CHAPTER IV

CALIBRATION IN A NON-STATIONARY FIELD

4.1 INTRODUCTION

In this chapter we examine the problem of statistical calibration in a non-stationary field. A linear statistical model of the form $\mathbf{V} = \mathbf{W}\beta + \xi$ is assumed to exist at each of a number of locations. Furthermore, the vector of parameters, β at a particular location $\mathbf{X}$ is assumed to be given by the model $\beta = \mathbf{AX}$, where $\mathbf{A}$ is a matrix of deterministic scalars. Given V,W data we could estimate β at each location via OLS. The resulting collection of β's could then be used to similarly estimate $\mathbf{A}$. An alternative to this 'two-stage regression' approach is to determine that $\mathbf{A}$ which minimizes an appropriate criterion such as $(\mathbf{V}-\hat{\mathbf{V}})^{\top}(\mathbf{V}-\hat{\mathbf{V}})$. The details of this latter method are given and it is shown to be equivalent to the two-stage regression approach in the case of a single location. The method is then extended to cater for the case of multiple sampling locations.

We commence by first solving the necessary equations for the case where it is assumed that observations between locations are independent and then investigate the case where the covariance structure between locations is taken into consideration using the Seemingly Unrelated Regression (SUR) procedures discussed in the previous chapter. The computations are illustrated with the use of an example.

4.2 DEFINITIONS AND NOMENCLATURE

Consider a random field in which we measure at some location X, V(X) and W(X). Here V(X) is a random function and W(X) is deterministic. For example, V(X) might be a neutron-probe reading (count) at position X (X possibly three dimensional) and W(X) is the water content at X. Thus, both V and W vary with respect to X and ,at least initially, we will assume W does so in a deterministic way.

The mean, cross-covariance, and variance functions are respectively defined to be :

$$\mu(X) = \mathbb{E}[V(X)] \tag{4.1}$$

$$\mathbb{C}(X,X') = \mathbb{E}\left[[V(X)-\mu(X)][V(X')-\mu(X')]\right] \tag{4.2}$$

$$\mathbb{C}(X,X) = \mathbb{E}\left[[V(X)-\mu(X)]^2\right] = \sigma^2(X) \tag{4.3}$$

At a given location X, we assume the following linear model :

$$V(X) = W(X)\beta(X) + \xi(X) \tag{4.4}$$

where

W(X) is an (n x p) matrix

$\beta(X)$ is a (p x 1) vector

V(X) is an (n x 1) vector

and

$$\mathbb{E}[\xi(X)] = 0 \qquad \mathrm{Var}[\xi(X)] = \sigma^2(X) \tag{4.5}$$

Furthermore, we assume that the process is non-stationary and that $\beta(X)$ varies linearly with X. Thus, if we let dim(X) = ν-1 then

$$\beta(X) = AX \tag{4.6}$$

where A is a (p x ν) matrix of (unknown) constants and X is the (ν x 1) position vector. Note that X has been augmented with the additional row containing the element 1 which corresponds to a constant term in the model given by equation (4.6).

In this chapter we are concerned with the following problem : Given sample data for V(X) and W(X) at known locations X how do we estimate A?

We consider two approaches. In the first case we perform a 2-stage OLS regression to determine the estimate of A whilst the second method estimates A in a more direct manner by performing a least-squares fit with respect to the matrix A. It will be subsequently shown that the two approaches yield the same estimator in the case of a single location. When there are multiple sampling locations it will be demonstrated that the latter method is superior.

4.3 ESTIMATION PROCEDURES FOR INDEPENDENT LOCATIONS

4.3.1 TWO -STAGE REGRESSION.

For convenience we will drop the explicit reference to the position vector X, it being understood hereon that $V \equiv V(X)$; $W \equiv W(X)$; $\beta \equiv \beta(X)$; and $\xi \equiv \xi(X)$.

Our model is thus

$$V = W\beta + \xi$$

Given sample data for V and W we may estimate β in the usual manner as :

$$\hat{\beta} = (W^T W)^{-1} W^T V \qquad (4.7)$$

This corresponds to the first stage of the estimation procedure. Next we assume that $\hat{\beta} = \hat{A}X$ by virtue of equation (4.6). Again, given data for $\hat{\beta}$ from stage 1 and values for X we find the OLS estimator $\hat{A}$ is derived as follows :

$$\hat{A}X = \hat{\beta}$$

$$\Rightarrow \quad \hat{A}X = (W^T W)^{-1} W^T V \qquad (4.8)$$

and thus

$$\hat{A} = (W^T W)^{-1} W^T V \, X^- \qquad (4.9)$$

where X^- is the generalized inverse of X.

