Technical Report 86-1

Correction Factors For F Ratios in Nonlinear Regression

by

David Hamilton & Douglas Wiens

CORRECTION FACTORS FOR F RATIOS IN NONLINEAR REGRESSION

DAVID HAMILTON and DOUGLAS WIENS
Department of Mathematics, Statistics
and Computing Science
Dalhousie University
Halifax, Nova Scotia

Summary

Multiplicative correction factors are derived for the limiting F distributions of several test statistics for parameter subsets in nonlinear regression. The factors depend on the first and second derivatives of the model and are related to measures of intrinsic nonlinearity. An example is given for which the correction is substantial. A similar factor is obtained for the lack of fit test in nonlinear regression.

Some key words: intrinsic nonlinearity; lack of fit; limiting distribution; nonlinear regression; parameter subsets

1. Introduction

Recently, Hamilton (1986) discussed several approaches to hypothesis testing and confidence region construction for parameter subsets in nonlinear regression models. Accurate approximations were obtained using a geometric approach and first and second derivatives of the expectation function. The limiting statistical properties of the various approaches are examined in this paper, and correction factors are derived for the commonly used, but approximate F distribution. An example shows that these factors can have a large effect. A similar correction factor is obtained for the lack of fit test.

The nonlinear regression model is $\mathbf{y} = \eta(\theta) + \epsilon$, where y, $\eta(\theta)$ and ϵ are n dimensional vectors respectively containing the observed responses, their expectations and a spherically normal error with zero mean and variance $\sigma^2\mathbf{I}$. The expectation vector depends nonlinearly on the p parameters $\theta^T = (\theta_1^T, \theta_2^T)$, and also on some explanatory variables. In this article, the null hypothesis of interest is that the first p_1 parameters, θ_1, equal specified values $\theta_{1,0}$. Hamilton (1986) discusses four statistics used for this test and its associated confidence region:

$$F_1 = \frac{(|\tilde{\mathbf{e}}|^2 - |\hat{\mathbf{e}}|^2)/p_1}{|\hat{\mathbf{e}}|^2/(n-p)}$$

$$F_2 = \frac{|(\tilde{\mathbf{P}} - \tilde{\mathbf{P}}_2)\tilde{\mathbf{e}}|^2/p_1}{|(\mathbf{I} - \tilde{\mathbf{P}})\tilde{\mathbf{e}}|^2/(n-p)}$$

$$F_3 = \frac{|(\dot{\mathbf{P}} - \dot{\mathbf{P}}_2)\dot{\mathbf{e}}|^2/p_1}{|(\mathbf{I} - \dot{\mathbf{P}})\dot{\mathbf{e}}|^2/(n-p)}$$

$$F_4 = \frac{|(\mathbf{I} - \hat{\mathbf{P}}_2)\hat{\mathbf{V}}_1(\theta_1 - \hat{\theta}_1)|^2/p_1}{|\hat{\mathbf{e}}|^2/(n-p)}$$

In these expressions $\mathbf{e} = \mathbf{y} - \eta(\theta)$ is the residual vector for a given θ. The matrix $\mathbf{P} = \mathbf{V}(\mathbf{V}^T\mathbf{V})^{-1}\mathbf{V}^T$ denotes the projection matrix onto the subspace spanned by the columns of $\partial\eta/\partial\theta = \mathbf{V} = (\mathbf{V}_1, \mathbf{V}_2)$, and $\mathbf{P}_2$ is the similar projection onto the subspace spanned by the columns of $\partial\eta/\partial\theta_2 = \mathbf{V}_2$. The symbols $\tilde{}$ and $\dot{}$ above a character indicate evaluation using $\theta_{1,0}$ and, respectively, θ_2 equal to its restricted maximum likelihood estimate $\tilde{\theta}_2$, or its unrestricted maximum likelihood estimate $\hat{\theta}_2$. Similarly, $\hat{}$ denotes evaluation with $\hat{\theta}_1$ and $\hat{\theta}_2$, the maximum likelihood estimates. F_1 is the familiar likelihood ratio statistic, F_2 and F_3 are derived from score statistics using different estimators for the nuisance parameters θ_2, and F_4 uses the large sample normality of $\hat{\theta}_1$ and its independence from the residuals $\hat{\mathbf{e}}$. For linear models the four F ratios are equal and follow an F distribution with p_1 and $n-p$ degrees of freedom. For nonlinear models the F distribution is only an approximation except in certain cases.

