bar{\bar{P}}_Z u$$

$$E(u'Nu)^2 = (T - a)(T - a + 2) + 2(1 - k)(T - a)a$$
$$+ (1 - k)^2 a(a + 2)$$

$$E(u'(I - \lambda\bar{\bar{P}}_Z)u)^2 = E(u'P_X u)^2 + 0(\sigma) = c(c + 2) + 0(\sigma)$$

Summarizing, from (A14),

(A15)

$$E\{V'N*uu'N*V\} = \begin{bmatrix} (T - a)(T - a + 2) + 2(1 - k)(T - a)a + (1 - k)^2 a(a + 2) \\ \\ c(c + 2) + 0(\sigma) \end{bmatrix}^* C_1$$

$$+ \begin{bmatrix} T - a + (k - 1)^2 a \\ \\ T - a + \frac{b(b + 2)}{a - 2} + 0(\sigma) \end{bmatrix}^* C_2$$

Next, from (A13),

(A16) $E\{SQX'uu'N*V\} = E\{(V'X + X'V)QX'uu'N*V\}$

$= E\{(W'X + X'W)QX'uu'N*W\} + E\{(qu'X + X'uq')QX'uu'N*uq'\}$

$= E\{u'N*XQX'u\}C_2 + E\{X'N*uu'XQ\}C_2 + Eu'[P_X]uu'N*uC_1$

$+ [X'uu'N*uu'XQC_1$

(since $N*X = X$)

$$= (c + 1) C_2 + \begin{bmatrix} (T + 2 - ka)c \\ c(c + 2) + 0(\sigma) \end{bmatrix}^* C_1 + \begin{bmatrix} T + 2 - ka \\ c + 2 + 0 (\sigma) \end{bmatrix}^* C_1$$

$$= (c + 1) C_2 + (c + 1) \begin{bmatrix} T + 2 - ka \\ c + 2 + 0(\sigma) \end{bmatrix}^* C_1$$

Finally, the last term from (A13),

(A17) $ESQX'uu'XQS = E(V'X + X'V)QX'uu'XQ(V'X + X'V)$

$$= E(W'X + X'W)QX'uu'XQ(W'X + X'W)$$

$$+ E\{(qu'X + X'uq')QX'uu'XQ(qu'X + X'uq')\}$$

$$= C_2[QX'IXQX']X + X'(XQX'IXQ)C_2$$

$$+ tr\ (XQX'IXQX')C_2 + X'(tr\ (QX'IXQC_2)X$$

$$+ E\{qu'XQX'uu'XQX'uq'\} + E\{X'uq'QX'_{uu}'XQqu'X\}$$

$$+ EX'uq'QX'uu'XQX'uq' + Equ'XQX'uu'XQqu'X$$

$$= 2C_2 + tr\ (P_X)C_2 + tr\ (QC_2)Q^{-1} + E(u'P_Xu)^2C_1$$

$$+ X'E(uu'XQC_1QX'uu')X + X'Euu'P_Xuu'XQC_1$$

$$+\quad C_1QX'uu'P_Xuu'X$$

$$= (c + 2)C_2 + tr\ (QC_2)Q^{-1} + c(c + 2)C_1 + tr(QC_1)Q^{-1}$$

$$+ 2(c + 2)C_1 + 2C_1$$

$$= (c + 2)C_2 + tr\ (QC_2)Q^{-1} + ((c + 2)^2 + 2)C_1 + tr\ (QC_1)Q^{-1}$$

Returning to (A13) and using (A15), (A16) and (A17),

$$E(B_3) = \sigma^4 \left\{ \begin{bmatrix} [b + (1 - k)a]^2 + 2a(k - 1)k + 2 \\ 2 + 0(\sigma) \end{bmatrix}^* C_1 + tr\ (QC_1)Q^{-1} \right.$$

(A18)
$$+\left(b+\left[\begin{array}{c}(k-1)^2a\\[2mm]\dfrac{b(b+2)}{a-2}+0(\sigma)\end{array}\right]^*\right)C_2+(\text{tr }QC_2)Q^{-1}\Bigg\}\circ$$

Finally, B_4 must be computed.

(A19) $E(SQSQX'u - V'N*VQX'u - SQV'N*u)(u'X)$

$$= E(SQSQX'uu'X) - E(V'N*VQX'uu'X) - E(SQV'N*uu'X)$$

(A20) $E(SQSQX'uu'X) = E(V'X + X'V)Q(V'X + X'V)QX'uu'X$

$$= E(W'X + X'W)Q(W'X + X'W)QX'uu'X$$

$$+ E(qu'X + X'uq')Q(qu'X + X'uq')QX'uu'X$$

$$= E(W'XQW'X) + E(W'XQX'W) + E(X'WQW'X) + E(X'WQX'W)$$

$$+ E(qu'XQqu'XQX'uu'X) + E(qu'XQX'uq'XQ'uu'X)$$

$$+ E(X'uq'Qqu'XQX'uu'X) + E(X'uq'QX'uq'QX'uu'X)$$

$$= C_2QX'X + \text{tr }(P_X)C_2 + \text{tr }(C_2Q)Q^{-1} + X'XQC_2$$

$$+ E(qq'QX'uu'P_Xuu'X) + Eqq'QX'uu'P_Xuu'X$$

$$+ E(\text{tr }(C_1Q)X'uu'P_Xuu'X) + E(X'uu'XQqq'QX'uu'X)$$

$$= (c + 2)C_2 + \text{tr }(C_2Q)Q^{-1} + 2(c + 2)C_1 + (c + 2)\text{ tr }(C_1Q)Q^{-1}$$

