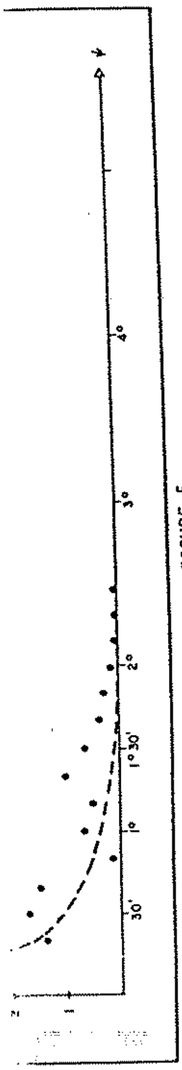

FIGURE 5

A STATISTICAL TEST
for
SIGNIFICANCE OF PEAKS
IN THE
LEAST SQUARES SPECTRUM

by

R.R. Steeves

Geodetic Survey Division
Surveys and Mapping Branch
Department of Energy, Mines and Resources
615 Booth Street
Ottawa, K1A 0E9

In "Collected Papers, Geodetic Survey." Surveys and Mapping Branch, Dept. of Energy, Mines & Resources, 1981.

ABSTRACT

After a review of the method of least squares spectral analysis (Vanicek, 1971), with the inclusion of an arbitrary non-singular covariance matrix for the observed time series, a statistical test is developed for testing the significance of peaks in the least squares spectrum. A by-product of this investigation is the derivation of the probability density function of the least squares spectrum under the hypothesis that the observed time series is derived from a vector of statistically independent random variables whose probability density function is the multi-dimensional normal distribution $N_n(0,\sigma^2 I)$. A simple formula is given for computing a $100(1-\alpha)\%$ critical value for significant peaks in the spectrum.

I INTRODUCTION

In Vanicek (1971) it was suggested that a statistical criterion could be derived for testing the statistical significance of individual peaks in the least squares spectrum. In this paper such a criterion is derived after a review of the method of least squares spectral analysis. For a more detailed discussion of the method, especially concerning its desirable properties for analysis of non-stationary time series, the reader is referred to the paper cited above.

We consider an observed time series which we represent by a vector f (dim(f) = n) of values f_i, i=1, 2, ..., n observed at respective times t_i, i=1, 2, ... n. Although equally spaced times result in a more efficient algorithm for computing the least squares spectrum (Wells and Vanicek (1978)) we do not make this assumption here. We will assume that we may have knowledge about the accuracy of f in the form of a non-singular covariance matrix C_f.

The basic objective of the least squares spectral analysis is to detect unknown periodic signals in f especially in the case that f also contains various systematic variations of unknown magnitude whose functional forms are known. After Vanicek (1971) we call these known constituents "systematic noise" since their presence is a nuisance from the point of view of detecting the unknown periodic signal. We call the unknown periodic signal "systematic signal". It is felt that we need not restrict ourselves to periodic signals or indeed to systematic signals. However an extension of the method to consider arbitrary functional forms for the signal or to consider zero mean stochastic correlated signal constituents is beyond the scope of the present paper.

Once a statistically significant signal is detected it should be related to the physics of the system being observed. Once it is understood the detected signal becomes systematic noise and we may then search for further hidden periodicities. From this point of view the method is an iterative one and is meaningful only if we do not lose sight of the physics of the situation (as with any method of spectral analysis of observed time series).

II LEAST SQUARES SPECTRAL ANALYSIS

Least squares spectral analysis has its basis in the least squares estimation of the magnitudes of the systematic noise constituents and the coefficients a and b in the trigonometric term $a.\cos \omega t + b.\sin \omega t$ where ω (angular frequency) is the argument of the spectrum (defined below). It will be shown here that for each ω_i in the spectrum we determine the simultaneous least squares estimates of a and b along with the systematic noise magnitudes. Thus we see that the least squares spectrum will not be distorted by the systematic noise (see also [Taylor and Hamilton, 1972]). The systematic noise may consist of various functional forms examples of which are constant terms (datum biases), linear (quadratic, exponential, etc.) trends and trigonometric terms of known angular frequency. Note that this treatment allows for gaps in the observed series which are very common in observed time series in the physical sciences.

