Working Paper 99-19
Statistics and Econometrics Series 04
February 1999

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (34-91) 624-9849

DISTRIBUTIONAL ASPECTS IN PARTIAL LEAST SQUARES REGRESSION.

Rosario Romera*

Abstract

This paper presents some results about the asymptotic behaviour of the estimate of a regression model obtained by Partial Least Squares (PLS) Methods. Because the nonlinearity of the regression estimator on the response variable, local linear approximation through the δ-method for the PLS regression vector is carried out. A new implementation of the PLS algorithm is developed for this purpose.

Keywords: Partial Least Squares Regression; Asymptotic Distribution.

*Romera, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid. C/ Madrid, 126 28903 Getafe -Madrid-. Spain. Ph: 34-91-624.98.64, Fax: 34-91-624.98.49, e-mail: mrromera@est-econ.uc3m.es.

DISTRIBUTIONAL ASPECTS IN PARTIAL LEAST SQUARES REGRESSION

Rosario Romera
Department of Statistics and Econometrics
Universidad Carlos III de Madrid
c/ Madrid 126. 28903 Getafe. (Madrid). Spain
mrromera@est-econ.uc3m.es

ABSTRACT

This paper presents some results about the asymptotic behaviour of the estimate of a regression model obtained by Partial Least Squares (PLS) Methods. Because the nonlinearity of the regression estimator on the response variable, local linear approximation through the δ-method for the -PLS regression vector is carried out . A new implementation of the PLS algorithm is developed for this purpose.

INTRODUCTION

Partial Least Squares (PLS) regression methods were initially proposed by Herman Wold (1966, 1975) in algorithmic form as a modification of the NIPALS algorithm for computing Principal Components. In Wold (1984) there is a survey of PLS methods with emphasis on Social and Economic Sciences. PLS is also heavily promoted and used by chemometricians as the most popular regression method in presence of high degree of collinearity. It is pointed out by the increasing attention paid to PLS in specialised journals; see Frank and Friedman (1993) for a thorough survey. From the field of Marketing Research there is also potential interest in this PLS methodology, see Fornell and Bookstein (1982). Although this increasing interest from applied fields, there is a little research about the statistical properties of PLS methods. Some aspects of the mathematical and statistical structure are in Höskuldsson (1988), Helland (1988, 1992), Manne (1987) and Denham (1997). For geometric aspects of PLS see Goutis (1996) and Phatak and De Jong (1997).

The aim of this paper is to present some asymptotic results of the PLS regression estimator. Because the high nonlinearity of the PLS regression vector, there is no exact distribution available in the PLS literature for this biased estimator. Standard regression theory based in gaussian errors is not suitable in this case. Although there is some inference for the regression vector (also for the predictor) in Phatak, Reilly and Penlidis (1992) and Denham (1997). They are based in several approaches such as bootstrapping or linearization techniques. Linearization in both cases is carried out around an initial value, estimated by fitted values from a PLS regression. This paper presents a linearization of the regression vector around the population vector of the covariances between response and predictors and covariances between predictors. Maximum likelihood estimator for each component of this vector is obtained directly from the sample and not previous PLS regression is needed. Then we obtain the regression vector as a transformation which is differentiable at this population parameter vector. Assuming the multinormal model we can use distributional theory for the covariance matrix. For this purpose we develop a new expression of the PLS algorithm. Consistency of the PLS regression vector follows directly.

The paper is organised as follows. Section 1 describes the PLS model in the regression context. Section 2 introduces the new version of the population PLS algorithm. Sample PLS algorithm is in Section 3. Strong consistency for the PLS estimator is obtained. In Section 4 approximate and asymptotic variances of the regression vector are constructed and asymptotic normality holds. Conclusions and extensions are in section 5.

1. MODEL AND SET-UP

We consider the prediction of a univariate random variable y from a p-variate random variable x with

$$E[y] = \mu_y, Var[y] = \sigma_y^2, \quad (1)$$

$$E[x] = \mu_x, Var(x) = \Sigma_x \quad (2)$$

In the linear prediction context under quadratic loss, the best linear predictor is given by

$$\hat{y} = \mu_y + \beta'(x - \mu_x) \text{ with } \beta = \Sigma_x^{-1}\sigma_{xy} \quad (3)$$

where σ_{xy} is the p-vector of the covariance between the response and the predictor variables.
If p is relatively large, there is often a high degree of collinearity among the predictors. It is well known that in this case, ordinary least squares can be considerably improved by biased regression methods based on a small number of linear combinations of the predictor variables.

In the general form, this regression vector can be written as

$$\beta = R(R'\Sigma_x R)^{-1} R'\sigma_{xy} \quad (4)$$

for some matrix R defining the subspace onto which x is projected. Such subspaces are considered as "components" in the general regression context with high collinearity. Regression methods suitable for this case are Principal Components Regression, Ridge Regression and PLS.

As in Naes and Martens (1985) we consider a decomposition of x onto two orthogonal subspaces and we define a set of components to be "irrelevant" (for prediction purposes) if they have no correlation with y and no correlation with the other subspace. More precisely, we state the decomposition of x into

$$x = R\,z + U\,v \quad (5)$$

where R and U are non-random matrices verifying R'U = 0, and with cov(v, y) = 0 and cov (z, v) = 0.

This implies that there are eigendirections that can be ignored for best linear prediction of y. After the spectral decomposition of Σ_x we obtain

$$\beta_{OLS} = \Sigma_x^{-1}\sigma_{xy} = \sum_{i=1}^{r} \lambda_i^{-1} e_i e_i' \sigma_{xy} \tag{6}$$

where r is the rank of $\Sigma_x; \lambda_i$ are the positive eigenvalues of the eigendecomposition and e_i are the corresponding eigenvectors. So we can consider

$$\beta_{OLS} = \Sigma_{i=1}^{m} \lambda_i^{-1} e_i e_i' \sigma_{xy} \qquad \text{and } e_i' \sigma_{xy} = 0 \tag{7}$$

for i = m+1,..., r. Then the subspace spanned by (e_1, ..., e_m) is exactly the spanned by R. This is the m-relevant components model.