Johansen (1983) showed that, as σ^2 goes to zero, the limiting distribution of F_1 when all the parameters are of interest is, to terms of order σ^3,

$$1 - \sigma^2 \gamma n / p(n-p) \tag{1.1}$$

times that of an F with p and $n-p$ degrees of freedom. The factor γ depends on $\partial\eta/\partial\theta$ and $\partial^2\eta/\partial\theta^2$ evaluated at θ_0 and is closely related to measures of intrinsic nonlinearity proposed by Beale (1960) and Bates and Watts (1980). Similar multiplicative factors are presented in the following section for the F ratios enumerated above.

2. The correction factors

The correction factors are obtained by determining the limiting distributions of the F ratios as σ^2 goes to zero, using an approach similar to that of Johansen (1983).

Details of the derivation are shown in the appendix. Separate expansions are obtained for the squared lengths in the numerator and denominator of each F ratio about the true value of θ up to order four in ϵ. Evaluation of these expansions is complicated, but is facilitated by use of the *vec* operator, which places the column vectors of a matrix into one long vector, the *vec* permutation matrix, and associated identities related to differentation (see e.g. Henderson and Searle, 1979; Magnus and Neudecker, 1979; Wiens, 1985). In all four cases the numerator and denominator expansions have the form

$$|(P-P_2)\epsilon|^2 + c_{i1}(\epsilon) + f_{i1}(\epsilon) + O_p(\sigma^5) \qquad (2.1)$$

and

$$|(I-P)\epsilon|^2 + c_{i2}(\epsilon) + f_{i2}(\epsilon) + O_p(\sigma^5) \qquad (2.2)$$

respectively, where c_{ij} and f_{ij} are homogeneous polynomials of degree three and four in the elements of ϵ, for $j = 1, 2$ and $i = 1,..., 4$. When divided by σ^2, (2.1) and (2.2) are, to terms of $O_p(\sigma^3)$, independently distributed as $(1+\alpha_{i1}\sigma^2/p_1)\chi^2_{p_1}$ and $\{1+\alpha_{i2}\sigma^2/(n-p)\}\chi^2_{n-p}$, where $\alpha_{ij} = \sigma^{-4}E[f_{ij}]$. Thus, to the same order of approximation, F_i is distributed as

$$(1-\gamma_i\sigma^2)F^{p_1}_{n-p} \qquad (2.3)$$

where

$$\gamma_i = \frac{\alpha_{i2}}{n-p} - \frac{\alpha_{i1}}{p_1} \qquad (2.4)$$

For $i = 3$ and 4 the expressions for α_{i1} are prohibitively complicated, involving numerous terms and up to third derivatives of η with respect to θ. Because it is unlikely such terms could be used in practice, they are not reported here. For $i = 1$ and 2, however, the expectations reduce to manageable expressions which depend on $\partial\eta/\partial\theta$ and $\partial^2\eta/\partial\theta^2$ but not on $\partial^3\eta/\partial\theta^3$. Their values are

$$\alpha_{11} = \alpha_{21} + \alpha_{22} - \alpha_{12} \tag{2.5}$$

$$\alpha_{12} = -\frac{1}{2}b^T\{(V^TV)^{-1} \otimes (V^TV)^{-1} \otimes I-P\}b + \frac{1}{4}b^T[\{vec(V^TV)^{-1}\}\{vec(V^TV)^{-1}\}^T \otimes I-P]b \tag{2.6}$$

$$\alpha_{21} = -\frac{1}{2}b_2^T\{(V_2^TV_2)^{-1} \otimes (V_2^TV_2)^{-1} \otimes P-P_2\}b_2 + \frac{1}{4}b_2^T[\{vec(V_2^TV_2)^{-1}\}\{vec(V_2^TV_2)^{-1}\}^T \otimes P-P_2]b_2 \tag{2.7}$$

$$\alpha_{22} = -\frac{1}{2}b_2^T\{(V_2^TV_2)^{-1} \otimes (V_2^TV_2)^{-1} \otimes I-P\}b_2 + \frac{1}{4}b_2^T[\{vec(V_2^TV_2)^{-1}\}\{vec(V_2^TV_2)^{-1}\}^T \otimes I-P]b_2 \tag{2.8}$$