We can thus attempt to model the time variations of an observed time series f by

$$\tilde{f} = A_N x_N + A_S x_S \tag{2.1}$$

where A_N (dim A_N = (n,m), m<n+2) is the Vandermonde matrix of functional values of the systematic noise with unknown magnitudes x_N (dim x_N = m) and A_S (dim A_S = (n,2) is the matrix of functional values ($\cos\omega t_i$ and $\sin \omega t_i$, i = 1,2, ..., n) of the systematic signal with unknown magnitudes $x_S = [a, b]^T$. (Superscript (^{T}) denotes matrix transposition and ($^{-1}$) will denote matrix inversion). We define the residual vector

$$r = f - \hat{f} \quad . \tag{2.2}$$

Minimization of the quadratic form

$$r^T C_f^{-1} r = (f - \tilde{f})^T C_f^{-1} (f - \tilde{f}) \tag{2.3}$$

with respect to

$$x = [x_N \vdots x_S]^T \tag{2.4}$$

yields the familiar expression

$$\hat{x} = (A^T C_f^{-1} A)^{-1} \quad A^T C_f^{-1} f \tag{2.5}$$

for the least squares estimates of x where

$$A = [A_N \vdots A_S].$$

It is assumed here that $A^T C_f^{-1} A$ is non-singular which has to be confirmed in each specific case in practice.

Similarly we get

$$\hat{x}_P = (A_N^T C_f^{-1} A_N)^{-1} \; A_N^T C_f^{-1} f \tag{2.6}$$

for the least squares estimate of the systematic noise magnitudes x_N when the systematic signal is ignored.

We thus have

$$\hat{r} = f - A^T \hat{x} \tag{2.7}$$

and

$$\hat{r}_P = f - A_N^T \; \hat{x}_P \tag{2.8}$$

as the least squares estimates of the residual vectors r and r_p respectively.

We see that $\hat{r}_p^T C_f^{-1} \hat{r}_p$ is a measure of the variance of f which is not modelled by the systematic noise. Thus the difference

$$S^*(\omega) = \hat{r}_p^T C_f^{-1} \hat{r}_p - \hat{r}^T C_f^{-1} \hat{r} \qquad (2.9)$$

is a measure of the variance of f absorbed by the trigonometric term represented by $A_S x_S$ or, in other words, $s^*(\omega)$ gives a measure of the maximum (since $\hat{r}^T C_f^{-1} \hat{r}$ is minimized) contribution of $A_S x_S$ to the variance of f. Since $s^*(\omega) \in [0, \hat{r}_p^T C_f^{-1} \hat{r}_p]$ we can define the "normalized" least squares spectrum

$$s(\omega) = \frac{\hat{r}_p^T C_f^{-1} \hat{r}_p - \hat{r}^T C_f^{-1} \hat{r}}{\hat{r}_p^T C_f^{-1} \hat{r}_p}, \quad \hat{r}_p^T C_f^{-1} \hat{r}_p \neq 0 \qquad (2.10)$$

or

$$s(\omega) = 1 - \frac{\hat{r}^T C_f^{-1} \hat{r}}{\hat{r}_p^T C_f^{-1} \hat{r}_p} \qquad (2.11)$$

whose values lie in the interval [0,1]. We see that $100.s(\omega)$ is a measure of the percentage of the variance of f not accounted for by $A_N \hat{x}_N$ that is accounted for by $A_S \hat{x}_S$.