One of the most well-known methods for regression with high collinearity is Principal Components Regression (PCR). A population version of PCR would choose the $e_i's$ corresponding to the largest eigenvalues $\lambda_i's$ and then we have $\beta^{\alpha}_{PCR} = \sum_{i=1}^{\alpha} \lambda_i^{-1} e_i e_i' \sigma_{xy}$.
The population version of PLS chooses (e_1, ..., e_m) via the Krylov sequence

$$\sigma_{xy}, \Sigma_x \sigma_{xy}, \Sigma_x^2 \sigma_{xy} \ldots \tag{8}$$

This method has a clear advantage over PCR because we do not need to care about which components should be included, only the number of components must be determined.

Maximum Likelihood PLS-regression estimates under the m-relevant components model, can be found in Helland (1992), and expressions of the asymptotic prediction Minimum Squared Error with *m*-relevant components are in Helland and Almoy (1994). All these results assumed that the number of components relevant for prediction *m* is known. In practice it is not, and it is commonly selected from the data by cross-validation methods or leverage corrected sum of squares methods. In fact, it is random and some questions about this uncertainty should be proposed. Assuming multinormal distribution for (y,x), likelihood ratio tests have been obtained for the *m*-relevant hypothesis in Höskuldsson (1988) and in Helland (1992).

In practice, from the nx(1+p) matrix of data (y, X) we estimate

$$\hat{\Sigma}_x = S_x, \hat{\sigma}_{xy} = s_{xy} \tag{9}$$

and the PLS regression vector $\beta_{PLS}{}^{\alpha}$ is estimated by

$$\hat{\beta}_{PLS}{}^{\alpha} = \hat{R}(\hat{R}' S_x \hat{R})^{-1} \hat{R}' s_{xy} \qquad \text{for } 1 \le \alpha \le \text{m} \tag{10}$$

$$\text{where } \hat{R} = (s_{xy}, S_x s_{xy}, \ldots, S_x^{\alpha-1} s_{xy}) \tag{11}$$

with α determined empirically.

2. A POPULATION PLS ALGORITHM

2.1. ALGORITHM

In this section we give an alternative formulation to the population version of the PLS algorithm. The goal is to obtain scores, loadings and residuals as recursive expressions of the initial covariance matrix Σ_x and vector σ_{xy}.

In the canonical PLS version (see Helland (1990) or Stone and Brooks (1990)), we construct for each step α the bilinear representation.

$$x = \mu_x + p_1 t_1 + \ldots + p_\alpha t_\alpha + e_\alpha$$
$$y = \mu_y + q_1 t_1 + \ldots + q_\alpha t_\alpha + f_\alpha$$

following the PLS principles and starting with $x - \mu_x = e_0$, $y - \mu_y = f_0$
For i =1,2,... we have

- scores (t_i) are linear combinations of the last step x-residuals with weights selected as covariances with the last step y-residuals

let $\Sigma_i = \text{var}(e_i)$, $\sigma_i = \text{cov}(e_i, f_i)$
$t_{i+1} = e_i' \sigma_i$
it follows

var (t_{i+1})= $\sigma_i' \Sigma_i \sigma_i$ and cov(e_i, t_{i+1}) = $\Sigma_i \sigma_i$

- loadings (p_i and q_i) are obtained by OLS regression of the last step residuals on t_i

$p_{i+1} = (\sigma_i' \Sigma_i \sigma_i)^{-1} \Sigma_i \sigma_i$

$q_{i+1} = (\sigma_i' \Sigma_i \sigma_i)^{-1} \sigma_i' \ \sigma_i$

- residuals are constructed as

$e_{i+1} = e_i - p_{i+1} t_{i+1}$
$f_{i+1} = f_i - q_{i+1} t_{i+1}$

ALGORITHM

Let introduce at each PLS step the pxp-matrix A_i in order to reformulate the PLS algorithm as follows:

starting values: $e_o = x - \mu_x$, $f_0 = y - \mu_y$

covariances: $\sigma_0 = \text{cov}(e_0, f_0)$, $\Sigma_0 = \text{var}(e_0)$

first score: $t_1 = e_0' \sigma_0$

notice that

$\mathrm{cov}(e_0, t_1) = \mathrm{cov}(e_0, e_0' \sigma_0) = \mathrm{cov}(e_0, \sigma_0' e_0) = \Sigma_0 \sigma_0$
$\mathrm{var}(t_1) = \mathrm{var}[e_0' \sigma_0] = \sigma_0' \Sigma_0 \sigma_0$
$\mathrm{cov}(f_0, t_1) = \mathrm{cov}(f_0, \sigma_0' e_0) = \sigma_0' \sigma_0$

loadings by least squares:

$$p_1 = \frac{\mathrm{cov}(e_0, t_1)}{\mathrm{var}[t_1]} = \Sigma_0\sigma_0 / \sigma_0'\Sigma_0\sigma_0, \qquad q_1 = \frac{\mathrm{cov}(f_0, t_1)}{\mathrm{var}[t_1]} = \sigma_0'\sigma_0 / \sigma_0'\Sigma_0\sigma_0$$

$$e_1 = \left(I - \frac{\Sigma_0\sigma_0\sigma_0'}{\sigma_0'\Sigma_0\sigma_0} \right).e_0 = A_0 e_0$$

$$f_1 = f_0 - \frac{\sigma_0'\sigma_0}{\sigma_0'\Sigma_0\sigma_0}\sigma_0' e_0 = f_0 - b' e_0$$

In general for i = 0, 1,... we obtain

$\sigma_i = \mathrm{cov}(e_i, f_i)$ p-vector and

$\Sigma_i = \mathrm{cov}(e_i, e_i)$ p x p-matrix

- loadings

$$P_{i+1} = \frac{\Sigma_i\sigma_i}{\sigma_i'\Sigma_i\sigma_i}, \qquad q_{i+1} = \frac{\sigma_i'\sigma_i}{\sigma_i'\Sigma_i\sigma_i}$$

- scores (linear combination of previous residual)

$t_{i+1} = e_i' \sigma_i$

- and residuals

$e_{i+1} = A_i e_i$
$f_{i+1} = f_i - b_i' e_i$

with

$$A_i = I - \frac{\Sigma_i\sigma_i\sigma_i'}{\sigma_i'\Sigma_i\sigma_i}$$

$$b_i' = \frac{\sigma_i'\sigma_i}{\sigma_i'\Sigma_i\sigma_i'}$$

Notice that $\sigma_i = A_{i-1}\,\sigma_{i-1}$ and $\Sigma_i = A_{i-1}\,\Sigma_{i-1}\,A_{i-1}'$ are obtained recursively through the A matrices. Unfortunately there is not a clear intuition about the A matrices.