$$+ \text{tr }(C_1Q)Q^{-1} + 2C_1$$

$$= (c + 2)C_2 + \text{tr }(C_2Q)Q^{-1} + 2(c + 3)C_1 + (c + 3)\text{ tr }(C_1Q)Q^{-1}$$

$$E\{V'N*VQX'uu'X\} = E\{W'N*WQX'uu'X\} + E\{qu'N*uq'QX'uu'X\}$$

$$= E(\text{tr }N*)C_2QX'uu'X + Eqq'QX'uu'N*uu'X$$

$$= E\left[\begin{array}{c}T - ka\\[2mm]T - \dfrac{u'\bar{P}_Xu}{u'\bar{P}_Zu}a + 0(\sigma)\end{array}\right]^*C_2QX'uu'X + \left[\begin{array}{c}T + 2 - ka\\[2mm]c + 2 + 0(\sigma)\end{array}\right]^*C_1$$

Now

$$EC_2QX'u\left(T - \frac{u'\bar{P}_X u}{u'\bar{P}_Z u}a\right)u'X = TC_2 - aEC_2QX'uu'X - EaC_2QX'u\left(\frac{u'(\bar{P}_X - \bar{P}_Z)}{u'\bar{P}_Z u}\right)u'X$$

$$= (T - a)C_2 - \frac{abC_2}{a-2} = \left(T - a - \frac{ab}{a-2}\right)C_2 \ .$$

Then

$$(A21) \quad E\{V'N*VQX'uu'X\} = \begin{bmatrix} T + 2 - ka \\ c + 2 + 0(\sigma) \end{bmatrix}^* C_1 + \begin{bmatrix} T - ka \\ T - a - \frac{ab}{a-2} + 0(\sigma) \end{bmatrix}^* C_2$$

Finally,

$$(A22) \quad E(SQV'N*uu'X) = E(V'X + X'V)QV'N*uu'X$$

$$= E(W'X + X'W)QW'N*uu'X + Equ'XQqu'N*uu'X$$

$$+ EX'uq'Qqu'N*uu'X$$

$$= EC_2QX'N*uu'X + E(tr\ C_2Q)X'N*uu'X$$

$$+ C_1QX'Euu'N*uu'X + tr\ (C_1Q)X'Euu'N*uu'X$$

$$= C_2 + tr\ (C_2Q)Q^{-1} + \begin{bmatrix} T + 2 - ka \\ c + 2 + 0(\sigma) \end{bmatrix}^* C_1 + \begin{bmatrix} T + 2 - ka \\ c + 2 + 0(\sigma) \end{bmatrix}^* tr(C_1Q)Q^{-1}$$

$$(A23)\ EB_4 = \sigma^4\left\{\begin{bmatrix} 2\{a(k-1) - b + 1\} \\ 2 + 0(\sigma) \end{bmatrix}^* C_1 + \begin{bmatrix} (k-1)a - b + 1 \\ -b + \frac{ab}{a-2} + 1 + 0(\sigma) \end{bmatrix}^* C_2 \right.$$

$$\left. + \begin{bmatrix} a(k-1) - b + 1 \\ 1 + 0(\sigma) \end{bmatrix}^* (tr\ C_1Q)Q^{-1}\right\}$$

Finally, Theorem 2 follows using (A6), (A8), (A12), (A18) and (A23).

Appendix B: Certain Expectations

<u>Lemma B1:</u> Let A be a T x T constant matrix. Then

$$E(W'AW) = (tr\ A)C_2 \ .$$

<u>Proof:</u>

$$(EW'AW)_{ij} = E \sum_{t,t'} w_{it} a_{tt'} w_{t'j} = E \sum w_{it} w_{tj} a_{tt}$$

$$= (C_2)_{ij} \sum_t a_{tt} = (tr\ A)(C_2)_{ij} \ .$$

<u>Lemma B2.</u> Let A be a c x c constant matrix. Then

$$E(WAW') = tr\ (C_2 A)I \ .$$

<u>Proof:</u>

$$[EWAW']_{tt'} = E \sum_{j,j'} w_{tj} a_{jj'} w_{j't'} = E\sum (C_2)_{jj'} \delta_{tt'} a_{jj'}$$

$$= tr\ (C_2 A)\ \delta_{tt'} \ .$$

<u>Lemma B3:</u> Let A be a c x T constant matrix. Then

$$E(WAW) = A'C_2 \ .$$

<u>Proof:</u>

$$E(WAW)_{tj} = E\ w_{tk} A_{kt'} w_{t'j} = A_{kt} \sum w_{tk} w_{tj} = (A'C_2)_{tj} \ .$$