We will show now that the evaluation of the spectrum does not require the explicit computation of equation (2.5). We can rewrite equation (2.5) more explicitly as

$$\begin{bmatrix} \hat{x}_N \\ \hat{x}_S \end{bmatrix} = \left\{ \begin{bmatrix} A_N^T \\ A_S^T \end{bmatrix} C_f^{-1} [A_N^T \quad A_S^T] \right\}^{-1} \begin{bmatrix} A_N^T \\ A_S^T \end{bmatrix} f \qquad (2.12)$$

$$= \begin{Bmatrix} N_{NN} & N_{NS} \\ N_{SN} & N_{SS} \end{Bmatrix}^{-1} \begin{bmatrix} u_N \\ u_S \end{bmatrix} \qquad (2.13)$$

where we have let $N_{ij} = A_i^T C_f^{-1} A_j$ and $u_j = A_j^T C_f^{-1} f$; $i,j = N,S$. Equation (2.13) can be written (see the Appendix)

$$\begin{bmatrix} \hat{x}_N \\ \hat{x}_S \end{bmatrix} = \begin{bmatrix} (N_{NN}-N_{NS}N_{SS}^{-1}N_{SN})^{-1} & -N_{NN}^{-1} N_{NS}(N_{SS}-N_{SN}N_{NN}^{-1}N_{NS})^{-1} \\ -N_{SS}^{-1}N_{SN}(N_{NN}-N_{NS}N_{SS}^{-1}N_{SN})^{-1} & (N_{SS}-N_{SN}N_{NN}^{-1} N_{NS})^{-1} \end{bmatrix} \begin{bmatrix} u_N \\ u_S \end{bmatrix} \quad (2.14)$$

Using the identities (A.3) and (A.4) we have

$$\hat{x}_N = (N_{NN}^{-1} + N_{NN}^{-1}N_{NS}(N_{SS}-N_{SN}N_{NN}^{-1}N_{NS})^{-1}N_{SN}N_{NN}^{-1})u_N - N_{NN}^{-1}N_{NS}(N_{SS}-N_{SN} N_{NN}^{-1} N_{NS})^{-1} u_S \quad (2.15)$$

and

$$\hat{x}_S = -(N_{SS}-N_{SN}N_{NN}^{-1}N_{NS})^{-1} N_{SN}N_{NN}^{-1} u_N +(N_{SS}-N_{SN}N_{NN}^{-1}N_{NS})^{-1} u_S \,. \quad (2.16)$$

Subtracting equation (2.6) from (2.15) and combining the result with (2.16) we have, with

$$\hat{x}_R = \hat{x}_N - \hat{x}_P \,, \quad (2.17)$$

$$\begin{bmatrix} \hat{x}_R \\ \hat{x}_S \end{bmatrix} = \begin{bmatrix} N_{NN} & N_{NS} \\ N_{SN} & N_{SS} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ u_S - N_{SN}N_{NN}^{-1} u_N \end{bmatrix} . \quad (2.18)$$

Since

$$A_N^T C_f^{-1} f - (A_N^T C_f^{-1} A_N) N_{NN}^{-1} u_N = 0 \quad (2.19)$$

we have

$$\begin{bmatrix} \hat{x}_R \\ \hat{x}_S \end{bmatrix} = \begin{bmatrix} N_{NN} & N_{NS} \\ N_{SN} & N_{SS} \end{bmatrix}^{-1} \begin{bmatrix} A_N^T C_f^{-1} f - A_N^T C_f^{-1} A_N N_{NN}^{-1} u_N \\ A_S^T C_f^{-1} f - A_S^T C_f^{-1} A_N N_{NN}^{-1} u_N \end{bmatrix} . \quad (2.20)$$

Thus, using equation (2.8),

$$\begin{bmatrix} \hat{x}_R \\ \hat{x}_S \end{bmatrix} = \begin{bmatrix} N_{NN} & N_{NS} \\ N_{SN} & N_{SS} \end{bmatrix}^{-1} \begin{bmatrix} A_N^T \\ A_S^T \end{bmatrix} C_f^{-1} \hat{f}_P \,. \quad (2.21)$$

From equations (2.18) and (2.21) we have

$$\hat{x}_S = (N_{SS} - N_{SN} N_{NN}^{-1} N_{NS})^{-1} A_S^T C_f^{-1} \hat{f}_P . \qquad (2.22)$$