2.2. REMARKS

REMARK 1. Scores t_i, t_j are incorrelated for $i \neq j$

Proof:

Because the properties of the OLS residuals e_i

$\mathrm{cov}\,(t_i, e_i) = 0$ and then
$\mathrm{cov}\,(t_i, t_{i+1}) = \mathrm{cov}\,(t_i, e_i'\,\sigma_i) = \sigma_i'\,\mathrm{cov}\,(t_i, e_i) = 0$

REMARK 2. It holds that $\sigma_i = \mathrm{cov}\,(e_i, f_i) = \mathrm{cov}\,(e_i, y)$

Proof:

Notice that $\mathrm{cov}\,(e_i, f_i) = \mathrm{cov}\,(e_i, f_{i-1} - b_{i-1}'\,e_{i-1}) = \mathrm{cov}\,(e_i, f_{i-1})$

it follows because

$\mathrm{cov}\,(e_i, b_{i-1}'\,e_{i-1}) = A_{i-1}\,\mathrm{cov}\,(e_{i-1}, e_{i-1})\,b_{i-1} = A_{i-1}\,\Sigma_{i-1}\,b_{i-1} = 0$

and applying induction we have

$\mathrm{cov}\,(e_i, f_i) = \mathrm{cov}\,(e_i, f_0)$

REMARK 3. We obtain a recursive expression for σ_i in terms of the A matrices

$$\sigma_i = \mathrm{cov}\,(e_i, f_0) = \mathrm{cov}\,(A_{i-1}\,e_{i-1}, f_0) = A_{i-1}\,A_{i-2}\ldots\,\mathrm{cov}\,(e_0, f_0) = \prod_{j=0}^{i-1} A_j\sigma_0$$

REMARK.4 We also obtain a useful recursive expression for $\sigma_{T,y} = \mathrm{cov}\,(t_i, y)$

$$\sigma_{T,y} = \mathrm{cov}\,(t_i, y) = \mathrm{cov}\,(t_i, f_0) = \mathrm{cov}\,(e'_{i-1}\,\sigma_{i-1}, f_0) = \sigma'_{i-1}\,\mathrm{cov}\,(e_{i-1}, f_0) = \sigma'_{i-1}\prod_{j=0}^{i-2} A_j\sigma_0$$

For the population version of PLS we have the following result

<u>Proposition 1</u>

For each step α, the PLS regression vector is given by

$$\beta_{PLS}^{\alpha} = \sum_{i=0}^{\alpha} \frac{\sigma_i' A_{i-1}\ldots A_0\sigma_0}{\sigma_i'\Sigma_i\sigma_i} A_0\ldots A_{i-1}\sigma_0 \qquad (12)$$

Proof

Let introduce γ_{PLS}^{α} the same vector as β_{PLS}^{α} but in terms of the t-variables rather than x-variables. After (4) with weights R= $W_\alpha = [w_1, \dots w_\alpha]$, where $w_i = \text{cov}(e_i, f_0) = \sigma_i$ we obtain

$$\beta_{PLS}^{\alpha} = W_\alpha (W_\alpha' \Sigma_x W_\alpha)^{-1} W_\alpha' \sigma_{xy} = W_\alpha \Sigma_T^{-1} \sigma_{T,y}$$

Σ_T is the diagonal matrix of cov (t_i, t_j) and $\sigma_{T,y}$ is the vector of cov (t_i, y)

For α steps we define $\gamma_{PLS}^{\alpha} = \Sigma_T^{-1} \sigma_{T,y}$. Notice that this vector contains the PLS loadings for the response variable up to step α

$$\gamma_{PLS}^{\alpha} = \begin{pmatrix} \sigma_0' \Sigma_0 \sigma_0 & 0 \\ 0 & 0 \\ 0 & \sigma_{\alpha-1}' \Sigma_{\alpha-1} \sigma_{\alpha-1} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_0' \sigma_0 \\ \sigma_1' A_0 \sigma_0 \\ \sigma_{\alpha-1}' A_{\alpha-2} \dots A_0 \sigma_0 \end{pmatrix} = \begin{pmatrix} \dfrac{\sigma_0' \sigma_0}{\sigma_0' \Sigma_0 \sigma_0} \\ \\ \dfrac{\sigma_{\alpha-1}' A_{\alpha-2} \dots A_0 \sigma_0}{\sigma_{\alpha-1}' \Sigma_{\alpha-1} \sigma_{\alpha-1}} \end{pmatrix} = \begin{pmatrix} q_1 \\ . \\ q_\alpha \end{pmatrix}$$

The following relationship then holds

$$\beta_{PLS}^{\alpha} = W_\alpha \gamma_{PLS}^{\alpha} \tag{13}$$

Following Remark 3

$$\beta_{PLS}^{\alpha} = (\sigma_0 \quad A_0 \sigma_0 \dots \prod_{i=0}^{\alpha-2} A_i \sigma_0) \ . \ \gamma_{PLS}^{\alpha} = \sum_{i=0}^{\alpha} \frac{\sigma_i' A_{i-1} \dots A_0 \sigma_0}{\sigma_i' \Sigma_i \sigma_i} A_0 \dots A_{i-1} \sigma_0$$

where for i=0 the matrix A_{-1} . A_0 is taken equal to the I_{pxp} (identity matrix). ■

REMARK 5. All the components in β_{PLS}^{α} vector for each α, are obtained as continuous functions of σ_i' s vectors and Σ_i's matrices with i = 0, 1,... α.

REMARK 6. We use the fact that $T_a = Xw_a$, based on the equivalence between PLS formulations. It has been proved by Helland (1988), theorem 2.1. Then it follows for the covariances

$$\Sigma_T = W_\alpha' \Sigma_x W_\alpha \text{ and } \quad \sigma_{T,y} = W_\alpha' \sigma_{xy}$$

REMARK 7. Notice that β_{PLS}^{α} should be equal to $\beta_{OLS} = \Sigma_x^{-1} \sigma_{xy}$ if all relevant components are included but it would be not if α<m.

3. SAMPLE PLS ALGORITHM. ESTIMATION

The sample version of the PLS algorithm in section 2.1. is obtained replacing the population values by their estimators. The covariances are then estimated by

$$\hat{\sigma}_{xy} = s_{xy} = x'y / n - 1$$

$$\hat{\Sigma}_x = S_x = X'X / n - 1$$

with X and y centred. For large n, consistency (strong) follows by direct application of the law of Large Numbers to the Method of Moments estimators.