where $b = vec\ \partial^2\eta/\partial\theta^2$ and $b_2 = vec\ \partial^2\eta/\partial\theta_2^2$. In these expressions, the matrix differentiation conventions of Wiens (1985) are followed. Thus $\partial\eta/\partial\theta$ is an $n\times p$ matrix and $\partial^2\eta/\partial\theta^2 = \partial(vec\ \partial\eta/\partial\theta)/\partial\theta$ is an $np\times p$ matrix. The vector b is therefore $np^2\times 1$ and contains all the second derivatives $\partial^2\eta_i/\partial\theta_j\partial\theta_k$, for $i = 1,\dots, n$, $j = 1,\dots, p$ and $k = 1,\dots, p$. The $np_2^2\times 1$ vector b_2 contains only the second derivatives with respect to the nuisance parameters θ_2. Each of the terms α_{12}, α_{21} and α_{22} is the difference between two positive terms and can be positive or negative. For tests of composite hypotheses the correction factors are evaluated using the hypothesized value for θ_1 and the restricted estimate $\tilde{\theta}_2$. For the associated confidence regions the factors are calculated using the unrestricted maximum likelihood estimates $\hat{\theta}_1$ and $\hat{\theta}_2$. Spjøtvoll points out in the discussion of Johansen's (1983) paper that it may be more accurate to allow the correction factors to vary over the confidence region, although to do so complicates matters numerically.

3. Relationship to previous work

Johansen (1983) considered the likelihood ratio statistic F_1 when all the parameters are of interest. In this case α_{21} and α_{22} disappear, so from (2.4) and (2.5), $\gamma_1 = n\alpha_{12}/p(n-p)$, and then a comparison of (1.1) and (2.3) shows that $\alpha_{12} = \gamma$. Johansen's (1983) result refines that of Beale (1960), whose more conservative factor replaces γ by $-(p+2)N_{min}/\sigma^2$ in (1.1). Beale (1960) showed that N_{min} is a measure of intrinsic nonlinearity of the expectation surface, which consists of all points $\eta(\theta)$. The formula for N_{min} is the same as $\sigma^2\gamma/(p+2)$ if the sign of the first term in (2.6) is changed from negative to positive. Bates and Watts (1980) showed that N_{min} is one quarter of their mean square intrinsic curvature. Inspection of (2.7) and (2.8) shows that $\alpha_{21}+\alpha_{22}$ gives an expression identical to that for α_{12}, but for the model with only p_2 parameters obtained by fixing θ_1 at the hypothesized value. The sum $\alpha_{21}+\alpha_{22}$ therefore measures the intrinsic curvature of the restricted expectation surface, and α_{11} measures the difference in intrinsic curvature between the entire and restricted expectation surfaces.

The awkwardness of the correction factors for F_3 and F_4 reflects the fact that both these statistics are based on rather crude methods of eliminating θ_2. Hamilton (1986) showed that a quadratic expansion for F_3 in terms of the tangent plane coordinates is more complicated than similar expansions for F_1 and F_2, while the expansion for F_4 is unrealistically simple.

4. Numerical Method

The α_{ij} can be calculated accurately and efficiently using a technique similar to that of Bates et al (1983). Orthogonal matrices $U = (U_1, U_2)$, N_1 and N_2 are obtained in such a way that their columns form bases for the tangent plane to the expectation surface at $\hat{\theta}$, the normal acceleration space and the remainder of the sample space. Then $(V^T V)^{-1} = LL^T$, where $U = VL$ and L is lower triangular, $I-P = N_1 N_1^T + N_2 N_2^T$, and $P-P_2 = U_1 U_1^T$. These decompositions simplify (2.6) to

$$\alpha_{12} = -\frac{1}{2} |\text{vec}\, A|^2 + \frac{1}{4} \left| \sum_{i=1}^{p} (e_i^T \otimes I_r) A e_i \right|^2$$

where e_i is the vector of length p with 1 in the i'th row and zeroes elsewhere, and $A = (L^T \otimes N_1^T)(\partial^2\eta/\partial\theta^2)L$. The matrix A is $rp \times p$, where $r \leqslant p(p+1)/2$ is the dimension of the normal acceleration space, and contains the nonzero entries in the intrinsic acceleration array, A^N, of Bates and Watts (1980). The first part of α_{12} is minus one half the sum of squares of all the entries in A^N, and the second is one quarter the sum of squares of the traces of the faces of A^N.