Rewriting equation (2.9) as

$$s^*(\omega) = (\hat{f}_p - \hat{r})^T C_f^{-1} (\hat{f}_p + \hat{r}) \qquad (2.23)$$

and using equations (2.7), (2.8) and (2.17) we have

$$s^*(\omega) = \hat{x}_R^T A_N^T C_f^{-1} \hat{f}_p + \hat{x}_R^T A_N^T C_f^{-1} \hat{r} + \hat{x}_S^T A_S^T C_f^{-1} \hat{f}_p + \hat{x}_S^T A_S^T C_f^{-1} \hat{r} . \qquad (2.24)$$

Since the first, second and fourth terms of the right hand side of equation (2.24) are identically zero (Mikhail, 1976) we have

$$s^*(\omega) = \hat{f}_P^T C_f^{-1} A_S \hat{x}_S \qquad (2.25)$$

and thus

$$s(\omega) = \frac{\hat{f}_P^T C_f^{-1} A_S \hat{x}_S}{\hat{f}_P^T C_f^{-1} \hat{f}_p} . \qquad (2.26)$$

Substituting for $\hat{x}_S$ from equation (2.22) we have finally

$$s(\omega) = \frac{\hat{f}_P^T C_f^{-1} A_S (N_{SS} - N_{SN} N_{NN}^{-1} N_{NS})^{-1} A_S^T C_f^{-1} \hat{f}_P}{\hat{f}_P^T C_f^{-1} \hat{f}_P} \qquad (2.27)$$

which has been detailed further for computer implementation by Wells and Vanicek (1978). Note that we have shown explicitly where the covariance matrix C_f enters into the computation of the least squares spectrum. Previous treatments have tacitly assumed $C_f = I$. However it is beyond the scope of this discussion to consider the effects of an arbitrary C_f on the least squares spectrum. (Preliminary investigations indicate that even a sinusoidal covariance function has little or no effect).

III STATISTICAL SIGNIFICANCE OF SPECTRAL PEAKS

We will now investigate the response of the least squares spectrum to "random noise". Specifically we derive the probability density function (pdf) of $s(\omega)$ under the hypothesis that f (dim $f = n$) is a vector of statistically independent random variables with pdf $N_n(0,I)$, that is each $f_i, i = 1,2,\ldots, n$ has a normal pdf with zero mean and unit variance. (We can assume arbitrary mean and variance which will not affect our results since the arbitrary mean can be modelled by $A_N x_N$ and an arbitrary variance δ^2 would cancel out of the expression (3.2) below). Under this hypothesis we have, according to equation (2.26), and since $\hat{r}_P^T A_N = 0$,

$$s'(\omega) = \frac{\hat{r}_P^T \; A_N \hat{x}_N + \hat{r}_P^T \; A_S \hat{x}_S}{r_P^T \hat{r}_P} \tag{3.1}$$

or

$$s'(\omega) = \frac{\hat{r}_P^T G \hat{r}_P}{\hat{r}_P^T \; \hat{r}_P} \tag{3.2}$$

where we have let (with $A = [A_N \vdots A_S]$ as before)

$$G = A \; (A^T A)^{-1} A^T \; . \tag{3.3}$$

We can show (Rao and Mitra (1971), theorem 9.4.1, p178) that the random variables $\hat{r}_p^T G \hat{r}_p$ and $\hat{r}_p^T \hat{r}_p$ are not independently distributed. However, using the same theorem, we can show that the random variables $\hat{r}_p^T (I - G) \hat{r}_p$ and $\hat{r}_p^T G \hat{r}_p$ are independently distributed. Following Jeudy (1980) we can make use of this fact by writing

$$\hat{r}_p^T \hat{r}_p = \hat{r}_p G \hat{r}_p + \hat{r}_p^T (I - G) \hat{r}_p \tag{3.4}$$

and substituting this in equation (3.2) giving

$$s'(\omega) = \frac{1}{1 + \hat{r}_p^T (I - G) \hat{r}_p / (\hat{r}_p^T G \hat{r}_p)} \; . \tag{3.5}$$

Since (Mikhail, 1976)

$$\hat{r}_p \sim N_n(0, D) \; , \tag{3.6}$$

(where "~" is read "has the probability density function"), where

$$D = (I - A_N(A_N^T A_N)^{-1} A_N) \; . \tag{3.7}$$

With

$$H = A_N(A_N^T A_N)^{-1} A_N^T \tag{3.8}$$

we have

$$D = I - H. \tag{3.9}$$

Therefore (Rao and Mitra, 1971; theorem 9.2.1, p171)

$$\hat{r}_p^T \, G \, \hat{r}_p \sim \chi^2(k_1) \tag{3.10}$$

with

$$k_1 = tr(GD) \tag{3.11}$$

where "tr" denotes the trace of a matrix.