By the same procedure we estimate the covariances of the residuals Σ_i, σ_i by sample covariances S_i, s_i, for each step $i = 0 \ldots \alpha$. After α steps of the PLS algorithm we obtain the regression vector

$$\hat{\beta}^{\alpha}_{PLS} = \hat{W}_{\alpha}(\hat{W}_{\alpha}' S_x \hat{W}_{\alpha})^{-1} \hat{W}_{\alpha} s_{xy}$$

REMARK 8. Notice That Proposition 1 gives expressions to be used to see the asymptotic bias of $\hat{\beta}^{\alpha}_{PLS}$ if fewer than necessary steps are chosen.

Proposition 2

For any α

$$\lim_{n \to \infty} \hat{\beta}^{\alpha}_{PLS} = \beta^{\alpha}_{PLS} \quad \text{a.s} \tag{14}$$

Proof

As in Proposition 1, let introduce $\hat{\gamma}^{\alpha}_{PLS}$ as the sample version of γ^{α}_{PLS}. So we estimate

$$\hat{\gamma}^{\alpha}_{PLS} = \begin{pmatrix} \hat{q}_1 \\ \\ \hat{q}_{\alpha} \end{pmatrix} \tag{15}$$

and it holds

$$\hat{\beta}^{\alpha}_{PLS} = \hat{W}_{\alpha} \hat{\gamma}^{\alpha}_{PLS} \tag{16}$$

Population and sample algorithms give similar expressions for loadings, so the strong consistency argument for sample covariances transformed by continuous functions gives strong convergence for each component, so it holds

$$\lim_{n \to \infty} \hat{\gamma}^{\alpha}_{PLS} = \gamma^{\alpha}_{PLS} \quad \text{a.s} \tag{17}$$

By the other hand we estimate consistently W_{α} weights by $\hat{W}_{\alpha} = [\hat{w}_1 \ldots \hat{w}_{\alpha}]$ replacing $cov(e_i, f_i) = \sigma_i$ by $E_i' g_i$ (s_i off 1/n), where E_i are the residuals in the sample PLS algorithm.

The strong Law of Large Numbers insures

$1/n\ E_i\ g_i \to \sigma_i$ a.s (see details in Appendix 1)

and strong consistency holds

$$\lim_{n \to \infty} \hat{\beta}^{\alpha}_{PLS} = \lim_{n \to \infty} \hat{W}_{\alpha} \lim_{n \to \infty} \hat{\gamma}^{\alpha}_{PLS} = W_{\alpha} \gamma^{\alpha}_{PLS} = \beta^{\alpha}_{PLS} \quad \text{a.s}$$ ■

4. ASYMPTOTICS

Our next question is if something more can be said about the asymptotic distribution of the PLS regression vector $\hat{\beta}^{\alpha}_{PLS}$ for any α; specially if fewer steps than m are chosen. We first search the asymptotic variance of the regression estimator.

In what follows we assume that the y,x variables have a joint multinormal distribution

$N\left(\underline{0}, \begin{pmatrix} \sigma_y^2 & \sigma_{xy}' \\ \sigma_{xy} & \Sigma_x \end{pmatrix}\right)$. Let denote $\Sigma = \left(\begin{pmatrix} \sigma_y^2 & \sigma_{xy}' \\ \sigma_{xy} & \Sigma_x \end{pmatrix}\right)$

Notice that under multinormal hypothesis, sample covariances are the Maximum Likelihood (ML) estimators of the population covariances. By the invariance principle the ML estimators of γ^{α}_{PLS} and β^{α}_{PLS} are $\hat{\gamma}^{\alpha}_{PLS}$ and $\hat{\beta}^{\alpha}_{PLS}$ respectively which are asymptotically unbiased. They also are strong consistent as is proved in Proposition 2.

Results for the PLS regression vector on t-variables, could be extended to $\hat{\beta}^{\alpha}_{PLS}$. We will obtain first asymptotic results for $\hat{\gamma}^{\alpha}_{PLS}$.

In the sequel we will omit the PLS subscript in the regression vectors, so γ^{α} and β^{α} will stand for γ^{α}_{PLS} and β^{α}_{PLS}, as well as in the sample case.

In subsection 4.1 we introduce the vector notation for covariance matrices and the principles used to obtain the asymptotic variance for the regression vector on the t-variables. In section 4.2 we obtain the asymptotic distribution for this estimator

4.1 Notation and Methodology

4.1.1. Notation.

We stacked columnwise the elements of the covariance vector σ_{xy} and the covariance matrix Σ_x. Let denote $r = p + p\,(p + 1)/2$ and

$B^0 = vec\,(\sigma_{xy}, \Sigma_x) = vec\,(\sigma_0, e_0)$.

Then B^0 is the r-dimensional vector which contains once all the covariances between response and predictor variables, and between predictor variables

$$B^0 = (\sigma_{1y}\ \sigma_{2y} \ldots\ \sigma_{py}\ \rho_{11}\ \rho_{12} \ldots \rho_{1p}\ \rho_{22} \ldots\ \rho_{2p} \ldots \rho_{pp})' \qquad (18)$$

where $\sigma_{jy} = cov\,(x_j\ y)$ and $\rho_{jk} = cov\,(x_j\ x_k)$, $1 \leq j, k \leq p$

Let denote by b^0 the same thing as B^0 but in the sample. Notice that b^0 is a strong consistent estimator of B^0, provided the Strong Law of Large Numbers (Proposition 2).

Now we define for the each step i of the population PLS algorithm with $1 \leq i \leq \alpha$
the r-vector

$B^i = vec\,(\sigma_i, \Sigma_i)$

where σ_i, Σ_i are covariances of the residuals defined as in section 2

$\sigma_i = cov\,(e_i, y)$ and $\Sigma_i = cov\,(e_i, e_i)$

Let the r-vector b^i be the same thing as B^i but in the sample,

$b^i = vec\,(s_i, S_i)$

where $s_i = E_i'\,g_i$ and $S_i = E_i'\,E_i$ (off 1/n)

By the multivariate Cramer central limit theorem, we know that the sample covariance matrix is asymptotically normally distributed as the sample size. Then we know the mean and the covariance matrix of the b^i vector.