4.3.2 LEAST-SQUARES WITH RESPECT TO A

In this case we seek that estimate of A (call it $\tilde{A}$) which minimizes $\Sigma(V-\hat{V})^2$ where $\hat{V} = W\tilde{A}X$ and the summation is taken over all observations at all locations.

Let $Q = [V-\hat{V}]^T[V-\hat{V}]$

$= [V - WAX]^T[V - WAX]$

$= V^TV - 2V^TWAX + X^TA^TW^TWAX$

Now,

$$\frac{\partial Q}{\partial A} = 0 \quad \Rightarrow \quad \frac{\partial}{\partial A}\left[V^TV - 2V^TWAX + X^TA^TW^TWAX \right] = 0$$

$$\Rightarrow \; -2\,\frac{\partial}{\partial A}\left[V^TWAX\right] + \frac{\partial}{\partial A}\left[X^TA^TW^TWAX\right] = 0.$$

Now, $\dfrac{\partial}{\partial A}\left[V^TWAX\right] = W^TV\,X^T$

(Theorem 10.8.3, p353 Graybill, 1983.)

Next, $\dfrac{\partial}{\partial A}\left[X^TA^TW^TWAX\right] = \dfrac{\partial}{\partial A}\left[(WAX)^T(WAX)\right].$

Writing the kernel of the last expression in scalar form we have :

$$(\mathbf{WAX})^{\top}(\mathbf{WAX}) = \sum_{s=1}^{n}\left\{\left[\sum_{v=1}^{p}\sum_{k=1}^{v} x_k a_{vk} w_{sv}\right]\left[\sum_{r=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{p} w_{sk} a_{lr} x_r\right]\right\}$$

$$\Rightarrow = \left\{\left[\sum_{s=1}^{n}\sum_{v=1}^{p}\sum_{k=1}^{v} x_k a_{vk} w_{sv}\right]\left[\sum_{r=1}^{n}\sum_{k=1}^{p}\sum_{l=1}^{p} w_{sk} a_{lr} x_r\right]\right\} \qquad (4.10)$$

Differentiating equation (4.10) with respect to a_{ij} we find :

$$\frac{\partial}{\partial \mathbf{A}}\left[(\mathbf{WAX})^{\top}(\mathbf{WAX})\right] = 2\mathbf{X}^{\top} \otimes \mathbf{W}^{\top}\mathbf{WAX} \qquad (4.11)$$

where the Kronecker-product in equation (4.11) is defined as

$$2[x_1\mathbf{W}^{\top}\mathbf{WAX} \ , \ x_2\mathbf{W}^{\top}\mathbf{WAX} \ , \ . \ . \ . \ x_{\nu}\mathbf{W}^{\top}\mathbf{WAX}]$$

Thus, returning to $\frac{\partial Q}{\partial \mathbf{A}} = 0$ we find :

$$\frac{\partial Q}{\partial \mathbf{A}} = 0 \ \Rightarrow \ -2\mathbf{W}^{\top}\mathbf{VX}^{\top} + 2\mathbf{X}^{\top} \otimes \mathbf{W}^{\top}\mathbf{WAX} = 0$$

$$\Rightarrow \ \mathbf{X}^{\top} \otimes \mathbf{W}^{\top}\mathbf{WAX} = \mathbf{W}^{\top}\mathbf{VX}^{\top}$$

taking transposes

$$\Rightarrow \ \mathbf{X} \otimes \mathbf{X}^{\top}\mathbf{A}^{\top}\mathbf{W}^{\top}\mathbf{W} = \mathbf{X}\,\mathbf{V}^{\top}\mathbf{W}$$

multiply both sides by $\mathbf{X}^{\top}$

$$\Rightarrow \ \mathbf{X}^{\top}\mathbf{X} \otimes \mathbf{X}^{\top}\mathbf{A}^{\top}\mathbf{W}^{\top}\mathbf{W} = \mathbf{X}^{\top}\mathbf{X}\,\mathbf{V}^{\top}\mathbf{W}$$

