Solution to Exercise 9.16

⋆9.16 The minimization of the GMM criterion function (9.87) yields the estimating equations (9.89) with $\boldsymbol{A} = \boldsymbol{\Psi}^\top \boldsymbol{W}$. Assuming that the $n \times l$ instrument matrix $\boldsymbol{W}$ satisfies the predeterminedness condition in the form (9.30), show that these estimating equations are asymptotically equivalent to the equations

$$\bar{\boldsymbol{F}}_0^\top \boldsymbol{\Psi} \boldsymbol{P}_{\boldsymbol{\Psi}^\top \boldsymbol{W}} \boldsymbol{\Psi}^\top \boldsymbol{f}(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}, \tag{9.123}$$

where, as usual, $\bar{\boldsymbol{F}}_0 \equiv \bar{\boldsymbol{F}}(\boldsymbol{\theta}_0)$, with $\boldsymbol{\theta}_0$ the true parameter vector. Next, derive the asymptotic covariance matrix of the estimator defined by these equations.

Show that the equations (9.123) are the optimal estimating equations for overidentified estimation based on the transformed zero functions $\boldsymbol{\Psi}^\top \boldsymbol{f}(\boldsymbol{\theta})$ and the transformed instruments $\boldsymbol{\Psi}^\top \boldsymbol{W}$. Show further that, if the condition $\mathcal{S}(\bar{\boldsymbol{F}}_0) \subseteq \mathcal{S}(\boldsymbol{W})$ is satisfied, the asymptotic covariance matrix of the estimator obtained by solving equations (9.123) coincides with the optimal asymptotic covariance matrix (9.83).

When $\boldsymbol{A} = \boldsymbol{\Psi}^\top \boldsymbol{W}$, the estimating equations (9.89) become

$$\boldsymbol{F}^\top(\hat{\boldsymbol{\theta}}) \boldsymbol{\Psi} \boldsymbol{P}_{\boldsymbol{\Psi}^\top \boldsymbol{W}} \boldsymbol{\Psi}^\top \boldsymbol{f}(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}. \tag{S9.28}$$

The only difference between equations (9.123) and equations (S9.28) is that $\bar{\boldsymbol{F}}_0$ appears in the former and $\boldsymbol{F}(\hat{\boldsymbol{\theta}})$ appears in the latter. In order to show that they give rise to asymptotically equivalent estimators, we multiply the estimating equations (S9.28) by $n^{-1/2}$ so as to put them in a form suitable for asymptotic analysis:

$$\frac{1}{n}\boldsymbol{F}^\top(\hat{\boldsymbol{\theta}}) \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W} \Big(\frac{1}{n}\boldsymbol{W}^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W}\Big)^{-1} n^{-1/2} \boldsymbol{W}^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{f}(\hat{\boldsymbol{\theta}}) = \boldsymbol{0}. \tag{S9.29}$$

First, we note that, since $\hat{\boldsymbol{\theta}}$ is consistent,

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{F}^\top(\hat{\boldsymbol{\theta}}) \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W} = \operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{F}_0^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W}.$$

Next, recall that, in (9.85), $\bar{\boldsymbol{F}}_0$ was defined so that

$$\mathrm{E}\big((\boldsymbol{\Psi}^\top \boldsymbol{F}_0)_t \,|\, \Omega_t\big) = (\boldsymbol{\Psi}^\top \bar{\boldsymbol{F}}_0)_t,$$

which means that

$$(\boldsymbol{\Psi}^\top \boldsymbol{F}_0)_t = (\boldsymbol{\Psi}^\top \bar{\boldsymbol{F}}_0)_t + \boldsymbol{V}_t,$$

where the $1 \times k$ vector $\boldsymbol{V}_t$ is such that $\mathrm{E}(\boldsymbol{V}_t \,|\, \Omega_t) = \boldsymbol{0}$. We then have that

$$\begin{aligned}
\frac{1}{n}\boldsymbol{F}_0^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W} &= \frac{1}{n}\sum_{t=1}^{n} (\boldsymbol{\Psi}^\top \boldsymbol{F}_0)_t^\top (\boldsymbol{\Psi}^\top \boldsymbol{W})_t \\
&= \frac{1}{n}\sum_{t=1}^{n} (\boldsymbol{\Psi}^\top \bar{\boldsymbol{F}}_0)_t^\top (\boldsymbol{\Psi}^\top \boldsymbol{W})_t + \frac{1}{n}\sum_{t=1}^{n} \boldsymbol{V}_t^\top (\boldsymbol{\Psi}^\top \boldsymbol{W})_t.
\end{aligned}$$

