Solution to Exercise 9.6

⋆9.6 Consider the model (9.20), where the matrix $\boldsymbol{\Psi}$ is chosen in such a way that the transformed error terms, the $(\boldsymbol{\Psi}^\top\boldsymbol{u})_t$, are innovations with respect to the information sets Ω_t. In other words, $\mathrm{E}\big((\boldsymbol{\Psi}^\top\boldsymbol{u})_t \,|\, \Omega_t\big) = 0$. Suppose that the $n \times l$ matrix of instruments $\boldsymbol{W}$ is predetermined in the usual sense that $\boldsymbol{W}_t \in \Omega_t$. Show that these assumptions, along with the assumption that $\mathrm{E}\big((\boldsymbol{\Psi}^\top\boldsymbol{u})_t^2 \,|\, \Omega_t\big) = \mathrm{E}\big((\boldsymbol{\Psi}^\top\boldsymbol{u})_t^2\big) = 1$ for $t = 1, \ldots, n$, are enough to prove the analog of (9.02), that is, that

$$\mathrm{Var}(n^{-1/2}\boldsymbol{W}^\top\boldsymbol{\Psi}^\top\boldsymbol{u}) = n^{-1}\mathrm{E}(\boldsymbol{W}^\top\boldsymbol{W}).$$

In order to perform just-identified estimation, let the $n \times k$ matrix $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{J}$, where $\boldsymbol{J}$ is an $l \times k$ matrix of full column rank. Compute the asymptotic covariance matrix of the estimator obtained by solving the moment conditions

$$\boldsymbol{Z}^\top\boldsymbol{\Psi}^\top(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{J}^\top\boldsymbol{W}^\top\boldsymbol{\Psi}^\top(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{0}. \tag{9.119}$$

The covariance matrix you have found will be a sandwich. Find the choice of $\boldsymbol{J}$ that eliminates the sandwich, and show that this choice leads to an asymptotic covariance matrix that is smaller, in the usual sense, than the asymptotic covariance matrix for any other choice of $\boldsymbol{J}$.

Compute the GMM criterion function for model (9.20) with instruments $\boldsymbol{W}$, and show that the estimator found by minimizing this criterion function is just the estimator obtained using the optimal choice of $\boldsymbol{J}$.

Let $\boldsymbol{v}$ be the n–vector with typical element $v_t \equiv (\boldsymbol{\Psi}^\top\boldsymbol{u})_t$. Then we have

$$\begin{aligned}
\mathrm{Var}(n^{-1/2}\boldsymbol{W}^\top\boldsymbol{\Psi}^\top\boldsymbol{u}) &= n^{-1}\mathrm{E}(\boldsymbol{W}^\top\boldsymbol{v}\boldsymbol{v}^\top\boldsymbol{W}) \\
&= \frac{1}{n}\sum_{t=1}^{n}\sum_{s=1}^{n}\mathrm{E}(v_t v_s \boldsymbol{W}_t^\top\boldsymbol{W}_s) \\
&= \frac{1}{n}\sum_{t=1}^{n}\mathrm{E}(v_t^2\boldsymbol{W}_t^\top\boldsymbol{W}_t) + \frac{1}{n}\sum_{s<t}\mathrm{E}(v_t v_s \boldsymbol{W}_t^\top\boldsymbol{W}_s) \\
&\quad + \frac{1}{n}\sum_{s>t}\mathrm{E}(v_t v_s \boldsymbol{W}_t^\top\boldsymbol{W}_s).
\end{aligned} \tag{S9.07}$$

Now observe that

$$\begin{aligned}
\mathrm{E}(v_t^2\boldsymbol{W}_t^\top\boldsymbol{W}_t) &= \mathrm{E}\big(\mathrm{E}(v_t^2\boldsymbol{W}_t^\top\boldsymbol{W}_t \,|\, \Omega_t)\big) \\
&= \mathrm{E}\big(\boldsymbol{W}_t^\top\boldsymbol{W}_t\,\mathrm{E}(v_t^2 \,|\, \Omega_t)\big) = \mathrm{E}(\boldsymbol{W}_t^\top\boldsymbol{W}_t),
\end{aligned}$$

whereas, for $s < t$,

$$\begin{aligned}
\mathrm{E}(v_t v_s \boldsymbol{W}_t^\top\boldsymbol{W}_s) &= \mathrm{E}\big(\mathrm{E}(v_t v_s \boldsymbol{W}_t^\top\boldsymbol{W}_s \,|\, \Omega_t)\big) \\
&= \mathrm{E}\big(v_s\boldsymbol{W}_t^\top\boldsymbol{W}_s\,\mathrm{E}(v_t \,|\, \Omega_t)\big) = \boldsymbol{0}.
\end{aligned}$$