4.1.2 Methodology

Let define the vector function $\gamma^\alpha : R^r \rightarrow R^\alpha$ in such a way that
$\hat{\gamma}^\alpha = \gamma^\alpha(b^0)$
and similar consideration for the B^0 parameter vector
$\gamma^\alpha = \gamma^\alpha(B^0)$
where α is the number of the PLS components.

Notice that this function is a transformation totally differentiable at B^0. It is clear in the Proof of Proposition 1.

We apply the δ-method to $\gamma^{\alpha}(b^0)$ and we use linearization around the population parameter vector B^0. For better understanding of the δ-method see Barndorff-Nielsen and Cox (1989).

So we have the first-order approximation

$$\gamma^{\alpha}(b^0) \approx \gamma^{\alpha}(B^0) + J_{B0}(b^0 - B^0) \tag{19}$$

where J_{B^0} stands for the Jacobian J evaluated at the B^0 vector. The Jacobian J is a αx r- dimensional matrix where the elements are the partial derivatives of the α elements of γ^{α} with respect to the r elements of b^0.

After (19) we obtain the approximate covariance matrix

$$\mathrm{Var}\,[\gamma^{\alpha}(b^0)] \approx J_{B^0}\,\mathrm{Var}\,[b^0]\,J_{B^0}{}' \tag{20}$$

and then each component can be approximate by

$$\mathrm{cov}(\hat{q}_i, \hat{q}_j) \approx \left(\frac{\partial q_i}{\partial B^0}\right)' Var[b^0]\left(\frac{\partial q_j}{\partial B^0}\right), \qquad i,j = 1 \ldots \alpha \tag{21}$$

4.2 MAIN RESULTS

Our task is now to develop separated expressions for the variance and the partial derivatives in (21). We use statistical multivariate methods for multinormal sampling and differential techniques for matrices and vectors. We will obtain first the covariance matrix $\mathrm{Var}[b^0]$ applying Proposition 3.

Proposition 3

Let $Z_{nxp} \sim N(0, \Sigma)$ where $\Sigma = (\sigma_{ij})$, let $A = Z'Z$. It is known that $A \sim W_p(n, \Sigma)$.
We know already the moments of the elements of this Wishart matrix

$E\,[A] = n\,\Sigma$
$\mathrm{cov}\,(a_{ij}, a_{kl}) = n\,(\sigma_{ik}\,\sigma_{jl} + \sigma_{il}\,\sigma_{jk})$ for i,j,k,l = 1…p

The matrix of covariances between elements of A can be expressed in terms of the Kronecker product

$K = \sum_{i,j=1\ldots p} H_{ij} \otimes H'_{ij}$ where H_{ij} is the pxp matrix with $h_{ij}=1$ and all the other elements zero

then the covariance matrix of vec(A) is

$$\mathrm{cov}(\mathrm{vec}(A)) = n\,(I_{p2} + K)\,(\Sigma \otimes \Sigma) \tag{22}$$

Finally under general conditions the sample covariance matrix S(n) formed from a sample of size n+1 is asymptotically normal as $n \to \infty$.
In the case of normal sampling S(n) is $W_p(n, (1/n)\Sigma)$ so the asymptotic distribution of

$$n^{1/2}\,[\mathrm{vec}(S(n))-\mathrm{vec}(\Sigma)] \text{ is } N_p[0, (I_{p2}+K)(\Sigma \otimes \Sigma)] \tag{23}$$

Proof of these results are in Muirhead 82, Anderson 84 or Magnus and Neudecker 88.

REMARK 9. Notice that expression (22) give us the approximate variance for the PLS regression vector on the t-variables constructed from a random sample of size n of the (y,x) vector-variable. The asymptotic variance of b^0 can be derived after (23) under the multinormal sampling hypothesis.

JACOBIAN

We will derive a recursive expression for the r-column vector of partial derivatives in (21).

Applying the Chain Rule, we obtain

$$\frac{\partial q_i}{\partial B^0} = \frac{\partial q_i}{\partial B^{i-1}} \frac{\partial B^{i-1}}{\partial B^{i-2}} \cdots \frac{\partial B^2}{\partial B^1} \frac{\partial B^1}{\partial B^0}, \quad i = 1\ldots\alpha \tag{24}$$

Because the recursive feature of (24) we observe only two types of terms. We then apply techniques of Algebra of Matrices and we first find the r-column vector

$$\frac{\partial q_i}{\partial B^{i-1}} = \begin{pmatrix} \dfrac{\partial q_i}{\partial \sigma_{i-1}} \\ \dfrac{\partial q_1}{\partial \Sigma_{i-1}} \end{pmatrix}, \tag{25}$$

Using expressions from section 2.1 we have

$$q_i = \left(\sigma_{i-1}^{'}\Sigma_{i-1}\sigma_{i-1}\right)^{-1}\sigma_{i-1}^{'}\sigma_i, \quad p_i = (\sigma_{i-1}^{'}\Sigma_{i-1}\sigma_{i-1}^{-1})\Sigma_{i-1}\sigma_{i-1}$$

and, omitting subscripts, we obtain the p-vector.

$$\frac{\partial q}{\partial \sigma} = \frac{\partial}{\partial \sigma}\left(\frac{\sigma^{'}\sigma}{\sigma^{'}\Sigma\sigma}\right) = \frac{\partial}{\partial \sigma}(\sigma^{'}\sigma).\frac{1}{\sigma^{'}\Sigma\sigma} + \frac{\partial}{\partial \sigma}\left(\frac{1}{\sigma^{'}\Sigma\sigma}\right)\sigma^{'}\sigma = \frac{2}{\sigma^{'}\Sigma\sigma}\sigma - \frac{\sigma^{'}\sigma}{(\sigma^{'}\Sigma\sigma)^2}2\Sigma\sigma =$$

$$= -2q\left[I - \frac{\sigma\sigma^{'}}{\sigma^{'}\sigma}\right]p \tag{26}$$

Analogously the p(p+1)/2 vector of partial derivatives

$$\frac{\partial q}{\partial \Sigma} = -\frac{\sigma'\sigma}{(\sigma'\Sigma\sigma)^2}\cdot\frac{\partial(\sigma'\Sigma\sigma)}{\partial \Sigma} = -\frac{q}{\sigma'\Sigma\sigma}\frac{\partial}{\partial \Sigma}\left(\sum_{k,l}\Sigma_{kl}\sigma_k\sigma_l\right) =$$

$$= -\frac{q}{\sigma'\Sigma\sigma}\cdot\left(\sigma_1^2 2\sigma_1\sigma_2 \ldots 2\sigma_1\sigma_p\sigma_2^2 \ldots 2\sigma_2\sigma_k\sigma_3^2 \ldots \sigma_p^2\right)' \tag{27}$$

Then $\frac{\partial q_i}{\partial B^{i-1}}$ is the vector of the above elements stacked columnwise.