Copyright © 2003, Russell Davidson and James G. MacKinnon

By the predeterminedness condition (9.30), $(\boldsymbol{\Psi}^\top \boldsymbol{W})_t \in \Omega_t$. This implies that $\mathrm{E}(\boldsymbol{V}_t{}^\top(\boldsymbol{\Psi}^\top \boldsymbol{W}_t)) = \mathrm{E}\big(\mathrm{E}(\boldsymbol{V}_t{}^\top \,|\, \Omega_t)(\boldsymbol{\Psi}^\top \boldsymbol{W}_t)\big) = \mathbf{0}$. Therefore, the last term here tends to zero as $n \to \infty$ by a law of large numbers. Thus

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \boldsymbol{F}_0^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W} = \operatorname*{plim}_{n\to\infty} \frac{1}{n} \bar{\boldsymbol{F}}_0^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W}. \tag{S9.30}$$

This shows that equations (S9.29) are asymptotically equivalent to the equations

$$\frac{1}{n} \bar{\boldsymbol{F}}_0^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W} \Big(\frac{1}{n} \boldsymbol{W}^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{W}\Big)^{-1} n^{-1/2} \boldsymbol{W}^\top \boldsymbol{\Psi}\boldsymbol{\Psi}^\top \boldsymbol{f}(\hat{\boldsymbol{\theta}}) = \mathbf{0}, \tag{S9.31}$$

which express the estimating equations (9.123) in asymptotic form.

To obtain the asymptotic covariance matrix of the estimator defined by equations (9.123), the simplest method is to use the general formula (9.67), in which, since the zero functions are here the elements of $\boldsymbol{\Psi}^\top \boldsymbol{f}(\boldsymbol{\theta})$, $\boldsymbol{F}$ is to be replaced by $\boldsymbol{\Psi}^\top \boldsymbol{F}$, $\boldsymbol{Z}$ by $\boldsymbol{P}_{\boldsymbol{\Psi}^\top \boldsymbol{W}} \boldsymbol{\Psi}^\top \bar{\boldsymbol{F}}_0$, and $\boldsymbol{\Omega}$ by the identity matrix. The "bread" in the sandwich (9.67), written there as $\operatorname{plim}\big(n^{-1}\boldsymbol{Z}^\top \boldsymbol{F}(\boldsymbol{\theta}_0)\big)^{-1}$ or its transpose, becomes

$$\operatorname*{plim}_{n\to\infty} \Big(\frac{1}{n} \bar{\boldsymbol{F}}_0^\top \boldsymbol{\Psi} \boldsymbol{P}_{\boldsymbol{\Psi}^\top \boldsymbol{W}} \boldsymbol{\Psi}^\top \boldsymbol{F}_0\Big)^{-1}, \tag{S9.32}$$

in which, by the same reasoning as that which led to (S9.30), the final factor of $\boldsymbol{F}_0$ can be replaced by $\bar{\boldsymbol{F}}_0$. The "jam" in the sandwich is written in (9.67) as $\operatorname{plim} n^{-1}\boldsymbol{Z}^\top \boldsymbol{\Omega} \boldsymbol{Z}$, and it becomes

$$\operatorname*{plim}_{n\to\infty} \frac{1}{n} \bar{\boldsymbol{F}}_0^\top \boldsymbol{\Psi} \boldsymbol{P}_{\boldsymbol{\Psi}^\top \boldsymbol{W}} \boldsymbol{\Psi}^\top \bar{\boldsymbol{F}}_0. \tag{S9.33}$$

Since (S9.32) with $\bar{\boldsymbol{F}}_0$ in place of $\boldsymbol{F}_0$ and (S9.33) are mutually inverse matrices, the sandwich collapses, and we conclude that

$$\operatorname{Var}\Big(\operatorname*{plim}_{n\to\infty} n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\Big) = \operatorname*{plim}_{n\to\infty} \Big(\frac{1}{n} \bar{\boldsymbol{F}}_0^\top \boldsymbol{\Psi} \boldsymbol{P}_{\boldsymbol{\Psi}^\top \boldsymbol{W}} \boldsymbol{\Psi}^\top \bar{\boldsymbol{F}}_0\Big)^{-1}. \tag{S9.34}$$

The above result could also be obtained directly, at the cost of a little more work, by Taylor expansion of equation (S9.31).

