Extremum Estimators – Review Exercises

Ryan Benschop

1 Extremum Estimators

1. (Estimating probit model by NLS)
Consider the probit model, where y_i is $\{0,1\}$-valued, x_i is $\mathbb{R}^k$-valued, and $\mathbb{P}(y_i = 1|x_i;\theta_0) = \Phi(x_i'\theta_0)$, where Φ is the standard normal CDF. The exercise asks us to verify that $\mathbb{E}[y_i|x_i] = \Phi(x_i'\theta_0)$, and specify m in the M-estimator objective function, where the model is estimated by NLS.

Answer: $\mathbb{E}[y_i|x_i;\theta_0] = \Phi(x_i'\theta_0)$ follows trivially by noting that $y_i|x_i$ is a Bernoulli random variable with parameter $\Phi(x_i'\theta_0)$. Define $\varepsilon_i = y_i - \Phi(x_i'\theta_0)$. NLS chooses the estimator $\hat{\theta}$ that minimises the sum of squared residuals:

$$\hat{\theta} \in \arg\min_{\theta\in\Theta} \sum_{i=1}^{n} \left(y_i - \Phi(x_i'\theta)\right)^2 = \arg\min_{\theta\in\Theta} n^{-1} \sum_{i=1}^{n} \left(y_i - \Phi(x_i'\theta)\right)^2$$
$$= \arg\max_{\theta\in\Theta} -n^{-1} \sum_{i=1}^{n} \left(y_i - \Phi(x_i'\theta)\right)^2 = \arg\max_{\theta\in\Theta} n^{-1} \sum_{i=1}^{n} m(w_i,\theta)$$

where $w_i = (y_i, x_i)$, and $m(w_i,\theta) = -(y_i - \Phi(x_i'\theta))$.

2. (Estimating probit model by GMM)
Consider the probit model, described above. The exercise asks us to verify that:

$$\mathbb{E}[x_i(y_i - \Phi(x_i'\theta_0))] = 0$$

and to specify g in the corresponding GMM objective function.

Answer: We previously verified that $\mathbb{E}[y_i|x_i] = \Phi(x_i'\theta_0)$. Hence, $\mathbb{E}[y_i - \Phi(x_i'\theta_0)|x_i] = 0$. Pre-multiplying by x_i (which we can bring inside the expectation since we are conditioning on x_i), we have $\mathbb{E}[x_i(y_i - \Phi(x_i'\theta_0))] = 0$. The orthogonality condition $\mathbb{E}[x_i(y_i - \Phi(x_i'\theta_0))] = 0$ then follows by applying the law of iterated expectations. Applying GMM, based on this orthogonality condition

and a weighting matrix $\hat{W}$, we have:

$$\begin{aligned}\hat{\theta} &\in \arg\min_{\theta\in\Theta} \left(\sum_{i=1}^{n} x_i(y_i - \Phi(x_i'\theta))\right)' \hat{W} \left(\sum_{i=1}^{n} x_i(y_i - \Phi(x_i'\theta))\right) \\ &= \arg\max_{\theta\in\Theta} -\frac{1}{2}\left(n^{-1}\sum_{i=1}^{n} x_i(y_i - \Phi(x_i'\theta))\right)' \hat{W} \left(n^{-1}\sum_{i=1}^{n} x_i(y_i - \Phi(x_i'\theta))\right)\end{aligned}$$

Hence, $g(\theta) = n^{-1}\sum_{i=1}^{n} x_i(y_i - \Phi(x_i'\theta))$.

2 Consistency

1. (Necessity of density identification)

Claim: Suppose $\mathbb{E}[\ln f(y_i|x_i;\theta)]$ *is uniquely maximised on* Θ *at* $\theta_0 \in \Theta$*. Then, for every* $\theta \neq \theta_0$*,* $f(y_i|x_i;\theta) \neq f(y_i|x_i;\theta_0)$*.*

That is, conditional density identification is a necessary condition for identification in conditional ML.

Proof. (Using the Kullback-Leibler information inequality). Suppose $\mathbb{E}[\ln f(y_i|x_i;\theta)]$ is uniquely maximised on Θ at $\theta_0 \in \Theta$. Let $\theta_1 \in \Theta \setminus \{\theta_0\}$. Then, by the Kullback-Leibler information inequality, if $f(y_i|x_i|\theta_0) = f(y_i|x_i;\theta_1)$:

$$\mathbb{E}[\ln f(y_i|x_i;\theta_0)] = \mathbb{E}[\ln f(y_i|x_i;\theta_1)]$$

which contradicts our assumption that $\mathbb{E}[\ln f(y_i|x_i;\theta)]$ is uniquely maximised on Θ at $\theta_0 \in \Theta$. Hence, $f(y_i|x_i|\theta_0) \neq f(y_i|x_i;\theta_1)$. □