But

$$\begin{aligned} GD &= G(I - H) \\ &= G - GH. \end{aligned} \tag{3.12}$$

It can be shown (see Appendix) that $GH = H$ and that G and H are idempotent (that is $G = GG$ and $H = HH$).

Therefore

$$tr(G) = rank\ (G) = m + 2 \tag{3.13}$$

and

$$tr(GH) = rank\ (H) = m\ . \tag{3.14}$$

Thus $k_1 = m + 2 - m = 2$ and

$$\hat{r}_p^T \, G \, \hat{r}_p \sim \chi^2(2)\ . \tag{3.15}$$

Also,

$$\hat{r}_p^T(I - G)\, \hat{r}_p \sim \chi^2(k_2) \tag{3.16}$$

with

$$\begin{aligned} k_2 &= tr\ ((I-G)D) \\ &= tr\ (I - G - H + GH) \\ &= n - (m+2) - m + m \\ &= n - m - 2\ . \end{aligned}$$

Thus

$$\hat{r}_p^T\ (I - G)\ \hat{r}_p \sim \chi^2\ (n - m - 2) \tag{3.17}$$

and, from equations (3.5), (3.15) and (3.17),

$$s'(\omega) \sim \frac{1}{1 + \chi^2(n-m-2)/\ \chi^2(2)} \tag{3.18}$$

or

$$s'(\omega) \sim \frac{1}{1 + \frac{(n-m-2)}{2}\ F(n-m-2,2)} \tag{3.19}$$

where $F(\nu_1, \nu_2)$ denotes the Fisher distribution with ν_1 and ν_2 degrees of freedom. We can thus test the hypothesis that $f \sim N_n(0, \sigma^2 I)$ at the $100(1 - \alpha)\%$ confidence level by computing the critial value

$$c_{s'}(\omega) = 1/(1 + \frac{n-m-2}{2} F(n-m-2,2,\alpha)) \qquad (3.20)$$

(using the lower tail value of F because of the inverse relationship in equation (3.19)). If $S(\omega)$ is greater than $c_{s'(\omega)}$ we can reject the hypothesis that $f \sim N_n(0, \sigma^2 I)$.

Since (Rao, 1965)

$$F(\nu_1, \nu_2, \alpha) = (F(\nu_2, \nu_1, 1 - \alpha))^{-1} \qquad (3.21)$$

and also

$$F(2, \nu, 1 - \alpha) = \tfrac{1}{2}(\frac{\nu}{\alpha^{2/\nu}} - \nu) \qquad (3.22)$$

we have, with $\nu = n - m - 2$,

$$c_{s'(\omega)} = (1 + (\alpha^{-2/\nu} - 1)^{-1})^{-1} \quad . \qquad (3.23)$$

Furthermore we can determine the expected response of the spectrum as follows. Defining

$$\chi_1^2 = \hat{f}_p^T G \hat{f}_p \sim \chi^2(2)$$

and

$$\chi_2^2 = \hat{f}_p^T (I - G)\hat{f}_p \sim \chi^2(n - m - 2)$$

we have

$$s'(\omega) = \frac{\chi_1^2}{\chi_1^2 + \chi_2^2} \quad . \qquad (3.24)$$

Thus (Rao, 1965; #17, p175; (3b.1.12), p145 and 3a.3, p136) $s'(\omega)$ has a beta distribution with parameters γ and δ given by

$$\gamma = 1$$
$$\delta = (n-m-2)/2$$

and the expected value, or mean, of this distribution (ie the expected response of $s(\omega)$ to $f \sim N_n(0, \sigma^2 I)$) is

$$E\ (s'(\omega)) = \frac{\gamma}{\gamma + \delta} = \frac{1}{1 + \frac{n-m-2}{2}}$$

or

$$E\ (s'(\omega)) = \frac{2}{n-m} \tag{3.25}$$

which agrees with Vanicek (1971; p23).