Now we deal with $\frac{\partial B^{i+1}}{\partial B^i}$ because the terms in (24) after the second one have similar expressions. First notice that this is a matrix of the form

$$\begin{pmatrix} \frac{\partial \sigma_{i+1}}{\partial \sigma_i} & \frac{\partial \sigma_{i+1}}{\partial \Sigma_i} \\ \frac{\partial \Sigma_{i+1}}{\partial \sigma_i} & \frac{\partial \Sigma_{i+1}}{\partial \Sigma_i} \end{pmatrix}$$

We have explicit expressions for each block in Appendix 2. There is also information about the computational complexity of these calculus in terms of the order of the matrices operations requested.

Theorem

Let the vector function $\gamma^\alpha : R^r \to R^\alpha$ defined as the PLS regression vector on the t-variables (Proposition 1), and totally differentiable at $B^0 \in R^r$. Let b^0 the sample vector of covariances (as in 4.1.1.)
If $J_{B0} = \nabla\gamma^\alpha(B^0)$ has rank r, then if $n \to \infty$

$$n^{1/2}(\gamma^\alpha(b^0) - \gamma^\alpha(B^0)) \approx N_\alpha(0, \Omega) \tag{28}$$

with $\Omega = J_{B0}((I_{p2} + K)(\Sigma \otimes \Sigma))J_{B0}'$.

Proof

By the multivariate Cramer central limit theorem (see Proposition 3) we have the asymptotic distribution of the sample r-vector of covariances under normal sampling hypothesis

$$n^{1/2}(b^0 - B^0) \approx N_r(0, ((I_{p2} + K)(\Sigma \otimes \Sigma))$$

By the other hand the Jacobian matrix at B^0 exits and can be calculated as in Appendix 2. The nonsingularity of this Jacobian is directly related with the selection of α. This number of components of the PLS regression should not be greater that the dimension of the space generated by the Krylov vector sequence (8). This is the usual criteria in PLS .

Then we apply the Differential Transformation Theorem (see Appendix 3) to γ^{α} and we obtain the asymptotic distribution (28)

$$n^{1/2}(\gamma^{\alpha}(b^0) - \gamma^{\alpha}(B^0)) \approx N_{\alpha}(\,0,\, J_{B0}\,\mathrm{Var}(b^0)\,J_{B0}{}')\,.$$

where $\mathrm{Var}(b^0) \approx (I_{p2} + K)\,(\Sigma \otimes \Sigma)$. ■

5. CONCLUSIONS AND EXTENSIONS

We have obtained theoretical results that confirm what we have observed in simulations with our Matlab Code implementations. We simulated data generated using the real data AVA.VAR kindly sent by Tom Fearn (University College London). The code is available on request. Here we present only asymptotic theoretical research.

For the PLS regression estimator on the t-variables (or PLS-components) and for each $\alpha < m$, asymptotic normality has been proved. Strong consistency holds. Formulas for the asymptotic variance implementation have been developed.

It is obvious that analogous properties for the $\hat{\beta}^{\alpha}_{PLS}$ estimator hold given (16) in Proposition 1. In fact the only difference in the formulas is concerning one term in the Jacobian matrix. More precisely, partial derivatives in (25) change and can be obtained following the methodology in 4.1.2.

There is a very attractive recursive feature for implementation in the PLS algorithm developed here. Through the auxiliary A matrices we have recursive expressions of the PLS regression vector, which can be handled as a function of the initial covariances vector σ_{xy} and matrix Σ_x.

Given the asymptotic results, we could perform inference for the true PLS regression vector, asymptotic interval estimation or hypothesis testing.

We also provide an approximate variance of $\hat{\beta}^{\alpha}_{PLS}$ for small samples. Constructing interval estimates (for the regression vector or for prediction) with small sample sizes, as sometimes is seen in the recent PLS literature, care must be made. The reason is that the PLS regression estimator is clearly biased for small sample sizes.

If the sample size is not too small, resampling techniques such as bootstrapping could be compared with our methodology.

ACKNOWLEDGES

The origin of this paper is in the conversations that I had with Costas Goutis during the last few months when he was visiting my department. We were planing to work together, just before he left us forever.

This research was supported by DGES through grant PB96-0111.

REFERENCES

Anderson, T.W., "An Introduction to Multivariate Statistical Analysis, 2nd ed., John Wiley, New York, 1984.

Barndorff- Nielsen, O.E. and Cox, D.R., "Asymptotic Techniques for use in Statistics", Chapman and Hall, London, 1989.

Denham, M. C., Prediction Intervals in Partial Least Squares, *Journal of Chemometrics*, Vol. 11, (1997), pp. 39-52.

Fornell, Cl. and Bookstein, L., Two Structural Equation Models LISREL and PLS Applied to Consumer Exit- Voice Theory, *Journal of Marketing Research*, **19**, (1982), pp. 440-452.

Frank, I.E. and Friedman, J.H., A statistical view of some chemometrics regression tools (with discussion), *Technometrics* **35**, (1993), pp. 109-147.

Goutis, C., Partial Least Squares Algorithm yields Shrinkage estimators, The Ann. of Statist. (1996), pp.816-824.

Helland, I. S., On the structure of Partial Least Squares Regression, *Comm. Statist. B-Simulation Comput.* **17**, (1988), pp.581-607.

Helland, I. S., Partial Least Squares Regression and Statistical Models, *Scandinavian Journal of Statistics* **17**, (1990), pp.97-114.

Helland, I. S., and Almoy, T. , Comparison of Prediction Methods when only a few components are relevant, *Journal of the American Statistical Association*, **89**, (1994), pp.583-591.

Helland, I. S., Maximum Likelihood Regression on Relevant Components, *Journal of the Royal Statistical Society, Ser. B.* **54**, (1992), pp.637-647.

Höskuldsson, A., PLS regression methods, *Journal of Chemometrics* **2**, (1988), pp. 211-228.

