Not-for-Publication Appendix to "Optimal Asymptotic Least Aquares Estimation in a Singular Set-up"

Antonio Diez de los Rios
Bank of Canada
diez@bankofcanada.ca

December 2014

A Proof of Propositions

A.1 Proof of Proposition 1

This proof closely follows Peñaranda and Sentana (2012), where further details can be found.

Let the spectral decomposition of $\mathbf{V}_g(\boldsymbol{\theta}^0)$ be given by

$$\mathbf{V}_g(\boldsymbol{\theta}^0) = \left(\begin{array}{cc} \mathbf{T}_1 & \mathbf{T}_2 \end{array} \right) \left(\begin{array}{cc} \boldsymbol{\Lambda} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{array} \right) \left(\begin{array}{c} \mathbf{T}_1' \\ \mathbf{T}_2' \end{array} \right) = \mathbf{T}_1 \boldsymbol{\Lambda} \mathbf{T}_1',$$

where $\boldsymbol{\Lambda}$ is a $(G-S) \times (G-S)$ positive definite diagonal matrix; and, without loss of generality, let $\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)$ be the Moore-Penrose[1] generalized inverse of $\mathbf{V}_g(\boldsymbol{\theta}^0)$:

$$\mathbf{V}_g^{+}(\boldsymbol{\theta}^0) = \mathbf{T}_1 \boldsymbol{\Lambda}^{-1} \mathbf{T}_1'.$$

In order to simplify the notation, it is convenient to reparameterize the parameter space into the alternative K parameters $\boldsymbol{\alpha}$ $(S \times 1)$ and $\boldsymbol{\beta}$ $((K-S) \times 1)$ such that

$$\mathbf{R}(\boldsymbol{\theta}) = \left(\begin{array}{cc} \boldsymbol{\alpha}' & \boldsymbol{\beta}' \end{array} \right)',$$

where the first S elements of $\mathbf{R}(\boldsymbol{\theta})$ are such $\boldsymbol{\alpha} = \mathbf{r}(\boldsymbol{\theta})$. In particular, we can choose $\mathbf{R}(\boldsymbol{\theta})$ to be a regular transformation of $\boldsymbol{\theta}$ on an open neighbourhood of $\boldsymbol{\theta}^0$. Further, let $\mathbf{q}\left[\mathbf{R}(\boldsymbol{\theta})\right] = \boldsymbol{\theta}$ be the corresponding inverse transformation of $\mathbf{R}(\boldsymbol{\theta})$ that recovers $\boldsymbol{\theta}$ back. Let the Jacobians of the inverse transformation be given by

$$\mathbf{Q}(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \frac{\partial \mathbf{q}(\boldsymbol{\alpha}, \boldsymbol{\beta})}{\partial(\boldsymbol{\alpha}', \boldsymbol{\beta}')} = \left[\begin{array}{cc} \mathbf{Q}_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) & \mathbf{Q}_{\boldsymbol{\beta}}(\boldsymbol{\alpha}, \boldsymbol{\beta}) \end{array} \right].$$

[1] As noted by Peñaranda and Sentana (2012), it is possible to show that the results in this proposition hold for any generalized inverse of $\mathbf{V}_g(\boldsymbol{\theta}^0)$. While a similar argument would apply here, we focus on the Moore-Penrose generalized inverse for simplicity.

This transformation allows us to impose the parametric restrictions $\mathbf{r}(\boldsymbol{\theta}) = \boldsymbol{\alpha} = \mathbf{0}$ by simply working with the smaller set of parameters $\boldsymbol{\beta}$ and the distance functions $\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0}, \boldsymbol{\beta})\right]$. Thus the optimal ALS estimator can be defined as $\widehat{\boldsymbol{\theta}} = \mathbf{q}(\mathbf{0}, \widehat{\boldsymbol{\beta}})$ where

$$\widehat{\boldsymbol{\beta}} = \arg\min_{\boldsymbol{\beta}} T\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0}, \boldsymbol{\beta})\right]' \mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0}, \boldsymbol{\beta})\right].$$