but X^TX is a scalar and thus cancels on both sides of the previous expression leaving :

$$X^TA^TW^TW = V^TW \tag{4.12}$$

Now W is (n x p), $n \geq p$ and therefore the matrix (W^TW) is full rank. Thus equation (4.12) becomes :

$$X^TA^TW^TW(W^TW)^{-1} = V^TW(W^TW)^{-1}$$

$$\Rightarrow \quad X^TA = V^TW(W\ W)^{-1}$$

taking transposes of both sides

$$\Rightarrow \quad AX = (W^TW)^{-1}W^TV \tag{4.13}$$

Observe that equation (4.13) is exactly the same as equation (4.8) obtained for the two-stage regression. We therefore arrive at the same estimator , namely :

$$\tilde{A} = (W^TW)^{-1}W^TV\ X^- \tag{4.14}$$

That the two procedures yield the same estimator is somewhat surprising. Intuitively, one would have thought the two-stage approach to be less efficient on the grounds that the values of β used in the second stage are not actual values but instead only least-squares estimates and that

the error associated with using $\hat{\beta}$ would have in some way 'compounded' in the second stage. As it turns out, this intuition is correct in the case of multiple sampling locations.

If equation (4.14) were to be used in the case where we had observations at more than one location we would end up estimating a separate **A** matrix for each location with no guarantee that these would be the same. This situation would defeat the object of the exercise since what we require is a single matrix **A** that in some way 'best' represents the data as a whole. Once having obtained this matrix (or its estimate) we are then in a position to determine the model parameters for the regression of **V** on **W** at some new location *without taking any measurements on either*. This is a most attractive facility since all that we are required to know for any future prediction (or calibration) is the position vector **X**. We now consider how this estimation of a single **A** matrix using all the data may be achieved.

4.3.3 THE CASE OF MULTIPLE LOCATIONS

We now assume there are k locations and at each location we have n observations on both **W** and **V**. Thus :

$$\underset{(knx1)}{\mathbf{V}} = \begin{bmatrix} \mathbf{V}_1 \\ \mathbf{V}_2 \\ -- \\ \vdots \\ \mathbf{V}_k \end{bmatrix} \qquad \text{where } \mathbf{V}_i \text{ is } (n \times 1).$$

and

$$\underset{(nx1)}{V_i} = \underset{(nxp)}{W_i}\ \underset{(px1)}{\beta_i} \qquad \text{where} \qquad \underset{(px1)}{\beta_i} = \underset{(p\times\nu)}{A}\ \underset{(\nu x1)}{X_i}$$

and

$$\hat{V} = \begin{bmatrix} W_1AX_1 \\ W_2AX_2 \\ \hline \vdots \\ \hline V_kAX_k \end{bmatrix}$$

As before, let $Q = [V-\hat{V}]^T[V-\hat{V}]$ where

$$[V-\hat{V}] = \begin{bmatrix} V_1 \\ V_2 \\ \hline \vdots \\ \hline V_k \end{bmatrix} - \begin{bmatrix} W_1AX_1 \\ W_2AX_2 \\ \hline \vdots \\ \hline V_kAX_k \end{bmatrix} = \begin{bmatrix} V_1 - W_1AX_1 \\ V_2 - W_2AX_2 \\ \hline \vdots \\ \hline V_k - V_kAX_k \end{bmatrix}$$

and thus :

$$[V-\hat{V}]^T[V-\hat{V}] = \sum_{i=1}^{k} (V_i - W_iAX_i)^T(V_i - W_iAX_i)$$

Hence $Q = \sum_{i=1}^{k} Q_i$ where $Q_i = (V_i - W_iAX_i)^T(V_i - W_iAX_i)$.

We consider again the two approaches : (i) 2-stage regression and (ii) minimization of Q with respect to the matrix A .