The result that equations (9.123) are the optimal estimating equations for overidentified estimation based on the transformed zero functions $\boldsymbol{\Psi}^\top \boldsymbol{f}(\boldsymbol{\theta})$ and the transformed instruments $\boldsymbol{\Psi}^\top \boldsymbol{W}$ is intuitively obvious, since the covariance matrix (S9.34) is not a sandwich. To prove it, we need to compare (S9.34) with the asymptotic covariance matrix when the estimating equations are

$$\boldsymbol{Z}^\top \boldsymbol{\Psi} \boldsymbol{P}_{\boldsymbol{\Psi}^\top \boldsymbol{W}} \boldsymbol{\Psi}^\top \boldsymbol{f}(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$$

Copyright © 2003, Russell Davidson and James G. MacKinnon

for some $n \times k$ matrix $\boldsymbol{Z} = \boldsymbol{W}\boldsymbol{J}$. By arguments similar to the one that led to equation (S9.34), it can be shown that the covariance matrix of the inefficient estimator is

$$\begin{aligned}\operatorname{Var}\Big(\operatorname*{plim}_{n\to\infty} n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\Big) &= \Big(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{Z}^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0\Big)^{-1} \times \\ &\Big(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\boldsymbol{Z}^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\boldsymbol{Z}\Big)\Big(\operatorname*{plim}_{n\to\infty} \frac{1}{n}\bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\boldsymbol{Z}\Big)^{-1}.\end{aligned} \tag{S9.35}$$

This sandwich covariance matrix is to be compared with (S9.34).

As in the answer to Exercise 9.6, we will, for simplicity, ignore the probability limits and the factors of $1/n$ and reason in terms of precision matrices. For our purposes, then, the precision matrix for $\hat{\boldsymbol{\theta}}$ is

$$\bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0,$$

and the precision matrix for $\tilde{\boldsymbol{\theta}}$ is

$$\bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\boldsymbol{Z}(\boldsymbol{Z}^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\boldsymbol{Z})^{-1}\boldsymbol{Z}^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0.$$

The latter can evidently be rewritten as

$$\bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{P}_{\boldsymbol{C}}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0,$$

where $\boldsymbol{C} \equiv \boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\boldsymbol{Z}$. Thus the difference between the precision matrices for $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ is simply

$$\begin{aligned}&\bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0 - \bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{P}_{\boldsymbol{C}}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0 \\ &\quad = \bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Psi}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{M}_{\boldsymbol{C}}\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0.\end{aligned}$$

Since this expression is a quadratic form in an orthogonal projection matrix, it must be positive semidefinite, and this establishes the optimality of the estimating equations (9.123).

When $\mathcal{S}(\bar{\boldsymbol{F}}_0) \subseteq \mathcal{S}(\boldsymbol{W})$, $\mathcal{S}(\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0) \subseteq \mathcal{S}(\boldsymbol{\Psi}^{\top}\boldsymbol{W})$. Therefore, in this special case,

$$\boldsymbol{P}_{\boldsymbol{\Psi}^{\top}\boldsymbol{W}}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0 = \boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0,$$

and the right-hand side of (S9.34) reduces to

$$\operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Psi}\boldsymbol{\Psi}^{\top}\bar{\boldsymbol{F}}_0\Big)^{-1} = \operatorname*{plim}_{n\to\infty}\Big(\frac{1}{n}\bar{\boldsymbol{F}}_0^{\top}\boldsymbol{\Omega}^{-1}\bar{\boldsymbol{F}}_0\Big)^{-1}, \tag{S9.36}$$

which is just (9.83), as we wished to show.

Copyright © 2003, Russell Davidson and James G. MacKinnon