(i) Since $(\mathbf{T}_1, \mathbf{T}_2)$ is an orthogonal matrix, and the $rank\left[\mathbf{Q}(\boldsymbol{\alpha}, \boldsymbol{\beta})\right] = K$ given that $\mathbf{R}(\boldsymbol{\theta})$ is a regular transformation of $\boldsymbol{\theta}$ on open neighbourhood of $\boldsymbol{\theta}^0$, we have by the inverse function theorem that

$$rank\left[\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)\right] = rank\begin{bmatrix} \mathbf{T}_1'\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\alpha}}(\mathbf{0},\boldsymbol{\beta}) & \mathbf{T}_1'\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}) \\ \mathbf{T}_2'\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\alpha}}(\mathbf{0},\boldsymbol{\beta}) & \mathbf{T}_2'\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\pi}^0,\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}) \end{bmatrix} = K. \tag{A.1}$$

Note now that Assumptions 1 and 2 imply that $\boldsymbol{\Xi}'\left[\mathbf{l}(\mathbf{0}, \boldsymbol{\beta})\right]\sqrt{T}\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0}, \boldsymbol{\beta})\right] \xrightarrow{p} \mathbf{0}$ for all $\boldsymbol{\beta}$ in the neighbourhood. So, by differentiating this random process with respect to $\boldsymbol{\beta}$ and evaluating the derivatives at the true value $\boldsymbol{\beta}^0$ we have, by the continuous mapping theorem, that

$$\sqrt{T}\left\{\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right] \otimes \mathbf{I}_S\right\} \frac{\partial vec\left\{\boldsymbol{\Xi}'\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]\right\}}{\partial \boldsymbol{\beta}'} + \boldsymbol{\Xi}'\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]\sqrt{T}\frac{\partial\sqrt{T}\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]}{\partial \boldsymbol{\beta}'} \xrightarrow{p} \mathbf{0}$$

$$\left\{\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right] \otimes \mathbf{I}_S\right\} \frac{\partial vec\left\{\boldsymbol{\Xi}'\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]\right\}}{\partial \boldsymbol{\beta}'} + \boldsymbol{\Xi}'\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]\frac{\partial\sqrt{T}\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]}{\partial \boldsymbol{\beta}'} \xrightarrow{p} \mathbf{0},$$

since $1/\sqrt{T} \xrightarrow{p} \mathbf{0}$.

Using the chain rule, the previous expression can be written as

$$\left\{\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right] \otimes \mathbf{I}_S\right\} \frac{\partial vec\left\{\boldsymbol{\Xi}'\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]\right\}}{\partial \boldsymbol{\theta}'}\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)+\boldsymbol{\Xi}'\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]\mathbf{G}_{\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0},\boldsymbol{\beta}^{\mathbf{0}})\right]\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0),$$

which implies that

$$\boldsymbol{\Xi}'\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^0)\right]\mathbf{G}_{\boldsymbol{\theta}}\left[\mathbf{l}(\mathbf{0},\boldsymbol{\beta}^{\mathbf{0}})\right]\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0) = \mathbf{0}$$

with $\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \mathbf{G}_{\boldsymbol{\theta}}\left[\mathbf{p}(\boldsymbol{\theta}), \boldsymbol{\theta}\right]$ and where we have used that $\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0}, \boldsymbol{\beta}^0)\right] \xrightarrow{p} \mathbf{g}\left[\boldsymbol{\pi}^0, \mathbf{q}(\mathbf{0}, \boldsymbol{\beta}^0)\right] = \mathbf{g}(\boldsymbol{\pi}^0, \boldsymbol{\theta}^0) = \mathbf{0}$, and that $\mathbf{G}_{\boldsymbol{\theta}}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0},\boldsymbol{\beta}^{\mathbf{0}})\right] \xrightarrow{p} \mathbf{G}_{\boldsymbol{\theta}}\left[\mathbf{p}(\boldsymbol{\theta}^0), \mathbf{q}(\mathbf{0},\boldsymbol{\beta}^{\mathbf{0}})\right] = \mathbf{G}_{\boldsymbol{\theta}}\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^{\mathbf{0}})\right]$.

Finally, note that since $\mathbf{T}_2'\mathbf{V}_g(\boldsymbol{\theta}^0) = 0$, then $\mathbf{T}_2$ must be a full-column rank linear transformation of $\boldsymbol{\Xi}(\boldsymbol{\theta})$. Therefore, it has to be that

$$\mathbf{T}_2'\mathbf{G}_{\boldsymbol{\theta}}\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta}^{\mathbf{0}})\right]\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0) = \mathbf{0},$$

which implies that $rank\left[\mathbf{Q}_1'\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\pi}^0, \boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0}, \boldsymbol{\beta})\right] = K - S$ for (A.1) to be true. Thus, after imposing that $\boldsymbol{\alpha} = \mathbf{0}$, the reduced system of distance functions $\mathbf{Q}_1'\mathbf{g}\left[\widehat{\boldsymbol{\pi}}, \mathbf{q}(\mathbf{0}, \boldsymbol{\beta})\right]$ will first-order identify $\boldsymbol{\beta}$ at $\boldsymbol{\beta}^0$.