As examples the following table of values of $100.c_{s'(\omega)}$ for α=0.05 and 100.E $(s'(\omega))$ for various (n-m-2) is given.

(n-m-2)	$100.c_{s'(\omega)}$	100.E $(s'(\omega))$
1	99.75	66.67
2	95.00	50.00
5	69.83	28.57
10	45.07	16.67
20	25.89	9.09
50	11.29	3.85
100	5.82	1.96
200	2.95	0.99
500	1.19	0.40
1000	0.60	0.20
2000	0.30	0.10
5000	0.12	0.04
10000	0.06	0.02

IV CONCLUDING REMARKS

We have derived a statistical criterion for testing the significance of peaks in the least squares spectrum; the 100(1-α)% critical value is given in a simple form by equation (3.23).

We note that the critical value, for a given significance level α, is a function only of the degrees of freedom of the least squares estimation of equation (2.5) which is intuitively pleasing. The longer the time series the more significant are small peaks in the spectrum from a statistical point of view.

V ACKNOWLEDGEMENTS

The author wishes to express his gratitude to Dr. Petr Vanicek for his many helpful discussions and suggestions concerning this work. Also financial support made available by the National Research Council of Canada (Research Agreement A8300) is gratefully acknowledged. Mrs. Theresa Pearce is acknowledged for her excellent typing.

REFERENCES

JEUDY, L. (1980) *Theorie du Gyrocompas Suspendu et ses Applications a la Geodesie*. Ph.D. Thesis, Dept. de Geodesie et de Cartographie, Universite Laval.

MIKHAIL, E. (1976) *Observations and Least Squares*. IEP, Dun Donnelly, N.Y.

RAO, C.R. (1965) *Linear Statistical Inference and Its Applications*. John Wiley and Sons, Inc.

RAO, C.R. and MITRA, (1971) *Generalized Inverse of Matrices and Its Applications*. John Wiley & Sons Inc.

TAYLOR and HAMILTON (1972) *Some Tests of the Vanicek Method of Spectral Analysis*. Astrophysics and Space Science 17, pp 357-367.

VANICEK, P. (1971) *Further Development and Properties of the Spectral Analysis by Least Squares*. Astrophysics and Space Science 12, pp 10-33.

WELLS, D.E. and VANICEK P. (1978) *Least Squares Spectral Analysis*. Bedford Institute of Oceanography, Dartmouth, N.S., Canada, Report Series BI-R-78-8, December.

APPENDIX: SOME MATRIX IDENTITIES

We show here some matrix identities used in the developments of sections 2 and 3. Identities (A.3) and (A.4) are particularly useful and are used in deriving equation (A.14) which simplifies arguments concerning the derivation of certain probability density functions in section 3.

We consider the square, non-singular, symmetric matrix

$$Q = \begin{bmatrix} N_{NN} & N_{NS} \\ N_{SN} & N_{SS} \end{bmatrix} \tag{A.1}$$

with both N_{NN} and N_{SS} non-singular.

The inverse of Q may then be written (Mikhail, 1976; Appendix A)

$$Q^{-1} = \begin{bmatrix} (N_{NN}-N_{NS}N_{SS}^{-1}N_{SN})^{-1} & -N_{NN}^{-1}N_{NS}\;(N_{SS}-N_{SN}N_{NN}^{-1}N_{NS})^{-1} \\ -N_{SS}^{-1}N_{SN}(N_{NN}-N_{NS}N_{SS}^{-1}N_{SN}) & (N_{SS}-N_{SN}N_{NN}^{-1}N_{NS})^{-1} \end{bmatrix} . \tag{A.2}$$