The second last equality here holds because v_s, $\boldsymbol{W}_t$, and $\boldsymbol{W}_s$ all belong to Ω_t. Reversing the roles of s and t shows that the terms in the last sum in (S9.07) also vanish. Thus (S9.07) becomes

$$\operatorname{Var}(n^{-1/2}\boldsymbol{W}^\top\boldsymbol{\varPsi}^\top\boldsymbol{u}) = \frac{1}{n}\sum_{t=1}^{n}\mathrm{E}({\boldsymbol{W}_t}^\top\boldsymbol{W}_t) = n^{-1}\boldsymbol{W}^\top\boldsymbol{W}, \tag{S9.08}$$

as required.

The moment conditions (9.119) can be solved directly to yield

$$\hat{\boldsymbol{\beta}} = (\boldsymbol{Z}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X})^{-1}\boldsymbol{Z}^\top\boldsymbol{\varPsi}^\top\boldsymbol{y}.$$

Replacing $\boldsymbol{y}$ by $\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{u}$ and performing the usual asymptotic manipulations gives

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0) = (n^{-1}\boldsymbol{Z}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X})^{-1}n^{-1/2}\boldsymbol{Z}^\top\boldsymbol{\varPsi}^\top\boldsymbol{u}. \tag{S9.09}$$

Strong identification implies that $\operatorname{plim} n^{-1}\boldsymbol{Z}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X}$ is deterministic and nonsingular. From (S9.08), we have

$$\begin{aligned}\operatorname{Var}(n^{-1/2}\boldsymbol{Z}^\top\boldsymbol{\varPsi}^\top\boldsymbol{u}) &= \operatorname{Var}(n^{-1/2}\boldsymbol{J}^\top\boldsymbol{W}^\top\boldsymbol{\varPsi}^\top\boldsymbol{u})\\ &= n^{-1}\boldsymbol{J}^\top\mathrm{E}(\boldsymbol{W}^\top\boldsymbol{W})\boldsymbol{J} = n^{-1}\mathrm{E}(\boldsymbol{Z}^\top\boldsymbol{Z}).\end{aligned}$$

Thus we see from (S9.09) that

$$\begin{aligned}&\operatorname{Var}\Big(\operatorname*{plim}_{n\to\infty} n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0)\Big) =\\ &\quad\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X}\Big)^{-1}\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{Z}^\top\boldsymbol{Z}\Big)\Big(\operatorname*{plim}_{n\to\infty}\frac{1}{n}\boldsymbol{X}^\top\boldsymbol{\varPsi}\boldsymbol{Z}\Big)^{-1}.\end{aligned} \tag{S9.10}$$

This is the desired asymptotic covariance matrix, and it is indeed a sandwich.

The matrix part of (S9.10), without powers of n or limits, becomes, when $\boldsymbol{Z}$ is replaced by $\boldsymbol{W}\boldsymbol{J}$,

$$(\boldsymbol{J}^\top\boldsymbol{W}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X})^{-1}\boldsymbol{J}^\top\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{J}(\boldsymbol{X}^\top\boldsymbol{\varPsi}\boldsymbol{W}\boldsymbol{J})^{-1}. \tag{S9.11}$$

The sandwich can be eliminated if we can find $\boldsymbol{J}$ so as to satisfy $\boldsymbol{W}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X} = \boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{J}$. Thus, if we set $\boldsymbol{J} = (\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X}$, we find that

$$\boldsymbol{J}^\top\boldsymbol{W}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X} = \boldsymbol{X}^\top\boldsymbol{\varPsi}\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X} = \boldsymbol{X}^\top\boldsymbol{\varPsi}\boldsymbol{P}_{\boldsymbol{W}}\boldsymbol{\varPsi}^\top\boldsymbol{X},$$

and

$$\begin{aligned}\boldsymbol{J}^\top\boldsymbol{W}^\top\boldsymbol{W}\boldsymbol{J} &= \boldsymbol{X}^\top\boldsymbol{\varPsi}\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{\varPsi}^\top\boldsymbol{X}\\ &= \boldsymbol{X}^\top\boldsymbol{\varPsi}\boldsymbol{P}_{\boldsymbol{W}}\boldsymbol{\varPsi}^\top\boldsymbol{X}.\end{aligned}$$