(ii) Since the transformation from $\boldsymbol{\theta}$ to $(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is regular on an open neighbourhood of $\boldsymbol{\theta}^0$, a first-order expansion system of distance functions delivers:

$$\begin{aligned}\sqrt{T}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}^0) &= -\left[\mathbf{Q}_{\boldsymbol{\beta}}'(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}_{\boldsymbol{\theta}}'(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\right]^{-1} \\ &\quad \times\mathbf{Q}_{\boldsymbol{\beta}}'(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}_{\boldsymbol{\theta}}'(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}}, \boldsymbol{\theta}^0) + o_p(1). \end{aligned} \tag{A.2}$$

Therefore,

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0) \xrightarrow{d} N\left[\mathbf{0},\mathbf{V}_{\boldsymbol{\beta}}\right],$$

where

$$\mathbf{V}_{\boldsymbol{\beta}} = \left[\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\right]^{-1}. \quad \text{(A.3)}$$

In addition, note that since the optimal ALS estimator is given by $\widehat{\boldsymbol{\theta}} = \mathbf{q}(\mathbf{0},\widehat{\boldsymbol{\beta}})$, we can use the Delta method to compute its asymptotic distribution:

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}^0) \xrightarrow{d} N\left[\mathbf{0},\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{V}_{\boldsymbol{\beta}}\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\right]. \quad \text{(A.4)}$$

We now compare the asymptotic covariance matrix of this optimal estimator with the ALS estimator that uses $\mathbf{W}$ as a weighting matrix and does not impose the restrictions $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$. In particular, the asymptotic covariance matrix of such an estimator is given by

$$\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]^{-1}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{V}_{\mathbf{g}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]^{-1}.$$

Therefore, for $\widehat{\boldsymbol{\theta}}$ to be optimal, we need

$$\begin{aligned}&\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]^{-1}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{V}_{\mathbf{g}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]^{-1}\\&-\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{V}_{\boldsymbol{\beta}}\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\end{aligned}$$

to be positive semidefinite, which in turn requires

$$\begin{aligned}&\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{V}_{\mathbf{g}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\\&-\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{V}_{\boldsymbol{\beta}}\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]\end{aligned}$$

to be positive semidefinite as well.

It can be shown that this is the case given that this matrix is the asymptotic residual variance of the limiting least squares projection of $\sqrt{T}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{g}\left[\widehat{\boldsymbol{\pi}},\boldsymbol{\theta}^0\right]$ on $\sqrt{T}\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{g}(\widehat{\boldsymbol{\pi}},\boldsymbol{\theta})$. In particular:

$$\begin{aligned}&\lim_{T\longrightarrow\infty} Var\left[\begin{array}{c}\sqrt{T}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{g}\left[\widehat{\boldsymbol{\pi}},\boldsymbol{\theta}^0\right]\\ \sqrt{T}\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{g}(\widehat{\boldsymbol{\pi}},\boldsymbol{\theta})\end{array}\right] =\\&\left[\begin{array}{cc}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{V}_{\mathbf{g}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0) & \mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\\ \mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0) & \mathbf{V}_{\boldsymbol{\beta}}^{-1}\end{array}\right].\end{aligned}$$

Alternatively, we can consider the variance of a third ALS estimator that uses $\mathbf{W}$ as weighting matrix but imposes the restrictions $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$:

$$\begin{aligned}&\left[\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\right]^{-1}\\&\times\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{V}_g(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\\&\times\left[\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{W}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\right]^{-1},\end{aligned}$$

and the variance of a fourth estimator that uses the generalized inverse of $\mathbf{V}_g(\boldsymbol{\theta}^0)$ as a weighting matrix but does not impose $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$:

$$\begin{aligned}&\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]^{-1}\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]^{-1} =\\&\left[\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\right]^{-1}.\end{aligned}$$

Again, it is possible to prove that the difference between any of these two matrices and $\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{V}_{\boldsymbol{\beta}}\mathbf{Q}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)$ is positive semidefinite.