Also

$$(N_{NN}-N_{NS}N_{SS}^{-1}N_{SN})^{-1} = N_{NN}^{-1} + N_{NN}^{-1}N_{NS}(N_{SS}-N_{SN}N_{NN}^{-1}N_{NS})^{-1}\;N_{SN}N_{NN}^{-1} \tag{A.3}$$

and

$$N_{SS}^{-1}N_{SN}(N_{NN}-N_{NS}N_{SS}^{-1}N_{SN})^{-1} = (N_{SS}-N_{SN}N_{NN}^{-1}N_{NS})^{-1}\;N_{SN}N_{NN}^{-1} . \tag{A.4}$$

We now consider the product

$$GH = A(A^TA)^{-1}\;A^TA_N\;\;(A_N^TA_N)^{-1}\;A_N^T \tag{A.5}$$

where $G = A(A^TA)^{-1}\;A^T$, $H = A_N(A_N^T\;A_N)^{-1}\;A_N^T$, $A = [A_N \;\vdots\; A_S]$ with $\dim A_N = (n,\; m)$, $\dim A_S = (n,\; p)$ where $(m + p) < n$ and $\text{rank } A = \text{rank } G = (m + p)$, $\text{rank } A_N = \text{rank } H = m$ (and $\text{rank } A_S = p$). Using the notation of section 2 we have

$$A^T A_N = \begin{bmatrix} A_N^T A_N \\ A_S^T A_N \end{bmatrix} = \begin{bmatrix} N_{NN} \\ N_{SN} \end{bmatrix} \tag{A.6}$$

and

$$A^T A_N (A_N^T A_N)^{-1} = \begin{bmatrix} N_{NN} \\ N_{SN} \end{bmatrix} N_{NN}^{-1} = \begin{bmatrix} I \\ N_{SN} \; N_{NN}^{-1} \end{bmatrix}. \tag{A.7}$$

Using this result we have

$$A^T A_N (A_N^T A_N)^{-1} \; A_N^T = \begin{bmatrix} A_N^T \\ N_{SN} \; N_{NN}^{-1} \; A_N^T \end{bmatrix} \tag{A.8}$$

Using equation (A.2) for the inverse of $A^T A$ we have

$$(A^T A)^{-1} \; A^T A_N \; (A_N^T A_N)^{-1} \; A_N^T =$$

$$\begin{bmatrix} \{(N_{NN} - N_{NS} N_{SS}^{-1} N_{SN})^{-1} \; - N_{NN}^{-1} N_{NS} (N_{SS} - N_{SN} N_{NN}^{-1} N_{NS})^{-1} \; N_{SN} N_{NN}^{-1} \} \; A_N^T \\ \{-N_{SS}^{-1} N_{SN} (N_{NN} - N_{NS} N_{SS}^{-1} N_{SN})^{-1} + (N_{SS} - N_{SN} N_{NN}^{-1} N_{NS})^{-1} \; N_{SN} N_{NN}^{-1} \} A_N^T \end{bmatrix}. \tag{A.9}$$

But according to equations (A.3) and (A.4) equation (A.9) becomes

$$(A^T A)^{-1} \; A^T A_N \; (A_N^T A_N)^{-1} \; A_N^T = \begin{bmatrix} N_{NN}^{-1} \; A_N^T \\ 0 \end{bmatrix} \tag{A.10}$$

and we have finally, premultiplying equation (A.10) by A,

$$A(A^T A)^{-1} \; A^T A_N \quad (A_N^T A_N)^{-1} \quad A_N^T = [A_N \quad A_S] \begin{bmatrix} N_{NN}^{-1} \; A_N^T \\ 0 \end{bmatrix} \tag{A.11}$$

$$= A_N^T \quad (A_N^T A_N)^{-1} \; A_N^T \quad . \tag{A.12}$$

That is $\quad GH = H \qquad$ (A.13)

(and rank (GH) = rank H = rank A_N = m).

Furthermore we note that $G^T = G = GG$ and $H^T = H = HH$, ie G and H are symmetric and idempotent. It is not difficult to show then that any sequence of products of the matrices G and H results in H. For example $GHGHGH = HGGHG = GHGH = HGH = GHG = HG = GH = H$ and thus the rank of any of these products is m. (A.14)