4.3.3.1 Two-stage regression

The solution for the 2-stage regression is simple when the number

of locations is $\geq \nu$ (recall,the dimension of the field is ν-1). In this case X has a right-inverse and so the least-squares solution for matrix A is :

$$\tilde{A} = (W^T W)^{-1} W^T V \; X^T (X X^T)^{-1} \qquad (4.15)$$

4.3.3.2 Least-squares with respect to A

Differentiating Q with respect to the matrix A we have :

$$\frac{\partial Q}{\partial A} = 0 \;\Rightarrow\; \left[\!\!\left[\sum_{i=1}^{k} \frac{\partial Q_i}{\partial A} \right]\!\!\right] = 0$$

Now,

$$\frac{\partial Q_i}{\partial A} = -2 W_i^T V_i X_i^T + 2 X_i^T \otimes W_i^T W_i A X_i = 0$$

and therefore

$$\frac{\partial Q}{\partial A} = 0 \;\Rightarrow\; \sum_{i=1}^{k} \left[\!\!\left[X_i^T \otimes W_i^T W_i A X_i - W_i^T V_i X_i^T \right]\!\!\right] \qquad (4.16)$$

Since there is no closed-form solution for equation (4.16) we must solve numerically. This is readily achieved using the Newton-Raphson method as suggested below.

Newton-Raphson solution.

$$\text{Let } g(A) = \sum_{i=1}^{k} \left[X_i^T \otimes W_i^T W_i A X_i - W_i^T V_i X_i^T \right]$$

where { g(A) is a (p x ν) matrix}

Then the iterative procedure is written as :

$$\text{vec}(A^{m+1}) = \text{vec}(A^m) - \left[\nabla \text{vec}[g(A^m)] \right]^{-1} \text{vec}[g(A^m)] \qquad (4.17)$$

4.4 ESTIMATION PROCEDURES FOR CORRELATED LOCATIONS

We again consider the case of determining the matrix A in the system $\beta = AX$; $V = W\beta + \xi$, where V is an (nx1) vector of dependent variable values, W is a (nxp) matrix of independent variable values, β is a (px1) parameter vector for location X (νx1).

The methods of the previous sections implicitly assumed that observations between locations were independent. In many applications this will not be the case - particularly in a geostatistical context where some or all of the locations are within a zone of influence of each other. We now extend earlier results to allow for this between-location covariance effect.

4.4.1 SEEMINGLY UNRELATED REGRESSION AND GENERALIZED LEAST-SQUARES

For the ith. location :

$$\underset{(nx1)}{V_i} = \underset{(nxp)}{W_i}\ \underset{(px1)}{\beta_i} + \underset{(nx1)}{\xi_i} \tag{4.18}$$

where

$$\underset{(px1)}{\beta_i} = \underset{(px\nu)}{A}\ \underset{(\nu x1)}{X_i} \tag{4.19}$$

A convenient way to write this system of equations is :

$$\begin{bmatrix} V_1 \\ V_2 \\ \vdots \\ V_k \end{bmatrix} = \begin{bmatrix} W_1 & & & \\ & W_2 & & \\ & & \ddots & \\ & & & W_k \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_k \end{bmatrix} \tag{4.20}$$

or

$$V = Z\,\Upsilon + \xi \tag{4.21}$$

where

V is (knx1) , Z is (knxkp) , Υ is (kpx1), and ξ is (knx1).

Furthermore, we assume $\mathbb{E}[\xi_i] = 0$ and the covariance matrix of the joint disturbance vector is given by

$$\mathbb{E}[\xi\xi^T] = \Sigma \otimes I_n = \psi, \tag{4.22}$$

where

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{k1} & \sigma_{k2} & \cdots & \sigma_{kk} \end{bmatrix} \qquad (4.23)$$

and σ_{ij} is the covariance between V in location i and V in location j.

If we regard the n observations in V_i as representing a different point in time, then the covariance assumption in equation (4.22) implies that the disturbances in different equations are correlated at a given point in time but are not correlated over time. In econometric theory this is known as a contemporaneous correlation.

4.4.1.1 Estimation

When the system represented by equation (4.20) is viewed as the single equation (4.21), we can estimate $\mathbf{T}$ and hence all the β_i via generalized least squares (GLS). If Z is of rank kp and Σ is known and of rank k, the GLS estimator exists and is given by :

$$\hat{\mathbf{T}} = (Z^{\top}\psi^{-1}Z)^{-1}Z^{\top}\psi^{-1}V \qquad (4.24)$$

Within the class of all estimators that are unbiased and linear functions of V, this estimator is minimum variance and, if V is normally distributed, it is the maximum likelihood estimator and is minimum variance within the class of all unbiased estimators (Judge et.al.1980,p246).