With this choice of $\boldsymbol{J}$, (S9.10) becomes $\operatorname{plim}(n^{-1}\boldsymbol{X}^\top\boldsymbol{\Psi}\boldsymbol{P_W}\boldsymbol{\Psi}^\top\boldsymbol{X})^{-1}$, and it is no longer a sandwich. We see that the instruments $\boldsymbol{WJ}$ which eliminate the sandwich are the columns of the $n \times k$ matrix $\boldsymbol{P_W}\boldsymbol{\Psi}^\top\boldsymbol{X}$.

To show asymptotic efficiency, we again dispense with powers of n and limits. In addition, we argue in terms of precision matrices. The difference between the precision matrix with the sandwich-eliminating choice of $\boldsymbol{J}$ and the precision matrix with arbitrary $\boldsymbol{J}$ is

$$\begin{aligned}\boldsymbol{X}^\top\boldsymbol{\Psi}\boldsymbol{P_W}\boldsymbol{\Psi}^\top\boldsymbol{X} &- \boldsymbol{X}^\top\boldsymbol{\Psi}\boldsymbol{WJ}(\boldsymbol{J}^\top\boldsymbol{W}^\top\boldsymbol{WJ})^{-1}\boldsymbol{J}^\top\boldsymbol{W}^\top\boldsymbol{\Psi}^\top\boldsymbol{X}\\ &= \boldsymbol{X}^\top\boldsymbol{\Psi}\boldsymbol{P_W}\boldsymbol{\Psi}^\top\boldsymbol{X} - \boldsymbol{X}^\top\boldsymbol{\Psi}\boldsymbol{P_{WJ}}\boldsymbol{\Psi}^\top\boldsymbol{X}\\ &= \boldsymbol{X}^\top\boldsymbol{\Psi}(\boldsymbol{P_W} - \boldsymbol{P_{WJ}})\boldsymbol{\Psi}^\top\boldsymbol{X}.\end{aligned}$$

Since $\mathcal{S}(\boldsymbol{WJ}) \subset \mathcal{S}(\boldsymbol{W})$, it follows that $\boldsymbol{P_W} - \boldsymbol{P_{WJ}}$ is an orthogonal projection matrix, and so the last expression above is positive semidefinite. This shows the optimality of the sandwich-eliminating choice of $\boldsymbol{J}$ for estimators based on the instruments $\boldsymbol{W}$.

The moment conditions for the estimation of the model (9.20) using the matrix of instruments $\boldsymbol{W}$ are

$$\mathrm{E}\big(\boldsymbol{W}^\top\boldsymbol{\Psi}^\top(\boldsymbol{y} - \boldsymbol{X\beta})\big) = \boldsymbol{0};$$

compare (9.21). The weighting matrix needed for the GMM criterion function is thus the inverse of the (asymptotic) covariance matrix of $\boldsymbol{W}^\top\boldsymbol{\Psi}^\top\boldsymbol{u}$, which, by (S9.08), is $(\boldsymbol{W}^\top\boldsymbol{W})^{-1}$. Therefore, the GMM criterion function is

$$\begin{aligned}Q(\boldsymbol{\beta}, \boldsymbol{y}) &= (\boldsymbol{y} - \boldsymbol{X\beta})^\top\boldsymbol{\Psi}\boldsymbol{W}(\boldsymbol{W}^\top\boldsymbol{W})^{-1}\boldsymbol{W}^\top\boldsymbol{\Psi}^\top(\boldsymbol{y} - \boldsymbol{X\beta})\\ &= (\boldsymbol{y} - \boldsymbol{X\beta})^\top\boldsymbol{\Psi}\boldsymbol{P_W}\boldsymbol{\Psi}^\top(\boldsymbol{y} - \boldsymbol{X\beta}).\end{aligned}$$

The first-order conditions for the minimization of this function are equivalent to the estimating equations

$$\boldsymbol{X}^\top\boldsymbol{\Psi}\boldsymbol{P_W}\boldsymbol{\Psi}^\top(\boldsymbol{y} - \boldsymbol{X\beta}) = \boldsymbol{0}.$$

These equations are the sample moment conditions generated by using the columns of the matrix $\boldsymbol{P_W}\boldsymbol{\Psi}^\top\boldsymbol{X}$ as instruments. As we saw above, these are the optimal instruments.