Proof. (Without using the Kullback-Leibler information inequality). We can show that the contrapositive holds. Suppose there is $\theta_1 \in \Theta \setminus \{\theta_0\}$ such that $f(y_i|x_i;\theta_0) = f(y_i|x_i;\theta_1)$. Then, $\ln f(y_i|x_i;\theta_0) = \ln f(y_i|x_i;\theta_1)$, which implies:

$$\mathbb{E}[\ln f(y_i|x_i;\theta_0)] = \mathbb{E}[\ln f(y_i|x_i;\theta_1)]$$

Hence, $\mathbb{E}[\ln f(y_i|x_i;\theta)]$ is not uniquely maximised on Θ at $\theta_0 \in \Theta$. □

3. (Identification in NLS)

Claim: Consider the NLS model the NLS model, where $\mathbb{E}[y_i|x_i] = \varphi(x_i;\theta_0)$*. Suppose* φ *is linear. That is,* $\varphi(x_i;\theta_0) = x_i'\theta_0$*. Then,* θ_0 *is identified if* $\operatorname{rank}(\mathbb{E}[x_i x_i']) = k$*.*

Proof. By the CEF decomposition theorem, $y_i = \mathbb{E}[y_i|x_i;\theta_0] + \varepsilon_i$, where $\mathbb{E}[\varepsilon_i|x_i;\theta_0] = 0$. This implies $y_i = x_i'\theta_0 + \varepsilon_i$, and $\mathbb{E}[x_i\varepsilon_i] = 0$. Premultiplying both sides of the linear equation by x_i and taking expectations, we have $\mathbb{E}[x_i y_i] = \mathbb{E}[x_i x_i']\theta_0$. Since $\operatorname{rank}(\mathbb{E}[x_i x_i']) = k$, θ_0 is uniquely defined by:

$$\theta_0 = (\mathbb{E}[x_i x_i'])^{-1}\mathbb{E}[x_i y_i]$$

□

6. (Identification in linear GMM)

Claim: Consider the following linear GMM model. We have a sample, $\{(y_i, x_i)\}_{i\in[n]}$ *and a set of instruments* $\{z_i\}_{i\in[n]}$*, where, for each* $i \in [N]$*,* y_i *is* $\mathbb{R}$*-valued,* x_i *is* $\mathbb{R}^K$*-valed,* z_i *is* $\mathbb{R}^L$*-valed. Suppose there is* $\theta_0 \in \mathbb{R}^K$ *such that, for each* $i \in [N]$*,* $y_i = x_i'\theta_0 + \varepsilon_i$*. Additionally, suppose that, for each* $i \in [N]$*,* $\mathbb{E}[z_i\varepsilon_i] = 0$*. Then, the following identification conditions are equivalent:*

$$\mathbb{E}[z_i x_i'] \text{ has full column rank} \qquad \iff \qquad \text{for every } \theta \neq \theta_0 : \mathbb{E}[z_i(y_i - x_i'\theta)] \neq 0$$

Proof. Consider the set of values $\theta \in \mathbb{R}^K$ satisfying the equation $\mathbb{E}[z_i(y_i - x_i'\theta)] = 0$. We can rewrite this equation as $\mathbb{E}[z_i x_i']\theta = \mathbb{E}[z_i y_i]$. If $\mathbb{E}[z_i x_i']$ has full column rank, then the associated linear map is injective, and hence the equation $\mathbb{E}[z_i x_i']\theta = \mathbb{E}[z_i y_i]$ has a unique solution, θ_0. Conversely, if there is some $\theta \neq \theta_0$ satisfying this equation, then the linear map is not injective, and hence $\mathbb{E}[z_i x_i']$ does not have full column rank. □

3 Asymptotic Normality

1. (Score and Hessian of objective function)

Claim: Consider the M-estimation framework, where the M-estimator $\hat{\theta}$ maximises $Q_n(\theta)$ on Θ, and:

$$Q_n(\theta) = n^{-1} \sum_{i=1}^{n}$$

Suppose the conditions of theorem 7.9 hold. Define the score and Hessian of observation i by:

$$s(w_i;\theta) = D_\theta m(w_i;\theta) \qquad H(w_i;\theta) = D_\theta^2 m(w_i;\theta)$$

Define the score and Hessian (without qualification) by:

$$s_n(\theta) = D_\theta Q_n(\theta) \qquad H_n(\theta) = D_\theta^2 Q_n(\theta)$$

Then:

$$\mathbb{E}[s_n(\theta_0)] = 0 \qquad \textit{and} \qquad -n^{-1}\mathbb{E}[H_n(\theta_0)] = \mathbb{E}[s_n(\theta_0)s_n(\theta_0)']$$

Proof. By the definition of Q_n and linearity of the derivative operator, the score $s_n(\theta_0)$ is the average of the scores $s(w_i;\theta_0)$ across observations. Since the score of each observation at θ_0 has expectation zero (assumption 3), $\mathbb{E}[s_n(\theta_0)] = 0$.