The constants α_{21} and α_{22} are similar functions of a $p_1 p_2 \times p_2$ matrix $A_{2,1} = (L_{22}^T \otimes U_1^T)(\partial^2\eta/\partial\theta_2^2)L_{22}$ and an $rp_2 \times p_2$ matrix $A_{2,2} = (L_{22}^T \otimes N_1^T)(\partial^2\eta/\partial\theta_2^2)L_{22}$, which contain the same entries as the first p_1 and last r faces of the intrinsic curvature array for the restricted model with θ_1 fixed. The intrinsic arrays and therefore the α_{ij} are invariant under parameter transformations.

The necessary reduction of the model derivatives is achieved as follows. Let $W = (W_1, W_2)$ be an $n \times p(p+1)/2$ matrix containing the distinct vectors

$\partial^2\eta/\partial\theta_i\partial\theta_j$, where W_2 contains the second derivatives with respect to the components of θ_2. Obtain the QR decomposition $(V_1,V_2|W_1,W_2) = (U_1,U_2|N_1,N_2)S$, where

$$S = \begin{bmatrix} R_{11} & 0 & U_1^TW_1 & U_1^TW_2 \\ R_{21} & R_{22} & U_2^TW_1 & U_2^TW_2 \\ 0 & 0 & N_1^TW_1 & N_1^TW_2 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The columns of second derivatives can be pivotted to improve accuracy. Then $L = R^{-1}$, $L_{22} = R_{22}^{-1}$, and the second derivative contributions to A, $A_{2,1}$ and $A_{2,2}$ are N_1^TW, $U_1^TW_2$, and $N_1^TW_2$. Determination of r, and multiplications involving L and L_{22} can be achieved as in Bates et al (1983).

5. Testing for lack of fit

The usual test statistic for evaluating lack of fit in regression is

$$F_5 = \frac{\sum n_i(\hat{y}_i - \bar{y}_i)^2/(k-p)}{\sum\sum(y_{ij} - \bar{y}_i)^2/(n-k)}$$

where k is the number of distinct settings of the independent variables and n_i is the number of replications y_{ij} at the i'th setting, x_i. The average and fitted values of y at x_i are $\hat{y}_i$ and $\bar{y}_i$ respectively. The denominator sum of squares is distributed exactly as $\sigma^2\chi^2_{n-k}$; however, unlike the linear regression case, the numerator sum of squares is neither exactly $\sigma^2\chi^2_{k-p}$ nor exactly independent of the denominator.

The statistic F_5 is the likelihood ratio statistic for the test H_0: $y = \eta(\theta) + \epsilon$ versus H_a: $y = X\beta + \epsilon$, where $\beta_i = E[y|x_i]$, $i = 1,\dots, k$, and X is the matrix of indicators. The model under H_0 is a submodel of that under H_a, and a parameterization

$\phi = \phi(\theta)$ can be found for which the null hypothesis is H_0: $\phi_1 = 0$, where ϕ_1 is $(k-p)\times 1$. Thus, previous results can be applied, with $i = 1$, $p = k$, and $p_1 = p$. Fortunately it is not necessary to find ϕ or the model derivatives with respect to ϕ because of the invariance of the α_{ij} under reparameterizations. Indeed, under H_a the model is linear in β, and the residual sum of squares is exactly χ^2, so $\alpha_{12} = 0$ and $\alpha_{11} = \alpha_{21}+\alpha_{22}$. The sum $\alpha_{21}+\alpha_{22}$ can be evaluated using the derivatives of the non-linear model with respect to θ, and equals Johansen's (1983) γ. From (2.4) and (2.3), F_5 is distributed as

$$F_5 \sim (1+\sigma^2\gamma/p)F_{n-k}^{k-p} \tag{5.1}$$

The multiplicative correction factor for testing lack of fit is always closer to unity than is Johansen's (1983) factor (1.1) for the likelihood ratio test.

6. Examples

Data for the model $\eta_t = \theta_1(1-\theta_2 e^{-\theta_3 x_t})$ is given in problem N, pg. 524 of Draper and Smith (1981). The multiplicative factors for the likelihood ratio test and the score test are shown in Table 1 for their data set #1, using parameter estimates $\hat{\theta} = (263.71,\ 0.94801,\ 4.86911\times 10^{-04})$ and $s^2 = 50.98$.