(iii) Using a Taylor expansion of $\sqrt{T}\mathbf{g}\left[\widehat{\boldsymbol{\pi}},\mathbf{q}(\mathbf{0},\widehat{\boldsymbol{\beta}})\right]$ and equation (A.2), we have that

$$\begin{aligned}
&\sqrt{T}\mathbf{g}\left[\widehat{\boldsymbol{\pi}},\mathbf{q}(\mathbf{0},\widehat{\boldsymbol{\beta}})\right] \\
&= \sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}},\boldsymbol{\theta}^0) + \mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\sqrt{T}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}^0) + o_p(1) \\
&= \left[\mathbf{I}-\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{V}_{\boldsymbol{\beta}}\mathbf{Q}'_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\mathbf{G}'_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{T}_1\boldsymbol{\Lambda}^{-1}\mathbf{T}'_1\right]\sqrt{T}\mathbf{g}(\widehat{\boldsymbol{\pi}},\boldsymbol{\theta}^0) + o_p(1),
\end{aligned}$$

and rearranging the previous expression as

$$\begin{aligned}
&\sqrt{T}\mathbf{g}\left[\widehat{\boldsymbol{\pi}},\mathbf{q}(\mathbf{0},\widehat{\boldsymbol{\beta}})\right] \\
&= \mathbf{T}_1\boldsymbol{\Lambda}^{1/2}\left[\mathbf{I}_{G-S}-\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\right]\sqrt{T}\boldsymbol{\Lambda}^{-1/2}\mathbf{T}'_1\mathbf{g}(\widehat{\boldsymbol{\pi}},\boldsymbol{\theta}^0)+o_p(1),
\end{aligned}$$

where $\mathbf{H}=\boldsymbol{\Lambda}^{-1/2}\mathbf{T}'_1\mathbf{G}_{\boldsymbol{\theta}}(\boldsymbol{\theta}^0)\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)$. Therefore, the criterion function evaluated at the optimal ALS estimator is

$$T\mathbf{g}\left[\widehat{\boldsymbol{\pi}},\mathbf{q}(\mathbf{0},\widehat{\boldsymbol{\beta}})\right]'\mathbf{V}_g^{+}(\boldsymbol{\theta}^0)\mathbf{g}\left[\widehat{\boldsymbol{\pi}},\mathbf{q}(\mathbf{0},\widehat{\boldsymbol{\beta}})\right] = \widehat{\mathbf{z}}'\left[\mathbf{I}_{G-S}-\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\right]\widehat{\mathbf{z}}+o_p(1),$$

where $\widehat{\mathbf{z}}=\boldsymbol{\Lambda}^{-1/2}\mathbf{T}'_1\mathbf{g}(\widehat{\boldsymbol{\pi}},\boldsymbol{\theta}^0)$ is asymptotically distributed as a standard multivariate normal, which implies that the criterion function converges to a chi-square distribution with $G-K$ degrees of freedom, given that the matrix $\left[\mathbf{I}_{G-S}-\mathbf{H}(\mathbf{H}'\mathbf{H})^{-1}\mathbf{H}'\right]$ is idempotent with rank $(G-S)-(K-S)=G-K$.

A.2 Proof of Proposition 2

As in the proof of Proposition 1, we will work with the alternative set of K parameters of interest $\boldsymbol{\alpha}$ $(S\times 1)$ and $\boldsymbol{\beta}$ $((K-S)\times 1)$ such that

$$\mathbf{R}(\boldsymbol{\theta}) = \begin{pmatrix}\boldsymbol{\alpha}' & \boldsymbol{\beta}'\end{pmatrix}',$$

where the first S elements of $\mathbf{R}(\boldsymbol{\theta})$ are such that $\boldsymbol{\alpha}=\mathbf{r}(\boldsymbol{\theta})$. Again, let $\mathbf{q}\left[\mathbf{R}(\boldsymbol{\theta})\right]=\boldsymbol{\theta}$ be the inverse transformation of $\mathbf{R}(\boldsymbol{\theta})$ that recovers $\boldsymbol{\theta}$ back, and let its Jacobians be denoted by $\mathbf{Q}(\boldsymbol{\alpha},\boldsymbol{\beta})=\partial\mathbf{q}(\boldsymbol{\alpha},\boldsymbol{\beta})/\partial(\boldsymbol{\alpha}',\boldsymbol{\beta}')$. As noted earlier, this (regular) transformation allows us to impose the parametric restriction $\mathbf{r}(\boldsymbol{\theta})=\mathbf{0}$ by simply setting $\boldsymbol{\alpha}=\mathbf{0}$. In particular, the asymptotic distribution of the ML estimate of $\boldsymbol{\beta}$ subject to the restriction that $\boldsymbol{\alpha}=\mathbf{0}$ is given by