Furthermore :

$$E[\hat{\Upsilon}] = \Upsilon$$

and

$$E[(\hat{\Upsilon} - \Upsilon)(\hat{\Upsilon} - \Upsilon)^T] = (Z^T \Psi^{-1} Z)^{-1} = [Z^T(\Sigma^{-1} \otimes I_n)Z]^{-1}$$

If interest centers only the i^{th} equation and only estimators of V_i are to be considered, then the OLS estimator $\hat{b}_i = (W_i^T W_i)^{-1} W_i^T V_i$ is the minimum variance linear unbiased estimator. However we can improve on this estimator by considering a wider class, namely linear unbiased estimators that are a function of V. Within this class $\hat{\beta}_i$, the i^{th} vector component of $\hat{\Upsilon}$, is better than $\hat{b}_i$ because it uses information on explanatory variables that are included in the system but are excluded from the i^{th} equation. Zellner (1962) indicates that if $\sigma_{ij} = 0$ for all $i \neq j$ or if $W_1 = W_2 = \ldots W_k$ the estimators $\hat{b}_i$ and $\hat{\beta}_i$ will be identical, and so there will be no gain in efficiency. Also, the efficiency gain tends to be higher when the explanatory variables in different equations are not highly correlated but the disturbance terms corresponding to different equations are highly correlated.

In most applications Σ is unknown, and so the estimator $\hat{\Upsilon}$ cannot be employed. However, one can utilize the estimated generalized least squares (EGLS) estimator :

$$\hat{\Upsilon} = [Z^T(\hat{\Sigma}^{-1} \otimes I)Z]^{-1} Z^T(\hat{\Sigma}^{-1} \otimes I)V \qquad (4.25)$$

where the estimator $\hat{\Sigma}$ is based on the OLS residuals $\hat{\xi}_i = V_i - W_i \hat{b}_i$ and has elements given by

$$\hat{\sigma}_{ij} = \frac{\xi_i^T \xi_i}{n} \quad , \qquad i,j = 1,2,\ldots k \tag{4.26}$$

4.4.1.2 Least squares estimation of matrix A

Single location case

We now examine the problem of least squares estimation of matrix A appearing in equation (4.19). Consider a single location i, and thus

$$\beta_i = AX_i$$

where A is a ($p \times \nu$) matrix of unknown constants and X_i is a ($\nu \times 1$) position vector. Our objective is to find that A which minimizes

$$Q = (V_i - \hat{V}_i)^T \psi^{-1} (V_i - \hat{V}_i).$$

(in this case ψ represents the covariance between observations within a single location).

Now,

$$Q = (V_i - W_i A X_i)^T \psi^{-1} (V_i - W_i A X_i)$$

$$= V_i^T \psi^{-1} V_i - 2 V_i \psi^{-1} W_i A X_i + X_i^T A^T W_i^T \psi^{-1} W_i A X_i \tag{4.27}$$

Differentiating equation (4.27) with respect to the matrix A and setting the result equal to zero we have :

$$\frac{\partial Q}{\partial A} = 0 \Rightarrow \frac{\partial}{\partial A}\left[V_i^T\psi^{-1}V_i - 2V_i^T\psi^{-1}W_iAX_i + X_i^TA^TW_i\psi^{-1}W_iAX_i\right] = 0$$

$$\Rightarrow \; -2\frac{\partial}{\partial A}\left[V_i^T\psi^{-1}W_iAX_i\right] + \frac{\partial}{\partial A}\left[X_i^TA^TW_i\psi^{-1}W_iAX_i\right] = 0$$

