Solution to Exercise 12.16

*12.16* The \( k \times k \) matrix \( X_\bullet^T (\Sigma^{-1} \otimes P_W) X_\bullet \) given in expression (12.66) is positive semidefinite by construction. Show this property explicitly by expressing the matrix in the form \( A^T A \), where \( A \) is a matrix with \( k \) columns and at least \( k \) rows that should depend on a \( g \times g \) nonsingular matrix \( \Psi \) which satisfies the relation \( \Psi \Psi^T = \Sigma^{-1} \).

Show that a positive semidefinite matrix expressed in the form \( A^T A \) is positive definite if and only if \( A \) has full column rank. In the present case, the matrix \( A \) fails to have full column rank if and only if there exists a \( k \)-vector \( \beta \), different from zero, such that \( A \beta = 0 \). Since \( k = \sum_{i=1}^g k_i \), we may write the vector \( \beta \) as \( [ \beta_1 \cdots \beta_g ] \), where \( \beta_i \) is a \( k_i \)-vector for \( i = 1, \ldots, g \). Show that there exists a nonzero \( \beta \) such that \( A \beta = 0 \) if and only if, for at least one \( i \), there is a nonzero \( \beta_i \) such that \( P_W X_i \beta_i = 0 \), that is, if \( P_W X_i \) does not have full column rank.

Show that, if \( P_W X_i \) has full column rank, then there exists a unique solution of the estimating equations (12.60) for the parameters \( \beta_i \) of equation \( i \).

If we set \( A \equiv (\Psi^T \otimes P_W) X_\bullet \), then we see that

\[
A^T A = X_\bullet^T (\Psi \otimes P_W)(\Psi^T \otimes P_W) X_\bullet \\
= X_\bullet^T (\Psi \Psi^T \otimes P_W) X_\bullet = X_\bullet^T (\Sigma^{-1} \otimes P_W) X_\bullet.
\]

In these manipulations, we make use of the fact that the orthogonal projection matrix \( P_W \) is symmetric and idempotent. Since \( \Psi^T \otimes P_W \) is \( gn \times gn \), and \( X_\bullet \) is \( gn \times k \), it follows that \( A \) is a \( gn \times k \) matrix. The requirement on the number of rows is satisfied because \( gn > k \).

For the second part of the question, the positive semidefinite matrix \( A^T A \) is positive definite if and only if \( \beta^T A^T A \beta = 0 \) implies that \( \beta = 0 \). But

\[
\beta^T A^T A \beta = ||A \beta||^2,
\]

and so the quadratic form is zero if and only if \( ||A \beta|| = 0 \), that is, if and only if \( A \beta = 0 \). If this last relation implies that \( \beta = 0 \), then by definition \( A \) has full column rank.

The matrix \( A \) can be expressed explicitly as follows:

\[
A = (\Psi^T \otimes P_W) X_\bullet \\
= \begin{bmatrix}
\psi_{11} P_W & \cdots & \psi_{g1} P_W \\
\vdots & \ddots & \vdots \\
\psi_{1g} P_W & \cdots & \psi_{gg} P_W
\end{bmatrix}
\begin{bmatrix}
X_1 & \cdots & O \\
\vdots & \ddots & \vdots \\
O & \cdots & X_g
\end{bmatrix}
\]

\[
= \begin{bmatrix}
\psi_{11} P_W X_1 & \cdots & \psi_{g1} P_W X_g \\
\vdots & \ddots & \vdots \\
\psi_{1g} P_W X_1 & \cdots & \psi_{gg} P_W X_g
\end{bmatrix}
\]
where \( \psi_{ij} \) is the \( ij \)th element of \( \Psi \). We can postmultiply this by a \( k \)-vector \( \beta \) that is partitioned as \( [\beta_1 \mid \cdots \mid \beta_g] \). The result is

\[
A \beta = \begin{bmatrix}
\psi_{11} P_W X_1 & \cdots & \psi_{g1} P_W X_g \\
\vdots & \ddots & \vdots \\
\psi_{1g} P_W X_1 & \cdots & \psi_{gg} P_W X_g
\end{bmatrix}
\begin{bmatrix}
\beta_1 \\
\vdots \\
\beta_g
\end{bmatrix} = \begin{bmatrix}
\sum_{j=1}^g \psi_{j1} P_W X_j \beta_j \\
\vdots \\
\sum_{j=1}^g \psi_{jg} P_W X_j \beta_j
\end{bmatrix}.
\]