By a similar argument, the Hessian $H_n(\theta)$ is the average of the Hessians $H(w_i;\theta_0)$. Hence:

$$-\mathbb{E}[H_n(\theta_0)] = n^{-1} \sum_{i=1}^{n} \mathbb{E}[-H(w_i;\theta_0)]$$

By assumption 3, $-\mathbb{E}[H(w_i;\theta_0)] = \mathbb{E}[s(w_i;\theta_0)s(w_i;\theta_0)']$:

$$-\mathbb{E}[H_n(\theta_0)] = n^{-1} \sum_{i=1}^{n} \mathbb{E}[s(w_i;\theta_0)s(w_i;\theta_0)']$$

Using the fact that the score $s_n(\theta)$ is the average of the scores $s(w_i;\theta)$, we have:

$$\mathbb{E}[s_n(\theta_0)s_n(\theta_0)'] = \mathbb{E}\left[\left(n^{-1}\sum_{i=1}^{n} s(w_i;\theta_0)\right)\left(n^{-1}\sum_{i=1}^{n} s(w_i;\theta_0)\right)'\right]$$

Now, since the process is i.i.d. and the score of each observation at θ_0 have mean zero, the above sum can be simplified to:

$$\mathbb{E}[s_n(\theta_0)s_n(\theta_0)'] = n^{-2} \sum_{i=1}^{n} \mathbb{E}[s(w_i;\theta_0)s(w_i;\theta_0)']$$

Thus, we have:

$$-\mathbb{E}[H_n(\theta_0)] = n^{-1}\big(n^2\mathbb{E}[s_n(\theta_0)s_n(\theta_0)']\big)$$

which immediately implies:

$$-n^{-1}\mathbb{E}[H_n(\theta_0)] = \mathbb{E}[s_n(\theta_0)s_n(\theta_0)']$$

□

2. (Conditional information matrix equality for the linear regression model)

Claim: Consider the linear regression model with normal errors: $\{(y_i, x_i)\}_{i\in[n]}$ *is i.i.d.,* $y_i = x_i\beta_0 + \varepsilon_i$*, and* $\varepsilon_i|x_i \sim N(0, \sigma_0^2)$*. The score and Hessian of observation* i *are as defined above, where* $w_i = (y_i, x_i)$*,* $\theta = (\beta, \sigma^2)$*, and the* m *function is:*

$$m(w_i;\theta) = -\frac{1}{2}\ln(2\pi) - \ln(\sigma) - \frac{1}{2}\left(\frac{y_i - x_i'\beta}{\sigma}\right)^2$$

Then:

$$\mathbb{E}[s(w_i;\theta_0)|x_i] = 0 \qquad \textit{and} \qquad -\mathbb{E}[H(w_i;\theta_0)|x_i] = \mathbb{E}[s(w_i;\theta_0)s(w_i;\theta_0)'|x_i]$$

Proof. The score at θ_0 is:

$$s(w_i;\theta_0) = \begin{bmatrix} \frac{1}{\sigma_0^2}x_i\varepsilon_i \\ -\frac{1}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}{\varepsilon_i}^2 \end{bmatrix}$$

Since $\mathbb{E}[\varepsilon_i|x_i] = 0$, we have:

$$\mathbb{E}\left[\frac{1}{\sigma_0^2}x_i\varepsilon_i|x_i\right] = 0$$

Moreover, notice $\mathbb{E}[\varepsilon_i^2|x_i] = \mathbb{V}[\varepsilon_i|x_i] = \sigma_0^2$. Hence:

$$\mathbb{E}\left[-\frac{1}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}\varepsilon_i^2|x_i\right] = -\frac{1}{2\sigma_0^2} + \frac{1}{2\sigma_0^4}\mathbb{V}\left[\varepsilon_i^2|x_i\right] = 0$$

Thus, $\mathbb{E}[s(w_i;\theta_0|x_i)] = 0$. Now, the Hessian and outer product of the score at θ_0 of observation i are given by:

$$H(w_i;\theta_0) = \begin{bmatrix} -\frac{1}{\sigma_0^2}x_ix_i' & -\frac{1}{\sigma_0^4}x_i\varepsilon_i \\ -\frac{1}{\sigma_0^4}x_i'\varepsilon_i & \frac{1}{2\sigma_0^4} - \frac{3}{2\sigma_0^6}\varepsilon_i^2 \end{bmatrix}$$

$$s(w_i;\theta_0)s(w_i;\theta_0)' = \begin{bmatrix} \frac{1}{\sigma_0^4}x_ix_i'\varepsilon_i^2 & -\frac{1}{2\sigma_0^4}x_i\varepsilon_i + \frac{1}{2\sigma_0^6}x_i\varepsilon_i^3 \\ -\frac{1}{2\sigma_0^4}x_i'\varepsilon_i + \frac{1}{2\sigma_0^6}x_i'\varepsilon_i^3 & \frac{1}{4\sigma_0^4} - \frac{1}{2\sigma_0^6}\varepsilon_i^2 + \frac{1}{4\sigma_0^8}\varepsilon_i^4 \end{bmatrix}$$