Now,

$$\frac{\partial}{\partial A}\left[V_i^T\psi^{-1}W_iAX_i\right] = W_i^T\psi^{-1}V_iX_i^T$$

and

$$\frac{\partial}{\partial A}\left[X_i^TA^TW_i\psi^{-1}W_iAX_i\right] = \frac{\partial}{\partial A}\left[(W_iAX_i)^T\psi^{-1}(W_iAX_i)\right]$$

$$= \; 2X_i^T \otimes W_i^T\psi^{-1}W_iAX_i$$

and therefore

$$\frac{\partial Q}{\partial A} = 0 \Rightarrow -2W_i^T\psi^{-1}V_iX_i^T + 2X_i^T \otimes W_i^T\psi^{-1}W_iAX_i = 0$$

$$\Rightarrow \; X_i^T \otimes W_i^T\psi^{-1}W_iAX_i = W_i^T\psi^{-1}V_iX_i^T$$

Taking transposes :

$$\Rightarrow \; X_i \otimes X_i^TA^TW_i^T\psi^{-1}W_i = X_iV_i^T\psi^{-1}W_i$$

Multiplying both sides by X^T :

$$\Rightarrow \; X_i^TX_i \otimes X_i^TA^TW_i^T\psi^{-1}W_i = X_i^TX_iV_i^T\psi^{-1}W_i$$

$$\Rightarrow \; X_i^TA^T(W_i^T\psi^{-1}W_i) = V_i^T\psi^{-1}W_i \qquad \text{(since } X_i^TX_i \text{ is scalar)}$$

Now W_i is full rank and hence :

$$X_i^T A^T = (W_i^T \psi^{-1} W_i)^{-1} V_i^T \psi^{-1} W_i$$

Finally, taking transposes we obtain:

$$AX_i = (W_i^T \psi^{-1} W_i)^{-1} W_i^T \psi^{-1} V_i$$

$$\Rightarrow \hat{T}_i = (W_i^T \psi^{-1} W_i)^{-1} W_i^T \psi^{-1} V_i \qquad (4.28)$$

Notice that , in the case of a single location,the least squares solution given by equation (4.28) is the same as the GLS estimator in equation (4.24) (Z in the case of a single location reduces to W_i).

Multiple locations.

For the case where we have more than one location we may express the quantity Q as follows :

$$Q = [V - Z \text{ vec}(AX)]^T \psi^{-1} [V - Z \text{ vec}(AX)] \qquad (4.29)$$

where V and Z are respectively (knx1) and (knxkp) matrices. A is (pxν) and X is (νxk). Previously in the case of multiple locations we were able to express Q as $\sum_{i=1}^{k} Q_i$ where Q_i was the contribution from the i^{th} location and thus $\frac{\partial Q}{\partial A} = 0 \quad \Rightarrow \quad \sum_{i=1}^{k} \frac{\partial Q_i}{\partial A} = 0$.

However, this is no longer possible due to the presence of ψ in equation (4.29) and we are forced to consider (4.29) a whole. The existence of a closed form solution to equation (4.29) has not been established at this time. As a practical alternative we can solve equation (4.29) numerically using a Newton-Raphson technique.

The estimate of A at the $(m+1)^{th}$ step is given as :

$$A^{(m+1)} = A^{(m)} - [\nabla^2 Q(A)]^{-1} \nabla Q(A) \qquad (4.30)$$

where $\nabla^2 Q(A)$ is the Hessian of Q evaluated at the current vec(A) and $\nabla Q(A)$ is the gradient function of Q evaluated at vec(A).

The iterative procedure described by equation (4.30) has been programmed using the matrix-based language GAUSS. A listing of the source code is given in Appendix L.

It should be pointed out that the process for determining the least squares solution for A using the EGLS approach is very computationally intensive since the covariance matrix $\hat{\psi}$ is updated at the end of the Newton-Raphson iterations and the cycle repeated using this revised $\hat{\psi}$. The process terminates when the norm of the difference between two successive approximations to A differ by less than some prescribed ammount.

4.4.1.3 Two stage regression approach

With k locations and n observations per location in a ν-dimensional field we can derive an estimate for **A** which parallels the so-called two-stage regression procedure previously developed. This estimate is given in closed form as :

$$\hat{\mathbf{T}} = [Z^{\top}(\hat{\Sigma}^{-1} \otimes I)Z]^{-1}Z^{\top}(\hat{\Sigma}^{-1} \otimes I)V\ X^{\top}(XX^{\top})^{-1} \qquad (4.31)$$

provided $\nu \leq k$.

We now illustrate the computations associated with both approaches with the use of an example.

4.5 AN EXAMPLE

The data used for the purpose of illustration of the techniques discussed in this chapter were generated according to the following scheme.