Manne, R. , Analysis of two partial-least-squares algorithms for multivariate calibration, *Chemometrics and Intelligent Laboratory Systems* **1**, (1987), pp. 187-197.

Magnus, J.R. and Neudecker, H., "Matrix Differential Calculus with applications in Statistics and Econometrics", John Wiley, Chichester,1988.

Martens, H. and Naes, T., "Multivariate Calibration". Wiley, New York,1989.

Muirhead, R. B., "Aspects of Multivariate Statistical Theory", John Wiley, New York, 1982.

Naes, T. and Martens, H., Comparison of predictor methods for multicollinear data, *Comm. in Statist. Simulation Comput.* **14**, (1985), pp. 545-576.

Naes, T., and Helland, I.S., Relevant Components in Regression, *Scandinavian Journal of Statistics*, **20**, (1993), pp. 239-250.

Phatak, A., Reilly, P.M. and Penlidis, A., The geometry of 2-blok partial least squares regression, *Comm. Statist. Theory Methods* **21**, (1992), pp. 1517-1553.

Phatak, A: and De Jong, S., The Geometry of Partial Least Squares, *Journal of Chemometrics*, Vol.11, (1997), pp. 31-338.

Stone, M. and Brooks, R.J., Contimnuum regression: cross-validated sequentally constructed prediction embracing ordinary least squares, partial least squares and principal components regression (with discussion), *Journal of the Royal Statistical Society, Ser. B*, **52,** (1990), pp. 237-269.

Wold, H., Nonlinear estimation by iterative least squares procedures.In *Research Papers in Statistics. Festschrift for J. Neyman.* (F.N. David, de.), pp. 411-444.Wiley , New York, 1966.

Wold, H., Soft Modelling by Latent Variables: the Nonlinear Iterative Partial Least Squares Approach". In *Perspectives in Probability and Statistics, Papers in Honour of M.S. Barlett*, (J. Gani, de.). Academic Press, London, 1975.

Wold , H., PLS Regression. In *Enciclopaedia of Statitical Sciences*, Vol. 6. (N.L. Johnson and S. Kotz, eds.), pp. 581-591, John Wiley, New York, 1984.

APPENDIX 1

Proposition

If (X_i, Y_i) , i =1...n is a random sample of (X,Y) with cov (X, Y) < ∞, then the sample covariance is a consistent estimator.

Proof

we constract the estimator of the population covariance by the Method of the Moments

$$\hat{\text{cov}}(X,Y) = \sum_{i=1}^{n} \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n}$$

for the strong Law of Large Numbers if $E[(X-\mu_x)(Y-\mu_x)] < \infty$ we have for $n \to \infty$

$$\Sigma \frac{(X_i - \bar{X})(Y_i - \bar{Y})}{n} \longrightarrow \text{cov}(X,Y) \text{ a.s.}$$

Notice that if X, Y are centered then if $E(XY) < \infty$ is a sufficient condition for the strong consistency of the population covariance estimator.

APPENDIX 2

Expressions for the $\frac{\partial B^{i+1}}{\partial B^i}$ matrix are obtained as follows

$$\frac{\partial B^{i+1}}{\partial B^i} = \begin{pmatrix} \frac{\partial \sigma_{i+1}}{\partial \sigma_i} & \frac{\partial \sigma_{i+1}}{\partial \Sigma_i} \\ \frac{\partial \Sigma_{i+1}}{\partial \sigma_i} & \frac{\partial \Sigma_{i+1}}{\partial \Sigma_i} \end{pmatrix} = \begin{pmatrix} 1) & 2) \\ 3) & 4) \end{pmatrix}$$

1) First block is a p x p matrix

$$\frac{\partial \sigma_{i+1}}{\partial \sigma_i} = \frac{\partial}{\partial \sigma_i}(A_i \sigma_i) = \frac{\partial}{\partial \sigma_i}\left(\sigma_i - \frac{\Sigma_i \sigma_i \sigma_i' \sigma_i}{\sigma_i' \Sigma_i \sigma_i}\right) = I - \frac{\partial}{\partial \sigma_i}(\Sigma_i \sigma_i q_i) = I - \Sigma_i \sigma_i \frac{\partial q_i}{\partial \sigma_i} - \Sigma_i q_i$$

2) Second block is a $px\frac{p(p+1)}{2}$ matrix

$$\frac{\partial \sigma_{i+1}}{\partial \Sigma_i} = \frac{\partial}{\partial \Sigma_i}\left(\sigma_i - \frac{\Sigma_i \sigma_i \sigma_i' \sigma_i}{\sigma_i' \Sigma_i \sigma_i}\right) = -\sigma_i' \sigma_i \frac{\partial}{\partial \Sigma_i}\left(\frac{\Sigma_i \sigma_i}{\sigma_i' \Sigma_i \sigma_i}\right) =$$

$$= -\sigma_i^{'}\sigma_i\left[\frac{1}{\sigma_i^{'}\Sigma_i\sigma_i}\frac{\partial}{\partial\Sigma_i}\Sigma_i\sigma_i - \frac{\Sigma_i\sigma_i}{\left(\sigma_i^{'}\Sigma_i\sigma_i\right)^2}\left(\frac{\partial}{\partial\Sigma_i}\sigma_i^{'}\Sigma_i\sigma_i\right)^{'}\right] =$$

$$= -\sigma_i^{'}\sigma_i\left[\frac{1}{\sigma_i^{'}\Sigma_i\sigma_i}\left(\frac{\partial}{\partial\Sigma_{11}}\Sigma_i\sigma_i \quad \frac{\partial}{\partial\Sigma_{22}}\Sigma_i\sigma_i \ldots \frac{\partial}{\partial\Sigma_{pp}}\Sigma_i\sigma_i\right)\right]$$

$$+\Sigma_i\sigma_i\left(\sigma_1^2 \;\; 2\sigma_1\sigma_2 \;\; 2\sigma_1\sigma_3 \ldots \sigma_2^2 \ldots \sigma_p^2\right) \Big] =$$

$$=$$

$$-q_i\begin{pmatrix} \sigma_1\sigma_2\ldots & \sigma_p 0 & \ldots & 0 \\ \sigma_1 & & \sigma_2\sigma_3\ldots\sigma_p 0 & 0 \\ & \sigma_1 & \sigma_2 & \\ & & \sigma_2 & \sigma_p \end{pmatrix} + q_i p_i (\sigma_1^2 \;\; 2\sigma_1\sigma_2 \ldots 2\sigma_1\sigma_p \;\; \sigma_2^2 \ldots 2\sigma_2\sigma_p \ldots \sigma_p{}^2)$$