$$\sqrt{T}(\widehat{\boldsymbol{\beta}}_{ML}-\boldsymbol{\beta}^0)\xrightarrow{d}N\left[\mathbf{0},\boldsymbol{\Upsilon}^{-1}_{\beta\beta}(\mathbf{0},\boldsymbol{\beta}^0)\right],$$

where $\boldsymbol{\Upsilon}_{\beta\beta}(\boldsymbol{\alpha},\boldsymbol{\beta}) = -\frac{1}{T}E\left[\frac{\partial^2\log L(\boldsymbol{\alpha},\boldsymbol{\beta})}{\partial\boldsymbol{\beta}\partial\boldsymbol{\beta}'}\right]$ is the relevant block of the information matrix. Similarly, since the ML estimator of $\boldsymbol{\theta}$ that imposes the restriction $\mathbf{r}(\boldsymbol{\theta})=\mathbf{0}$ is given by $\widehat{\boldsymbol{\theta}}_{ML}=\mathbf{q}(\mathbf{0},\widehat{\boldsymbol{\beta}}_{ML})$, we can use the Delta method to compute its asymptotic distribution:

$$\sqrt{T}(\widehat{\boldsymbol{\theta}}_{ML}-\boldsymbol{\theta}^0)\xrightarrow{d}N\left[\mathbf{0},\mathbf{Q}_{\beta}(\mathbf{0},\boldsymbol{\beta}^0)\boldsymbol{\Upsilon}^{-1}_{\beta\beta}(\mathbf{0},\boldsymbol{\beta}^{\mathbf{0}})\mathbf{Q}'_{\boldsymbol{\beta}}(\mathbf{0},\boldsymbol{\beta}^0)\right].$$

In particular, the optimal ALS estimate of $\boldsymbol{\theta}$ will be asymptotically equivalent to ML if they have the same asymptotic variance. Comparing this expression with equation (A.4), it is straightforward to see that this will only occur when $\mathbf{V}_{\boldsymbol{\beta}} = \boldsymbol{\Upsilon}^{-1}_{\beta\beta}$.

In order to prove this result, we will work on an alternative set of G auxiliary parameters $\boldsymbol{\delta}$ $(S \times 1)$ and $\boldsymbol{\gamma}$ $((G-S) \times 1)$ such that

$$\mathbf{M}\left[\mathbf{p}(\boldsymbol{\theta})\right] = \left[\begin{array}{cc} \boldsymbol{\delta}(\boldsymbol{\theta})' & \boldsymbol{\gamma}(\boldsymbol{\theta})' \end{array}\right]',$$

where the first S elements of $\mathbf{M}(\boldsymbol{\pi})$ are such that $\boldsymbol{\delta} = \mathbf{r}(\boldsymbol{\pi})$. Let $\mathbf{l}\left[\mathbf{M}(\boldsymbol{\pi})\right] = \boldsymbol{\pi}$ be the corresponding inverse transformation of $\mathbf{M}(\boldsymbol{\pi})$ that recovers $\boldsymbol{\pi}$ back. Let the Jacobians of the inverse transformation be given by

$$\mathbf{L}(\boldsymbol{\delta},\boldsymbol{\gamma}) = \frac{\partial \mathbf{l}(\boldsymbol{\delta},\boldsymbol{\gamma})}{\partial(\boldsymbol{\delta}',\boldsymbol{\gamma}')} = \left[\begin{array}{cc} \mathbf{L}_{\delta}(\boldsymbol{\delta},\boldsymbol{\gamma}) & \mathbf{L}_{\gamma}(\boldsymbol{\delta},\boldsymbol{\gamma}) \end{array}\right].$$