At selected locations in a two-dimensional field compute values of the independent variables W_1 and W_2 as follows :

$$w_1 = 3\exp\{-\tfrac{1}{2}[\tfrac{x_1^2}{4} + \tfrac{x_2^2}{5}]\}$$

and

$$w_2 = 2\exp\{-[x_1 + x_2]/10\}$$

The parameter values are computed as :

$$\beta_0(X) = -2 + 3x_1 + 4x_2$$
$$\beta_1(X) = 5 - x_1 + 2x_2$$
$$\beta_2(X) = 2 + x_1 - 3x_2$$

and thus

$$A = \begin{bmatrix} -2 & 3 & 4 \\ 5 & -1 & 2 \\ 2 & 1 & -3 \end{bmatrix}$$

In table 23 below we have used 5 locations with 6 observations at each location. The values of W_1 and W_2 at a given location were randomly generated about the mean value given by the expressions above. Two sets of V data are also given. The first set (V_1) are those values of V that satisfy $V = W\beta$ exactly at each location while the second set (V_2) were obtained by adding to the V_1 values a normally distributed (0,1) random error.

4.5.1 ANALYSIS OF V_1 DATA

Using the V_1 data in the table as dependent we estimate the A matrix using the two approaches as follows :

Table 23. Sample data for non-stationary calibration.

X_1	X_2	W_1	W_2	V_1	V_2
1.0	1.0	2.439	1.701	19.6340	20.5147
1.0	1.0	2.419	1.733	19.5140	19.3308
1.0	1.0	2.343	1.595	19.0580	18.5321
1.0	1.0	2.409	1.667	19.4540	18.9472
1.0	1.0	2.297	1.638	18.7820	19.3467
1.0	1.0	2.456	1.807	19.7360	19.0564
1.5	-0.7	2.140	1.904	14.8564	16.4618
1.5	-0.7	2.156	1.884	14.7780	14.4764
1.5	-0.7	2.202	1.833	14.5890	13.9134
1.5	-0.7	2.195	1.717	13.9247	14.3847
1.5	-0.7	2.145	1.705	13.7525	13.8574
1.5	-0.7	2.194	1.867	14.7626	16.0857
-2.3	1.8	1.779	2.024	6.1543	4.9149
-2.3	1.8	1.067	2.096	-2.0169	-2.5244
-2.3	1.8	1.092	2.105	-1.7957	-1.9037
-2.3	1.8	1.103	2.129	-1.8126	-2.8437
-2.3	1.8	1.138	2.050	-0.9808	-1.5414
-2.3	1.8	1.136	2.056	-1.0368	0.0647
-1.4	-1.2	2.042	2.713	8.5626	8.8700
-1.4	-1.2	2.005	2.651	8.1542	7.7663
-1.4	-1.2	2.003	2.514	7.5708	9.1219
-1.4	-1.2	2.080	2.628	8.3576	8.0765
-1.4	-1.2	2.014	2.654	8.2028	9.9138
-1.4	-1.2	2.052	2.609	8.1658	6.6364
0.9	2.1	1.768	1.488	18.7152	17.4071
0.9	2.1	1.775	1.479	18.8039	18.3699
0.9	2.1	1.743	1.463	18.5927	17.4336
0.9	2.1	1.790	1.478	18.9318	20.7829
0.9	2.1	1.781	1.478	18.8571	17.6514
0.9	2.1	1.751	1.588	18.2341	19.8053

4.5.1.1 Two-stage regression

Using (equation 4.31) we obtain

$$\mathbf{A} = \begin{bmatrix} -2 & 3 & 4 \\ 5 & -1 & 2 \\ 2 & 1 & -3 \end{bmatrix}$$

which is in perfect agreement with the actual A matrix used to generate the V_1 data.

4.5.1.2 Least squares with respect to A

Using equation (4.30) we obtain

$$\mathbf{A} = \begin{bmatrix} -1.9996 & 3.0001 & 3.9997 \\ 4.9998 & -1.0000 & 2.0001 \\ 2.0000 & 1.0000 & -3.0000 \end{bmatrix}$$

Again, the estimate of A is in almost perfect agreement with the actual A.

4.5.2 ANALYSIS OF V_2 DATA

4.5.2.1.OLS Analysis

The output below is from the GUASS program appearing in the Appendix L.

