Solution to Exercise 12.16

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

Show that a positive semidefinite matrix expressed in the form $\boldsymbol{A}^\top\boldsymbol{A}$ is positive definite if and only if $\boldsymbol{A}$ has full column rank. In the present case, the matrix $\boldsymbol{A}$ fails to have full column rank if and only if there exists a k–vector $\boldsymbol{\beta}$, different from zero, such that $\boldsymbol{A\beta} = \mathbf{0}$. Since $k = \sum_{i=1}^{g} k_i$, we may write the vector $\boldsymbol{\beta}$ as $[\boldsymbol{\beta}_1 \,\vdots\, \cdots \,\vdots\, \boldsymbol{\beta}_g]$, where $\boldsymbol{\beta}_i$ is a k_i–vector for $i = 1, \ldots, g$. Show that there exists a nonzero $\boldsymbol{\beta}$ such that $\boldsymbol{A\beta} = \mathbf{0}$ if and only if, for at least one i, there is a nonzero $\boldsymbol{\beta}_i$ such that $\boldsymbol{P_W X}_i\boldsymbol{\beta}_i = \mathbf{0}$, that is, if $\boldsymbol{P_W X}_i$ does not have full column rank.

Show that, if $\boldsymbol{P_W X}_i$ has full column rank, then there exists a unique solution of the estimating equations (12.60) for the parameters $\boldsymbol{\beta}_i$ of equation i.

If we set $\boldsymbol{A} \equiv (\boldsymbol{\Psi}^\top \otimes \boldsymbol{P_W})\boldsymbol{X}_\bullet$, then we see that

$$\begin{aligned}\boldsymbol{A}^\top\boldsymbol{A} &= \boldsymbol{X}_\bullet^\top(\boldsymbol{\Psi} \otimes \boldsymbol{P_W})(\boldsymbol{\Psi}^\top \otimes \boldsymbol{P_W})\boldsymbol{X}_\bullet \\ &= \boldsymbol{X}_\bullet^\top(\boldsymbol{\Psi}\boldsymbol{\Psi}^\top \otimes \boldsymbol{P_W})\boldsymbol{X}_\bullet = \boldsymbol{X}_\bullet^\top(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{P_W})\boldsymbol{X}_\bullet.\end{aligned}$$

In these manipulations, we make use of the fact that the orthogonal projection matrix $\boldsymbol{P_W}$ is symmetric and idempotent. Since $\boldsymbol{\Psi}^\top \otimes \boldsymbol{P_W}$ is $gn \times gn$, and $\boldsymbol{X}_\bullet$ is $gn \times k$, it follows that $\boldsymbol{A}$ 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 $\boldsymbol{A}^\top\boldsymbol{A}$ is positive definite if and only if $\boldsymbol{\beta}^\top\boldsymbol{A}^\top\boldsymbol{A\beta} = 0$ implies that $\boldsymbol{\beta} = \mathbf{0}$. But $\boldsymbol{\beta}^\top\boldsymbol{A}^\top\boldsymbol{A\beta} = \|\boldsymbol{A\beta}\|^2$, and so the quadratic form is zero if and only if $\|\boldsymbol{A\beta}\| = 0$, that is, if and only if $\boldsymbol{A\beta} = \mathbf{0}$. If this last relation implies that $\boldsymbol{\beta} = \mathbf{0}$, then by definition $\boldsymbol{A}$ has full column rank.

The matrix $\boldsymbol{A}$ can be expressed explicitly as follows:

$$\begin{aligned}\boldsymbol{A} &= (\boldsymbol{\Psi}^\top \otimes \boldsymbol{P_W})\boldsymbol{X}_\bullet \\ &= \begin{bmatrix} \psi_{11}\boldsymbol{P_W} & \cdots & \psi_{g1}\boldsymbol{P_W} \\ \vdots & \ddots & \vdots \\ \psi_{1g}\boldsymbol{P_W} & \cdots & \psi_{gg}\boldsymbol{P_W} \end{bmatrix} \begin{bmatrix} \boldsymbol{X}_1 & \cdots & \mathbf{O} \\ \vdots & \ddots & \vdots \\ \mathbf{O} & \cdots & \boldsymbol{X}_g \end{bmatrix} \\ &= \begin{bmatrix} \psi_{11}\boldsymbol{P_W X}_1 & \cdots & \psi_{g1}\boldsymbol{P_W X}_g \\ \vdots & \ddots & \vdots \\ \psi_{1g}\boldsymbol{P_W X}_1 & \cdots & \psi_{gg}\boldsymbol{P_W X}_g \end{bmatrix},\end{aligned}$$

where ψ_{ij} is the ij^{th} element of $\boldsymbol{\Psi}$. We can postmultiply this by a k–vector $\boldsymbol{\beta}$ that is partitioned as $[\boldsymbol{\beta}_1 \vdots \cdots \vdots \boldsymbol{\beta}_g]$. The result is

$$\boldsymbol{A\beta} = \begin{bmatrix} \psi_{11}\boldsymbol{P_W X}_1 & \cdots & \psi_{g1}\boldsymbol{P_W X}_g \\ \vdots & \ddots & \vdots \\ \psi_{1g}\boldsymbol{P_W X}_1 & \cdots & \psi_{gg}\boldsymbol{P_W X}_g \end{bmatrix} \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_g \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^{g}\psi_{j1}\boldsymbol{P_W X}_j\boldsymbol{\beta}_j \\ \vdots \\ \sum_{j=1}^{g}\psi_{jg}\boldsymbol{P_W X}_j\boldsymbol{\beta}_j \end{bmatrix}.$$