Table 1: Multiplicative correction factors for hypothesis tests

parameters fixed	likelihood ratio	score
1	1.0023	1.0000
2	0.6500	0.6499
3	1.0024	1.0000
12	1.0014	1.0000
13	1.0014	1.0000
23	1.0014	1.0000
123	1.0011	1.0000

The first column of the table gives the indices of the parameters fixed by the null hypothesis, and all possible subsets are considered. The second and third columns give the multiplicative factors from (2.3), (2.4) and (2.5)- (2.8), calculated using the numerical techique described in section 4. The score test is exact for all the parameters in any model, and for the subset of nonlinear parameters in a partially linear model (Hamilton, Watts and Bates, 1982; Hamilton, 1986), and this explains why the score factor is exactly unity for the last three rows of the table. The factors are similar for the two tests and are very close to unity, apart from the case where θ_2 is of interest. A confidence interval for θ_2 without regard for the correction would be much too large.

The correction factor for testing lack of fit can be illustrated using the data from Hamilton (1986). For this data set $F_5 = \dfrac{462.6/4}{397/5} = 1.46$, and the degrees of freedom are 4 and 5. The correction factor (5.1) is 1.00006, which is so close to unity that it has little effect on the table values. Thus the conclusion of no lack of fit remains valid.

7. Discussion

Results like those in Table 1 were obtained for 89 data sets corresponding to 37 different models with up to 5 parameters. The likelihood ratio factor for all the parameters was between 1 and 1.025 in every case, and exceeded 1.01 only 5 times. As mentioned in section 5, the factor for testing lack of fit is even less extreme. Thus, on the basis of this sample of data sets, it may be concluded that the correction factor can be ignored when testing all the parameters or when testing lack of fit, unless the test statistic is very close to the critical value.

The correction factors for subsets were usually farther from 1 than the correction factors for all the parameters, and the smallest subsets usually had the most extreme factors for a given data set. The likelihood ratio factor was usually positive and farther from 1 than the score factor, which was usually less than 1. The subset likelihood ratio factors were all between .6500 and 1.0612, while the subset score factors were all between .6499 and 1.0021. Because the subset factors are occasionally substantially different from 1, their calculation is recommended.

Appendix

We first outline the passage from (2.1), (2.2) to (2.3), then derive (2.1), (2.2) and evaluate the α_{ij}.

Make the orthogonal decomposition $\epsilon = \nu+\omega+\zeta = (P-P_2)\epsilon+P_2\epsilon+(I-P)\epsilon$, and let Q_1, Q_2 represent the terms in (2.1), (2.2). Evidently, $f_j(\epsilon) = \sum m_{j,abc}^T \rho_{abc}$ for some $n^4\times 1$ vectors $\mathbf{m}_j$ of constants, where $\rho_{abc} = \nu^{[a]} \otimes \zeta^{[b]} \otimes \omega^{[c]}$, [] denotes an iterated Kronecker product, and the summation is constrained by $a+b+c=4$. With $\phi_p(s)=(1-2is\sigma^2)^{-p/2}$, the characteristic function $\psi(s,t)=E[exp(isQ_1+itQ_2)]$ may be expanded as

$$\psi(s,t)=\phi_{p_1}(s)\phi_{n-p}(t)+E[\{isf_1(\epsilon)+itf_2(\epsilon)\}exp(is|\nu|^2+it|\zeta|^2)]+O\{(s+t)\sigma^5\} .$$

The expectations over the c_j have vanished for reasons of symmetry. Similarly, those utilizing odd a or b are also zero. The remaining expectations are easily evaluated after noting that $|\nu|$, $|\zeta|$, $\nu/|\nu|$, $\zeta/|\zeta|$ are mutually independent. Thus

$$E[f_j(\epsilon)exp(is|\nu|^2+it|\zeta|^2)]=\sum m_{j,abc}^T\phi_{p_1+a}(s)\phi_{n-p+b}(t)E[\rho_{abc}]$$
$$=\alpha_j\phi_{p_1}(s)\phi_{n-p}(t)+O\{(s+t)\sigma^6\},$$

whence $\psi(\frac{s}{\sigma^2},\frac{t}{\sigma^2})=\phi_{p_1}\{\frac{s}{\sigma^2}(1+\alpha_1\sigma^2/p_1)\}\phi_{n-p}\{\frac{t}{\sigma^2}(1+\alpha_2\sigma^2/n-p)\}+O\{(s+t)\sigma^3\}$

and (2.3) follows.