3) Third block is a $\frac{p(p+1)}{2}xp$ matrix where each column corresponds to different cov(e_k, f_0), $1 \le k \le i$ element

$$\frac{\partial\Sigma_{i+1}}{\partial\sigma_i} = \frac{\partial}{\partial\sigma_i}vec\left[\left(I - \frac{\Sigma_i\sigma_i\sigma_i^{'}}{\sigma_i^{'}\Sigma_i\sigma_i}\right)\Sigma_i\left(I - \frac{\Sigma_i\sigma_i\sigma_i^{'}}{\sigma_i^{'}\Sigma_i\sigma_i}\right)^{'}\right] = \frac{\partial}{\partial\sigma_i}vec\left[\Sigma_i - \frac{\Sigma_i\sigma_i\sigma_i^{'}}{\sigma_i^{'}\Sigma_i\sigma_i}\right] =$$

$$= -\frac{1}{\sigma_i^{'}\Sigma_i\sigma_i}\left[\left(2\sigma_1^{'}\sigma_i\sigma_1 \quad \sigma_1\sigma_2^{'}\sigma_i + \sigma_2\sigma_1^{'}\sigma_i \ldots 2\sigma^{'}{}_p\sigma_i\sigma_p\right)^{'} - 2\left(\left(\sigma_1^{'}\sigma_i\right)^2 \ldots \left(\sigma_p^{'}\sigma_i\right)^2\right)^{'}\sigma_i^{'}\Sigma_i\right]$$

4) Forth block is the $\frac{p(p+1)}{2}xp$ dimensional matrix

$$\frac{\partial\Sigma_{i+1}}{\partial\Sigma_i} = \frac{\partial}{\partial\Sigma_i}vec\left[\Sigma_i - \frac{\Sigma_i\sigma_i\sigma_i^{'}\Sigma_i}{\sigma_i^{'}\Sigma_i\sigma_i}\right] = \frac{I_{p(p+1)}}{2} - \frac{1}{\sigma_i^{'}\Sigma_i\sigma_i}\frac{\partial}{\partial\Sigma_i}vec\left(\Sigma_i\sigma_i\sigma_i^{'}\Sigma_i\right) +$$

$$+\frac{1}{\left(\sigma_i^{'}\Sigma_i\sigma_i\right)^2} \ldots\ldots$$

let define $u_i = \Sigma_i\sigma_i$ a p-vector then

$$= I - \frac{1}{\sigma_i'\Sigma_i\sigma_i}\frac{\partial}{\partial u_i} vec[u_i u_i']\frac{\partial r}{\partial \Sigma_i} + \frac{1}{\left(\sigma_i'\Sigma_i\sigma_i\right)^2}.vec[u_i u_i'].\left(\sigma_1^2 \quad 2\sigma_1\sigma_2 ... 2\sigma_1\sigma_3 ... \sigma_k^2\right)$$

where $\frac{\partial}{\partial u} vec[u_i u_i']$ is the $\frac{p(p+1)}{2} xp$ matrix

$$\begin{array}{c} p \left[\vphantom{\begin{matrix}0\\0\\0\\0\\0\end{matrix}} \right. \\ (p-1) \left[\vphantom{\begin{matrix}0\\0\\0\\0\end{matrix}} \right. \\ (p-2) \left[\vphantom{\begin{matrix}0\\0\\0\\0\end{matrix}} \right. \\ \cdots\cdots \\ 1 \end{array} \left(\begin{array}{cccc} u + \begin{pmatrix} u_1 \\ 0\ldots \\ 0 \end{pmatrix} & \begin{matrix} 0 \\ u_1 \\ 0\ldots \\ 0 \end{matrix} & \begin{matrix} 0 \\ 0 \\ u_1\ldots \\ 0 \end{matrix} & \begin{matrix} 0 \\ \ldots \\ u_1 \end{matrix} \\ \begin{matrix} 0 \\ 0 \\ \ldots \end{matrix} & u_{-1} + \begin{pmatrix} u_2 \\ 0\ldots \\ 0 \end{pmatrix} & \begin{matrix} 0 \\ u_2 \\ 0 \\ 0 \end{matrix} & \begin{matrix} 0 \\ \ldots \\ u_2 \end{matrix} \\ \ldots & \ldots & u_{-2} + \begin{pmatrix} u_3 \\ 0 \\ 0 \end{pmatrix} & \begin{matrix} 0 \\ \ldots \\ u_3 \end{matrix} \\ & & & \cdots\cdots \\ \cdots\cdots & \cdots\cdots & \cdots & 2u_p \end{array} \right) \frac{p(p+1)}{2} xp$$

and the p x r- matrix $\frac{\partial \Sigma_i \sigma_i}{\partial \Sigma_i}$ is obtained as in paragraph 2) of this Appendix.

Implementing the Jacobian matrix through this methodology we carried out the following products $\frac{\partial B^{i+1}}{\partial B^i} \cdot \frac{\partial B^i}{\partial B^{i-1}}$. The shape of this (p+r)x(p+r)-matrix is as follows

$$\begin{pmatrix} 1) & 2) \\ 3) & 4) \end{pmatrix} \begin{pmatrix} 1)* & 2)* \\ 3)* & 4)* \end{pmatrix} = \begin{pmatrix} K & L \\ M & N \end{pmatrix}$$

Let search the order of matrix products involved

Element of the matrix	dimension	maximum order of product
K	pxp	(pxr)x(rxp)
L	pxr	(pxr)x (rxr)
M	rxp	(rxp)x(pxp)
N	rxr	(rxp)x (pxr)

Notice that the worst case is the (rxr)x(rxr) multiplications and it is avoid in our calculus.

APPENDIX 3

Differentiable Transformation Theorem

Suppose h: $R^q \rightarrow R^s$ is totally differentiable at $u_0 \in R^q$ and $\nabla h(u_0)$ has rank s. If $\sqrt{n}\,(U_n - u_0) \approx N_q(0, \Omega)$ asymptotically, then as $n \rightarrow \infty$

$$\sqrt{n}\,\{h(U_n) - h(u_0)\} \approx N_s(0, \nabla h(u_0)\Omega\nabla h(u_0)').$$

Proof can be found in Barndorff-Nielsen and Cox 1990 , theorem 2.6