DATA ENTRY
==========

Enter drive and (optionally) a path for stored matrices : e:

Do you want OLS estimation or GLS estimation (Type O or G) : o

Two-stage regression estimate of matrix A :

11.3359	-17.3812	-20.0568
-1.0255	4.6563	13.4781
2.8953	6.6474	-1.1142

(matrix A_1)

Strike a key when ready . . .

Enter tolerance for determining stopping criterion : ? 1e-6

Iteration 1.0000

16.7866
-1.3619
-6.4514
-4.6814
-0.2161
7.5284
3.1391
2.8161
-3.5875

Iteration 2.0000

19.6008
-0.5280
-9.4370
-6.6670
-0.2688
8.6732
3.8456
2.5658
-3.1661

Convergence established at iteration 5.0000

Least squares estimate

Matrix A =

19.6000	-0.5280	-9.4365
-6.6666	-0.2689	8.6730
3.8456	2.5658	-3.1661

(matrix A_2)

Will now compute the predicted values of dependent variable using this A matrix . . .

Actual matrix of V data :

20.5147	16.4618	4.9149	8.8700	17.4071
19.3308	14.4764	-2.5244	7.7663	18.3699
18.5321	13.9134	-1.9037	9.1219	17.4336
18.9472	14.3847	-2.8437	8.0765	20.7829
19.3467	13.8574	-1.5414	9.9138	17.6514
19.0564	16.0857	0.0647	6.6364	19.8053

Matrix of predicted V values using A_2

19.3935	16.1617	5.1456	8.5615	18.5595
19.4626	15.7533	-2.2217	8.9281	18.6431
18.8827	14.6433	-2.0524	8.4062	18.2893
19.2310	13.5857	-2.1333	7.5825	18.8132
18.9423	14.1238	-1.1860	8.7900	18.7114
19.7670	15.0854	-1.2516	7.9731	18.3179

4.5.2.2 EGLS procedure

DATA ENTRY
==========

Enter drive and (optionally) a path for stored matrices : e:

Do you want OLS estimation or GLS estimation (Type O or G) : g
Are observations WITHIN a location independent (y/n) : y

Two-stage regression estimate of matrix A :

5.2807	4.0204	-12.0021
4.4922	0.6046	4.6958
-0.0344	-0.0036	2.0038

(matrix A_3)

Strike a key when ready . . .

Enter tolerance for determining stopping criterion : ? 1e-6

Convergence established at iteration 5.0000

Matrix A =

7.3691	1.4302	2.2187
1.7679	3.0079	3.0224
-0.2485	-1.8380	-3.6782

(matrix A_4)

Will now compute the predicted values of dependent variable using this A matrix . . .

Strike a key when ready . . .

Actual matrix of V data :

20.5147	16.4618	4.9149	8.8700	17.4071
19.3308	14.4764	-2.5244	7.7663	18.3699
18.5321	13.9134	-1.9037	9.1219	17.4336
18.9472	14.3847	-2.8437	8.0765	20.7829
19.3467	13.8574	-1.5414	9.9138	17.6514
19.0564	16.0857	0.0647	6.6364	19.8053

Estimated V values using matrix A_4 :

20.2321	16.0520	3.2427	8.5912	18.1242
19.8917	16.1273	2.8459	8.3979	18.2866
20.0945	16.3408	2.8293	7.4869	18.0943
20.1941	16.3616	2.7691	7.7877	18.4586
19.4879	16.1586	2.9880	8.3635	18.3612
19.7536	16.2928	2.9716	7.8296	16.9775

Do you want to continue iterating on psi matrix (y/n) : n
Execution stopped in line 356

4.5.3 CALIBRATING AT A FUTURE LOCATION

We now examine how well the various estimated A matrices from the previous section "predict" the actual V data at a location which was not used in the model fitting.

Position X_1	Position X_2	$\hat{V}$ using matrix : A_1	A_2	A_3	A_4	V actual
2.0	-1.4	91.7878	28.0765	23.4553	14.7445	14.148
1.6	-2.3	88.2433	38.4083	22.4887	15.3834	14.582
1.8	1.4	76.9956	19.5034	19.7228	19.1832	18.289
	SSE :	14900.4	763.17	151.20	1.80	

Based on the performance at the three selected test locations we see that A_1, A_2, and A_3 performed particularly poorly.The estimate of A obtained using the EGLS approach performed substantially better than the other three.