We now obtain (2.1), (2.2) for F_1, F_2, and F_3. The derivation for F_4 requires a separate but similar development, not reported here. Those for F_1, F_2, and F_3 can be unified by the following device. Put $\mathbf{H}_{p_2\times p}=(\mathbf{O}_{p_2\times p_1}|\mathbf{I}_{p_2})$, and consider the three 6- tuples $(\mathbf{e}, \hat{\theta}_+, \hat{\theta}_*, \mathbf{J}, \mathbf{K}, q)=(\hat{\mathbf{e}}, \hat{\theta}, \hat{\theta}, \mathbf{I}_p, \mathbf{I}_p, p)$, $(\tilde{\mathbf{e}}, \tilde{\theta}, \tilde{\theta}_2, \mathbf{H}, \mathbf{H}, p_2)$, $(\dot{\mathbf{e}}, \dot{\theta}, \hat{\theta}_2, \mathbf{I}_p, \mathbf{H}, p_2)$. Each of the F_i is composed of terms of the form $|\overset{+}{\mathbf{Q}}\mathbf{e}|^2$, where $\mathbf{Q}=\mathbf{Q}(\theta)$ is a projection matrix, $\overset{+}{\mathbf{Q}}$ its evaluation at $\hat{\theta}_+$. For F_1, we require the choices $(\mathbf{e},\mathbf{Q})=(\hat{\mathbf{e}}, \mathbf{I}_n)$, $(\tilde{\mathbf{e}}, \mathbf{I}_n)$; for F_2 we use $(\tilde{\mathbf{e}}, \mathbf{P}-\mathbf{P}_2)$, $(\tilde{\mathbf{e}}, \mathbf{I}-\mathbf{P})$; for F_3 we use $(\dot{\mathbf{e}}, \mathbf{P}-\mathbf{P}_2)$, $(\dot{\mathbf{e}}, \mathbf{I}-\mathbf{P})$. In each case $\mathbf{K}\hat{\theta}_+=\hat{\theta}_*$ is $q\times 1$. For $\mathbf{e}$ either $\hat{\mathbf{e}}$ or $\tilde{\mathbf{e}}$, $\hat{\theta}_+$ is defined by the q normal equations $\mathbf{K}\overset{+}{\mathbf{V}}^T\{\mathbf{y}-\eta(\hat{\theta}_+)\}=0$. From this, we can expand $\hat{\theta}_*-\theta_*$ in Kronecker powers of ϵ. The case $\mathbf{e}=\dot{\mathbf{e}}$ is then included via the relationship

$\hat{\theta}_2 - \theta_2 = \mathrm{KJ}^T(\hat{\theta} - \theta)$. Series expansions of $\mathbf{e}$ and $\overset{+}{\mathbf{Q}}$ in powers of $\hat{\theta}_* - \theta_*$ then yield (2.1), (2.2).

Notation for, and properties of, the *vec* operator and *vec*- permutation matrix $\mathbf{I}_{n,p}$ are as in Wiens (1985). Write the normal equations as

$$[\mathbf{I}_q \otimes \{\mathbf{y} - \eta(\hat{\theta}_+)\}^T] vec\ \overset{+}{\mathbf{V}}\mathbf{K}^T = 0 \ . \tag{A.1}$$

Taylor series expansions around θ_* give

$$\eta(\hat{\theta}_+) = \eta(\theta_+) + (vec\ \mathbf{V})^T \{\mathbf{K}^T(\hat{\theta}_* - \theta_*) \otimes \mathbf{I}_n\} + \frac{1}{2}\mathbf{b}^T[\{\mathbf{K}^T(\hat{\theta}_* - \theta_*)\}^{[2]} \otimes \mathbf{I}_n]$$

$$+ \frac{1}{6}\mathbf{c}^T[\{\mathbf{K}^T(\hat{\theta}_* - \theta_*)\}^{[3]} \otimes \mathbf{I}_n] + O_p(\sigma^4) \ , \tag{A.2}$$

$$vec\ \overset{+}{\mathbf{V}}\mathbf{K}^T = vec\ \mathbf{V}\mathbf{K}^T + \{(\hat{\theta}_* - \theta_*)^T \otimes \mathbf{I}_{nq}\}(\mathbf{K}^{[2]} \otimes \mathbf{I}_n)\mathbf{b} + O_p(\sigma^2) \ . \tag{A.3}$$