Note that this second (regular) transformation of the auxiliary parameters allows us to impose the parametric restriction $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{0}$ on both the estimation of the auxiliary and parameters of interest. Specifically, we have that $\boldsymbol{\delta}(\boldsymbol{\theta}) = \mathbf{r}\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta})\right] = \mathbf{0}$ for all $\boldsymbol{\beta}$. Further, the asymptotic distribution of the ML estimate of $\boldsymbol{\gamma}$ subject to the restriction that $\boldsymbol{\delta} = \mathbf{0}$ is given by

$$\sqrt{T}(\widehat{\boldsymbol{\gamma}}_{ML} - \boldsymbol{\gamma}^0) \xrightarrow{d} N\left[\mathbf{0}, \boldsymbol{\Upsilon}^{-1}_{\gamma\gamma}(\mathbf{0},\boldsymbol{\gamma}_0)\right],$$

where $\boldsymbol{\Upsilon}_{\gamma\gamma}(\boldsymbol{\delta},\boldsymbol{\gamma}) = -\frac{1}{T}E\left[\frac{\partial^2 \log L(\boldsymbol{\delta},\boldsymbol{\gamma})}{\partial\boldsymbol{\gamma}\partial\boldsymbol{\gamma}'}\right]$ is the relevant block of the information matrix. Note that $\boldsymbol{\Upsilon}_{\beta\beta} = \frac{\partial\beta'}{\partial\gamma}\boldsymbol{\Upsilon}_{\gamma\gamma}\frac{\partial\beta}{\partial\gamma'}$.

Moreover, since the ML estimator of $\boldsymbol{\pi}$ that imposes the restriction $\mathbf{r}(\boldsymbol{\pi})$ is given by $\widehat{\boldsymbol{\pi}}_{ML} = \mathbf{l}(\mathbf{0},\widehat{\boldsymbol{\gamma}}_{ML})$, we can use the Delta method to compute its asymptotic distribution:

$$\sqrt{T}(\widehat{\boldsymbol{\pi}}_{ML} - \boldsymbol{\pi}^0) \xrightarrow{d} N\left[\mathbf{0}, \mathbf{L}_{\gamma}(\mathbf{0},\boldsymbol{\gamma}^0)\boldsymbol{\Upsilon}^{-1}_{\gamma\gamma}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{L}'_{\gamma}(\mathbf{0},\boldsymbol{\gamma}^0)\right]. \tag{A.5}$$

Finally, note that, since the system is complete, and the fact that both $\mathbf{R}(\cdot)$ and $\mathbf{M}(\cdot)$ are regular imply that $\mathbf{Q}(\boldsymbol{\alpha},\boldsymbol{\beta})$ and $\mathbf{L}(\boldsymbol{\delta},\boldsymbol{\gamma})$ have full rank, we can write that

$$\frac{\partial\boldsymbol{\pi}}{\partial\theta'} = \frac{\partial\boldsymbol{\pi}}{\left(\begin{array}{cc}\boldsymbol{\delta}' & \boldsymbol{\gamma}'\end{array}\right)} \times \frac{\partial\left(\begin{array}{c}\boldsymbol{\delta}\\ \boldsymbol{\gamma}\end{array}\right)}{\left(\begin{array}{cc}\boldsymbol{\alpha}' & \boldsymbol{\beta}'\end{array}\right)} \times \frac{\partial\left(\begin{array}{c}\boldsymbol{\alpha}\\ \boldsymbol{\beta}\end{array}\right)}{\partial\theta'}$$

$$-\mathbf{G}_{\boldsymbol{\pi}}^{-1}\mathbf{G}_{\boldsymbol{\theta}} = \mathbf{L}(\boldsymbol{\delta},\boldsymbol{\gamma})\left(\begin{array}{cc}\frac{\partial\boldsymbol{\delta}}{\partial\boldsymbol{\alpha}'} & \frac{\partial\boldsymbol{\delta}}{\partial\boldsymbol{\beta}'}\\ \frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\alpha}'} & \frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\beta}'}\end{array}\right)\mathbf{Q}^{-1}(\boldsymbol{\alpha},\boldsymbol{\beta})$$

$$-\mathbf{G}_{\boldsymbol{\pi}}^{-1}\mathbf{G}_{\boldsymbol{\theta}}\left(\begin{array}{cc}\mathbf{Q}_{\boldsymbol{\alpha}} & \mathbf{Q}_{\beta}\end{array}\right) = \left(\begin{array}{cc}\mathbf{L}_{\boldsymbol{\delta}} & \mathbf{L}_{\gamma}\end{array}\right)\left(\begin{array}{cc}\frac{\partial\boldsymbol{\delta}}{\partial\boldsymbol{\alpha}'} & \frac{\partial\boldsymbol{\delta}}{\partial\boldsymbol{\beta}'}\\ \frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\alpha}'} & \frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\beta}'}\end{array}\right),$$