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

$$\begin{aligned} &[\, \textstyle\sum_{j=1}^{g}\psi_{j1}\boldsymbol{P_W X}_j\boldsymbol{\beta}_j \quad \cdots \quad \textstyle\sum_{j=1}^{g}\psi_{jg}\boldsymbol{P_W X}_j\boldsymbol{\beta}_j \,] \\ &\qquad = [\, \boldsymbol{P_W X}_1\boldsymbol{\beta}_1 \quad \cdots \quad \boldsymbol{P_W X}_g\boldsymbol{\beta}_g \,] \begin{bmatrix} \psi_{11} & \cdots & \psi_{1g} \\ \vdots & \ddots & \vdots \\ \psi_{g1} & \cdots & \psi_{gg} \end{bmatrix} \\ &\qquad = [\, \boldsymbol{P_W X}_1\boldsymbol{\beta}_1 \quad \cdots \quad \boldsymbol{P_W X}_g\boldsymbol{\beta}_g \,]\,\boldsymbol{\Psi}. \end{aligned}$$

Clearly, the vector $\boldsymbol{A\beta}$ is zero if and only if the matrix in the last line above is zero. But $\boldsymbol{\Psi}$ is a nonsingular $g \times g$ matrix, and so $\boldsymbol{A\beta}$ is zero for arbitrary nonzero $\boldsymbol{\beta}$ if and only if the entire matrix $[\boldsymbol{P_W X}_1\boldsymbol{\beta}_1 \;\cdots\; \boldsymbol{P_W X}_g\boldsymbol{\beta}_g]$ is zero. But that can only be the case if $\boldsymbol{P_W X}_i\boldsymbol{\beta}_i$ is zero for all $i = 1, \ldots, g$. Consequently, $\boldsymbol{A\beta}$ is zero with a nonzero $\boldsymbol{\beta}$ if and only if there is at least one i such that $\boldsymbol{P_W X}_i\boldsymbol{\beta}_i = \boldsymbol{0}$ with nonzero $\boldsymbol{\beta}_i$.

We now turn to the third part of the question. If $\boldsymbol{X}_\bullet^\top(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{P_W})\boldsymbol{X}_\bullet$ is nonsingular, then the equations (12.60) have a unique solution for $\boldsymbol{\beta}_\bullet$, and the result is trivial. The only case that needs further study is therefore the one in which the matrix $\boldsymbol{A}$ does not have full column rank. Suppose then that $\boldsymbol{A}$ is of rank $r < k$. Then $\boldsymbol{A}$ can be partitioned, possibly after a reordering of its columns, as $\boldsymbol{A} = [\boldsymbol{A}_1 \;\; \boldsymbol{A}_1\boldsymbol{B}]$, where $\boldsymbol{A}_1$ is $gn \times r$ with full column rank, and $\boldsymbol{B}$ is $r \times (k-r)$. This result simply makes explicit the fact that $k-r$ columns of $\boldsymbol{A}$ 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

$$\boldsymbol{A}^\top\big(\boldsymbol{A\beta}_\bullet - (\boldsymbol{\Psi}^\top \otimes \boldsymbol{P_W})\boldsymbol{y}_\bullet\big) = \boldsymbol{0}. \tag{S12.22}$$

Partition $\boldsymbol{\beta}_\bullet$ as $[\boldsymbol{\beta}_\bullet^1 \vdots \boldsymbol{\beta}_\bullet^2]$, conformably with the partition of $\boldsymbol{A}$, so that $\boldsymbol{\beta}_\bullet^1$ has r elements, and $\boldsymbol{\beta}_\bullet^2$ has $k-r$ elements. Thus $\boldsymbol{A\beta}_\bullet = \boldsymbol{A}_1(\boldsymbol{\beta}_\bullet^1 + \boldsymbol{B\beta}_\bullet^2)$. Now set $\boldsymbol{\beta}_\bullet^2 = \boldsymbol{0}$. Then we can show that equations (S12.22) have a unique solution for $\boldsymbol{\beta}_\bullet^1$. Indeed, equation (S12.22) becomes

$$\begin{bmatrix} \boldsymbol{A}_1^\top \\ \boldsymbol{B}^\top\boldsymbol{A}_1^\top \end{bmatrix}\big(\boldsymbol{A}_1\boldsymbol{\beta}_\bullet^1 - (\boldsymbol{\Psi}^\top \otimes \boldsymbol{P_W})\boldsymbol{y}_\bullet\big) = \boldsymbol{0}. \tag{S12.23}$$

Note that if

$$\boldsymbol{A}_1{}^{\top}\big(\boldsymbol{A}_1\boldsymbol{\beta}_\bullet^1 - (\boldsymbol{\Psi}^{\top} \otimes \boldsymbol{P_W})\boldsymbol{y}_\bullet\big) = \boldsymbol{0}, \tag{S12.24}$$

then (S12.23) is true, since the last $k - r$ rows are just linear combinations of the first r rows. But $\boldsymbol{A}_1$ has full column rank of r, and so the $r \times r$ matrix $\boldsymbol{A}_1{}^{\top}\boldsymbol{A}_1$ is nonsingular. Thus the equations (S12.24) have a unique solution for the r−vector $\boldsymbol{\beta}_\bullet^1$, as claimed. Denote this solution by $\tilde{\boldsymbol{\beta}}_\bullet^1$, and by $\tilde{\boldsymbol{\beta}}_\bullet$ the k−vector $[\tilde{\boldsymbol{\beta}}_\bullet^1 \;\vdots\; \boldsymbol{0}]$.

In order to show uniqueness, observe that any other solution to equation (S12.22), say $\hat{\boldsymbol{\beta}}_\bullet$, is such that

$$\boldsymbol{A}^{\top}\boldsymbol{A}(\hat{\boldsymbol{\beta}}_\bullet - \tilde{\boldsymbol{\beta}}_\bullet) = \boldsymbol{0}. \tag{S12.25}$$

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