Here, $\mathbf{c} = vec\ \partial^3\eta/\partial\theta^3$ is $np^3 \times 1$. Insert (A.2) and (A.3) into (A.1), together with

$$\hat{\theta}_* - \theta_* = l(\epsilon) + q(\epsilon) + O_p(\sigma^3) \tag{A.4}$$

where $l(\epsilon)$ is $O_p(\sigma)$, $q(\epsilon)$ is $O_p(\sigma^2)$. Solving for l and q gives

$$l(\epsilon) = \mathbf{M}\epsilon \ , \ q(\epsilon) = (\mathbf{b}^T \otimes \mathbf{I}_q)\mathbf{N}(\epsilon \otimes \epsilon) \ , \tag{A.5}$$

where

$$\mathbf{M} = \mathbf{KJ}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \ , \ \mathbf{A} = \mathbf{VJ}^T \ , \ \mathbf{R} = \mathbf{A}(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \ ,$$

$$\mathbf{N} = (\mathbf{I}_p \otimes \mathbf{I}_{n,p} \otimes \mathbf{I}_q)\{\mathbf{J}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes \mathbf{I} - \mathbf{R} \otimes vec\ \mathbf{KJ}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{J}\}$$

$$-\frac{1}{2}\{\mathbf{J}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes \mathbf{J}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T \otimes vec\ \mathbf{KJ}^T(\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\} \ .$$

The use of $\mathbf{KJ}^T$ in the definitions of $\mathbf{M}$ and $\mathbf{N}$ extends the validity of (A.4) to the case $\mathbf{e} = \dot{\mathbf{e}}$. Now (A.4) and (A.5) in (A.2) give

$$e(\hat{\theta}_+) = y - \eta(\hat{\theta}_+) = S\epsilon - T(\epsilon \otimes \epsilon) - U(\epsilon \otimes \epsilon \otimes \epsilon) + O_p(\sigma^4) \qquad (A.6)$$

where

$$S = I_n - VK^T M \ , \ T = (b^T \otimes I_n)[(I_{np^2} \otimes VK^T)N + \frac{1}{2}\{(K^T M)^{[2]} \otimes vec\ I_n\}] \ ,$$

$$U = (b^T \otimes I_n)\{K^T M \otimes K^T(b^T \otimes I_q)N \otimes vec\ I_n\} + \frac{1}{6}(c^T \otimes I_n)\{(K^T M)^{[3]} \otimes vec\ I_n\}$$

Now expand $\overset{+}{Q}$ in powers of $\hat{\theta}_* - \theta_*$ and use (A.4), (A.5) to get

$$\begin{aligned} vec\ \overset{+}{Q} &= vec\ Q + (\partial Q/\partial\theta)K^T M\epsilon \\ &\quad + (\epsilon^T \otimes \epsilon^T \otimes I_{n^2})[\{N^T(b \otimes I_q)K \otimes I_{n^2}\}vec\ \partial Q/\partial\theta \\ &\quad + \frac{1}{2}\{(M^T K)^{[2]} \otimes I_{n^2}\}vec\ \partial^2 Q/\partial\theta^2] + O_p(\sigma^3) \ . \end{aligned} \qquad (A.7)$$

Inserting (A.6), (A.7) into $|\overset{+}{Q}e|^2 = (e^T \otimes e^T)vec\ \overset{+}{Q}$ gives

$$|\overset{+}{Q}e|^2 = |QS\epsilon|^2 + c(\epsilon) + f(\epsilon) + O_p(\sigma^5) \qquad (A.8)$$

where $c(\epsilon)$ is a homogeneous cubic polynomial in the elements of ϵ, and

$$\begin{aligned} f(\epsilon) &= (\epsilon^T)^{[4]}[(vec\ T^T QT - 2vec\ U^T QS) \\ &\quad + [\{N^T(b \otimes K) \otimes S^T \otimes S^T\} - 2(M^T K \otimes S^T \otimes T^T)]vec\ \partial Q/\partial\theta \\ &\quad + \frac{1}{2}\{(M^T K)^{[2]} \otimes S^T \otimes S^T\}vec\ \partial^2 Q/\partial\theta^2] \end{aligned} \qquad (A.9)$$

For each of the terms required in F_1, F_2 and F_3, QS is one of $I-P$, $I-P_2$, or $P-P_2$. The required versions of (2.1) and (2.2) are thus obtained as particular cases of (A.8) and (A.9).