<u>Lemma B4:</u> Let P_1 and P_2 be projections such that $P_1 P_2 = 0$. Then

$$E\ u\ \frac{u'P_1 u}{u'P_2 u}\ u' = \left(\frac{tr\ P_1 + 2}{tr\ P_2 - 2}\right)P_1 + \left(\frac{tr\ P_1}{tr\ P_2}\right)P_2 + \left(\frac{tr\ P_1}{tr\ P_2 - 2}\right)[I - P_1 - P_2] \ .$$

<u>Proof:</u> Since P_1 and P_2 commute, let Γ (orthogonal) simultaneously

diagonalize them.

$$\Gamma P_1 \Gamma' = \begin{bmatrix} I & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = I_1 \qquad \Gamma P_2 \Gamma' = \begin{bmatrix} 0 & 0 & 0 \\ 0 & I & 0 \\ 0 & 0 & 0 \end{bmatrix} = I_2 \ .$$

Also let $v' = u'\Gamma$, and $v = (v_1, v_2, v_3)$ partitioned conformably with I_1 and I_2 . Then

$$E\, u\, \frac{u'P_1 u}{u'P_2 u}\, u' = \Gamma\, E\, \left[v\, \frac{v_1'v_1}{v_2'v_2}\, v'\right]\, \Gamma'$$

$$= \Gamma\left[\frac{\mathrm{tr}\ P_1 + 2}{\mathrm{tr}\ P_2 - 2}\, I_1 + \frac{\mathrm{tr}\ P_1}{\mathrm{tr}\ P_2}\, I_2 + \frac{\mathrm{tr}\ P_1}{\mathrm{tr}\ P_2 - 2}\, (I - (I_1 + I_2))\right]\, \Gamma'$$

$$= \frac{\mathrm{tr}\ P_1 + 2}{\mathrm{tr}\ P_2 - 2}\, P_1 + \frac{\mathrm{tr}\ P_1}{\mathrm{tr}\ P_2}\, P_2 + \frac{\mathrm{tr}\ P_1}{\mathrm{tr}\ P_2 - 2}\, (I - P_1 - P_2) \quad \text{QED .}$$

In particular,

$$E\, u\, \frac{u'(\bar{P}_X - \bar{P}_Z)u}{u'\bar{P}_Z u}\, u' = \frac{b+2}{a-2}(\bar{P}_X - \bar{P}_Z) + \frac{b}{a}\, \bar{P}_Z + \frac{b}{a-2}\, P_X .$$

<u>Lemma B6</u>: Let R be a symmetric, constant matrix

$$E\, u\, u'Ru\, u' = (\mathrm{tr}\ R)\, I + 2R .$$

<u>Proof</u>:

Diagonalize R :

$$\Gamma R \Gamma' = D_\lambda \qquad v = \Gamma u \qquad v \sim N(0, I)$$

$$E\, u\, u'Ru\, u' = \Gamma'E(v\, v'D_\lambda v\, v')\Gamma$$

$$= \Gamma'(E\, v\, \Sigma\lambda_i\, v_i^2\, v')\Gamma$$

has the jth diagonal element

$$E(\Sigma\lambda_i v_i^2)v_j^2 = E\{\sum_{i\neq j} \lambda_i v_i^2 v_j^2 + \lambda_j v_j^4\}$$

$$= \sum_{i\neq j} \lambda_i + 3\lambda_j = \sum_i \lambda_i + 2\lambda_j \qquad \text{(i.e. (tr } (R)I + 2D_\lambda).$$

So the whole expectation is $(\mathrm{tr}\ R)I + 2R$. \qquad QED.

<u>Corollary</u>. Let R_1 be a symmetric, constant matrix and R_2 a constant matrix, then

$$E \; u' \; R_1 u \; u' R_2 u = (tr \; R_1)(tr \; R_2) + 2 \; tr \; (R_1 R_2) \; .$$

<u>Proof</u>:

$$E \; u'R_1 u \; u'R_2 u = E\{tr \; u'R_1 u \; u'R_2 u\}$$

$$= E \; tr \; u \; u'R_1 u \; u'R_2 = tr \; (\{E \; u \; u'R_1 \; u \; u'\}R_2)$$

$$= tr \; (((tr \; R_1)I + 2R_1)R_2) = tr \; R_1 \; tr \; R_2 + 2 \; tr \; R_1 R_2 \; . \quad \text{QED.}$$

$$E \; u \; u'N^*u \; u' = \begin{bmatrix} (T + 2 - ka)I - 2k\bar{\bar{P}}_Z \\[1em] (c + 2)I - 2\bar{P}_X + 0(\sigma) \end{bmatrix}^*$$

$$E \; u'Ru \; u'N^*u = \begin{bmatrix} (T + 2 - ka) \; tr \; R - 2k \; tr(R\bar{\bar{P}}_Z) \\[1em] (c + 2) \; tr \; R - 2 \; tr \; (R\bar{\bar{P}}_X) + 0(\sigma) \end{bmatrix}^* \quad .$$