which, since $\boldsymbol{\delta}(\boldsymbol{\theta}) = \mathbf{r}\left[\mathbf{q}(\mathbf{0},\boldsymbol{\beta})\right] = \mathbf{0}$ for all $\boldsymbol{\beta}$ implies that $\partial\boldsymbol{\delta}/\partial\boldsymbol{\beta}' = 0$, we have that

$$-\mathbf{G}_{\boldsymbol{\theta}}\mathbf{Q}_{\beta} = \mathbf{G}_{\boldsymbol{\pi}}\mathbf{L}_{\gamma}\frac{\partial\boldsymbol{\gamma}}{\partial\boldsymbol{\beta}'}. \tag{A.6}$$

Substituting equations (A.5) and (A.6) evaluated at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ in the expression for $\mathbf{V}_{\boldsymbol{\beta}}$ in (A.3) we have that

$$\mathbf{V}_{\boldsymbol{\beta}}^{-1} = \frac{\partial \boldsymbol{\gamma}'}{\partial \boldsymbol{\beta}} \left\{ \mathbf{L}'_{\gamma}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{G}'_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0) \left[\mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0)\mathbf{L}_{\gamma}(\mathbf{0},\boldsymbol{\gamma}^0)\boldsymbol{\Upsilon}_{\gamma\gamma}^{-1}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{L}'_{\boldsymbol{\gamma}}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0) \right]^{+} \right.$$
$$\left. \times \mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0)\mathbf{L}_{\gamma}(\mathbf{0},\boldsymbol{\gamma}^0) \right\} \frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\beta}'}.$$

Let $\mathbf{D}$ be the term inside the curly brackets. Premultiplying $\mathbf{D}$ by $\mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0)\mathbf{L}_{\gamma}(\mathbf{0},\boldsymbol{\gamma}^0)\boldsymbol{\Upsilon}_{\gamma\gamma}^{-1}(\mathbf{0},\boldsymbol{\gamma}^0)$, and postmultiplying it by $\boldsymbol{\Upsilon}_{\gamma\gamma}^{-1}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{L}'_{\boldsymbol{\gamma}}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0)$, we find that

$$\mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0)\mathbf{L}_{\gamma}(\mathbf{0},\boldsymbol{\gamma}^0)\boldsymbol{\Upsilon}_{\gamma\gamma}^{-1}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{D}\boldsymbol{\Upsilon}_{\gamma\gamma}^{-1}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{L}'_{\boldsymbol{\gamma}}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0) =$$
$$\mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0)\mathbf{L}_{\gamma}(\mathbf{0},\boldsymbol{\gamma}^0)\boldsymbol{\Upsilon}_{\gamma\gamma}^{-1}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{L}'_{\boldsymbol{\gamma}}(\mathbf{0},\boldsymbol{\gamma}^0)\mathbf{G}_{\boldsymbol{\pi}}(\boldsymbol{\theta}^0),$$

where we have used the fact that a generalized inverse must satisfy $\mathbf{W}\mathbf{W}^{+}\mathbf{W} = \mathbf{W}$. Thus, $\mathbf{D} = \boldsymbol{\Upsilon}_{\boldsymbol{\gamma\gamma}}$ for the last equation to be true. This implies that,

$$\mathbf{V}_{\boldsymbol{\beta}} = \left(\frac{\partial \boldsymbol{\gamma}}{\partial \boldsymbol{\beta}'} \boldsymbol{\Upsilon}_{\gamma\gamma} \frac{\partial \boldsymbol{\gamma}'}{\partial \boldsymbol{\beta}} \right)^{-1} = \boldsymbol{\Upsilon}_{\beta\beta}^{-1}.$$

Therefore, the optimal ALS estimator that uses a generalized inverse of $\mathbf{V}_g(\boldsymbol{\theta}^0)$ as the weighting matrix and that, simultaneously, imposes the restriction $\mathbf{r}(\boldsymbol{\theta}) = \mathbf{r}\left[\mathbf{p}(\boldsymbol{\theta})\right] = \mathbf{0}$ is asymptotically equivalent to the ML estimator that imposes that restriction.