To evaluate the $\alpha = \sigma^{-4}E[f(\epsilon)]$ and thus complete the derivation, we note that

$$\sigma^{-4}E[\epsilon^{[4]}] = \{I_{n^4} + (I_{n^2} \otimes I_{n,n})\}vec\ I_{n^2} + (vec\ I_n \otimes vec\ I_n) \qquad (A.10)$$

(Magnus and Neudecker, 1979). The derivatives of Q, which are not required for F_1, may be calculated from the results in section 2 of Wiens (1985). When Q is of the form $C(C^TC)^{-1}C^T$ for C an $n \times q$ matrix function of θ, then

$$\partial Q/\partial C = (I_{n^2}+I_{n,n})\{C(C^TC)^{-1} \otimes I-Q\} \;,\; \partial Q/\partial\theta = (\partial Q/\partial C)(\partial C/\partial\theta) \quad (A.11)$$

$$\partial^2 Q/\partial\theta^2 = \{(\partial C/\partial\theta)^T \otimes I_{n^2}\}(\partial^2 Q/\partial C^2)(\partial C/\partial\theta) + (I_q \otimes \partial Q/\partial C)(\partial^2 C/\partial\theta^2) \quad (A.12)$$

$$\begin{aligned} \partial^2 Q/\partial C^2 = \{I_{nq} \otimes (I_{n^2}+I_{n,n})\}(I_q \otimes I_{n,n} \otimes I_n) \times \\ [-\{I_{nq} \otimes (I_{n^2}+I_{n,n})\}\{vec\ C(C^TC)^{-1} \otimes C(C^TC)^{-1} \otimes I-Q\} \\ +\{(C^TC)^{-1} \otimes I-Q \otimes vec\ I-Q\} \\ -(I_{n,q} \otimes I_{n^2})\{C(C^TC)^{-1} \otimes (C^TC)^{-1}C^T \otimes vec\ I-Q\}] \end{aligned} \quad (A.13)$$

Inserting (A.10), together with the appropriate versions (or differences of versions) of (A.11) - (A.13) into (A.9) now yields (2.5)- (2.8), after a rather lengthly calculation.

Acknowledgement

This work was supported by the National Sciences and Engineering Research Council of Canada.

References

Bates, D. M., Hamilton, D. C. and Watts, D. G. (1983). Calculation of intrinsic and parameter-effects curvatures for nonlinear regression models. *Comm. Statist.* 12, 469- 77.

Bates, D. M. and Watts, D. G. (1980). Relative curvature measures of nonlinearity (with discussion). *J. R. Statist. Soc. B* 42, 1- 25.

Beale, E. M. L. (1960). Confidence regions in nonlinear estimation (with discussion).

J. R. Statist. Soc. B 22, 41-88

Draper, N. R. and Smith, H. (1981). *Applied Regression Analysis (Second Edition).* New York: John Wiley & Sons, Inc.

Hamilton, D. C. (1986). Confidence regions for parameter subsets in nonlinear regression. *Biometrika* 73, 1, xxx-xxx.

Hamilton, D. C., Watts, D. G. and Bates, D. M. (1982). Accounting for intrinsic nonlinearity in nonlinear regression parameter inference regions, *Ann. Statist.*, 10, 386-393.

Henderson, H. V. and Searle, S. R. (1979). *Vec* and *vech* operators for matrices, with some uses in Jacobians and multivariate statistics, *Canad. J. Statist.* 7, 65-81.

Johansen, S. (1983). Some topics in regression. *Scand. J. Statist.* 10, 161-94.

Magnus, J. R. and Neudecker, H. (1979). The commutation matrix: some properties and applications, *Ann. Statist.* 7, 381-394.

Wiens, D. P. (1985). On some pattern-reduction matrices which appear in statistics, *Linear algebra and its applications,* 67, 233-258.