If we stack the \( n \times 1 \) blocks of the \( gn \times 1 \) vector \( A \beta \) horizontally in an \( n \times g \) matrix rather than vertically, we obtain

\[
\begin{bmatrix}
\sum_{j=1}^g \psi_{j1} P_W X_j \beta_j \\
\vdots \\
\sum_{j=1}^g \psi_{jg} P_W X_j \beta_j
\end{bmatrix} = \begin{bmatrix}
\psi_{11} & \cdots & \psi_{1g} \\
\vdots & \ddots & \vdots \\
\psi_{g1} & \cdots & \psi_{gg}
\end{bmatrix}
\begin{bmatrix}
P_W X_1 \beta_1 \\
\vdots \\
P_W X_g \beta_g
\end{bmatrix} = [P_W X_1 \beta_1, \cdots, P_W X_g \beta_g] \Psi.
\]

Clearly, the vector \( A \beta \) is zero if and only if the matrix in the last line above is zero. But \( \Psi \) is a nonsingular \( g \times g \) matrix, and so \( A \beta \) is zero for arbitrary nonzero \( \beta \) if and only if the entire matrix \( [P_W X_1 \beta_1, \cdots, P_W X_g \beta_g] \) is zero. But that can only be the case if \( P_W X_i \beta_i \) is zero for all \( i = 1, \ldots, g \). Consequently, \( A \beta \) is zero with a nonzero \( \beta \) if and only if there is at least one \( i \) such that \( P_W X_i \beta_i = 0 \) with nonzero \( \beta_i \).

We now turn to the third part of the question. If \( X_\bullet ^\top (\Sigma^{-1} \otimes P_W) X_\bullet \) is nonsingular, then the equations (12.60) have a unique solution for \( \beta_\bullet \), and the result is trivial. The only case that needs further study is therefore the one in which the matrix \( A \) does not have full column rank. Suppose then that \( A \) is of rank \( r < k \). Then \( A \) can be partitioned, possibly after a reordering of its columns, as \( A = [A_1 \ A_2] \), where \( A_1 \) is \( gn \times r \) with full column rank, and \( B \) is \( r \times (k-r) \). This result simply makes explicit the fact that \( k-r \) columns of \( A \) are linear combinations of the other \( r \) columns.

Two points need to be established, namely, existence and uniqueness. For existence, observe that equations (12.60) can be written as

\[
A^\top (A \beta_\bullet - (\Psi^\top \otimes P_W)y_\bullet) = 0.
\] (S12.22)

Partition \( \beta_\bullet \) as \( [\beta_1^\bullet \mid \beta_2^\bullet] \), conformably with the partition of \( A \), so that \( \beta_1^\bullet \) has \( r \) elements, and \( \beta_2^\bullet \) has \( k-r \) elements. Thus \( A \beta_\bullet = A_1 (\beta_1^\bullet + B \beta_2^\bullet) \). Now set \( \beta_2^\bullet = 0 \). Then we can show that equations (S12.22) have a unique solution for \( \beta_1^\bullet \). Indeed, equation (S12.22) becomes

\[
\begin{bmatrix}
A_1^\top \\
B^\top A_1^\top
\end{bmatrix}
(A_1 \beta_1^\bullet - (\Psi^\top \otimes P_W)y_\bullet) = 0.
\] (S12.23)
Note that if
\[ A_1^\top (A_1\beta_1^* - (\Psi^\top \otimes P_W)y_*) = 0, \]
then (S12.23) is true, since the last \( k - r \) rows are just linear combinations of the first \( r \) rows. But \( A_1 \) has full column rank of \( r \), and so the \( r \times r \) matrix \( A_1^\top A_1 \) is nonsingular. Thus the equations (S12.24) have a unique solution for the \( r \)-vector \( \beta_1^* \), as claimed. Denote this solution by \( \tilde{\beta}_1^* \), and by \( \tilde{\beta}_* \) the \( k \)-vector \([\tilde{\beta}_1^* \ldots 0]\).

In order to show uniqueness, observe that any other solution to equation (S12.22), say \( \hat{\beta}_* \), is such that
\[ A^\top A(\hat{\beta}_* - \tilde{\beta}_*) = 0. \]

This follows by subtracting (S12.22) evaluated at \( \hat{\beta}_* \) from the same equation evaluated at \( \tilde{\beta}_* \). If we write \( \beta \equiv \hat{\beta}_* - \tilde{\beta}_* \), then, by an argument used earlier, (S12.25) implies that \( A\beta = 0 \), and, by the second part of this exercise, this implies that \( P_WX_i\beta_i = 0 \) for all \( i = 1, \ldots, g \), where \( \beta_i \) is defined as above as the \( i \)th block of \( \beta \). If, for some \( i \), \( P_WX_i \) has full column rank, then it follows that \( \beta_i = 0 \). Thus the two solutions \( \beta_1^* \) and \( \hat{\beta}_* \) have the same \( i \)th block, which is therefore defined uniquely, as we wished to show.