Now, equality of the four elements of the conditional expectations of the negative Hessian and outer product of the score at θ_0 of observation i can easily be verified by straightforward algebra, noting that $\mathbb{E}[\varepsilon_i^3|x_i] = 0$ and $\mathbb{E}[\varepsilon_i^4|x_i] = 3\sigma_0^4$. □

3. (Asy.$\mathbb{V}$ for the linear regression model)

Claim: Consider the linear regression with normal errors (described above). Let $(\hat{\beta}, \hat{\sigma}^2)$ denote the ML estimator of (β, σ^2). Define the following estimator of the asymptotic variance of $(\hat{\beta}, \hat{\sigma}^2)$.

$$\widehat{Asy.\mathbb{V}[(\hat{\beta}, \hat{\sigma}^2)]} = -\left[n^{-1} \sum_{i=1}^{n} H(w_i; \hat{\beta}, \hat{\sigma}^2) \right]^{-1}$$

Then:

$$\widehat{Asy.\mathbb{V}[\hat{\beta}]} = \hat{\sigma}^2 \left[n^{-1} \sum_{i=1}^{n} x_i x_i' \right]^{-1}$$

Proof. The Hessian $H(w_i; \hat{\beta}, \hat{\sigma}^2)$ is given by the expression above, replacing ε_i with the corresponding residual $\hat{\varepsilon}_i$. So, the average Hessian can be written as:

$$n^{-1} \sum_{i=1}^{n} H(w_i; \hat{\beta}, \hat{\sigma}^2) = \begin{bmatrix} -\frac{1}{\hat{\sigma}^2} n^{-1} \sum_{i=1}^{n} x_i x_i' & -\frac{1}{\hat{\sigma}^4} n^{-1} \sum_{i=1}^{n} x_i \hat{\varepsilon}_i \\ -\frac{1}{\hat{\sigma}^4} n^{-1} \sum_{i=1}^{n} x_i' \hat{\varepsilon}_i & \frac{1}{2\hat{\sigma}^4} - \frac{3}{2\hat{\sigma}^6} n^{-1} \sum_{i=1}^{n} \hat{\varepsilon}_i^2 \end{bmatrix}$$

Denote the (j,k)th block of the above matrix by $\hat{\mathbb{V}}_{jk}$. Using the formula for the inverse of a partitioned matrix, the upper left block of $\widehat{\text{Asy.}\mathbb{V}[(\hat{\beta}, \hat{\sigma}^2)]}$, which is equal to $\widehat{\text{Asy.}\mathbb{V}[\hat{\beta}]}$, we have:

$$\widehat{\text{Asy.}\mathbb{V}[\hat{\beta}]} = \left[\hat{\mathbb{V}}_{11} - \hat{\mathbb{V}}_{12} \hat{\mathbb{V}}_{22}^{-1} \hat{\mathbb{V}}_{21} \right]^{-1}$$

Recalling the equivalence of ML and OLS in the linear regression model with normal errors, we can note that $\hat{\beta}$ minimises the sum of squared residuals, which implies the residual is such that the prediction is the orthogonal projection of $(y_1, \ldots, y_n)$ into the column space of $(x_1', \ldots, x_n')$. Hence:

$$\hat{\mathbb{V}}_{12} \propto \sum_{i=1}^{n} x_i \hat{\varepsilon}_i = 0$$

Thus, this implies:

$$\widehat{\text{Asy.}\mathbb{V}[\hat{\beta}]} = \left[\hat{\mathbb{V}}_{11} \right]^{-1} = \hat{\sigma}^2 \left[n^{-1} \sum_{i=1}^{n} x_i x_i' \right]^{-1}$$

□

4. (GMM with optimal orthogonality conditions)

Claim: Suppose the density $f(w_i; \theta)$ is such that the information matrix equality holds. Let the score and Hessian be defined as follows:

$$s(w_i; \theta) = D_\theta \ln f(w_i; \theta) \qquad H(w_i; \theta) = D_\theta^2 \ln f(w_i; \theta)$$

Suppose we estimate θ by GMM, with $g(w_i; \theta)$ given by:

$$g(w_i; \theta) = s(w_i; \theta)$$

and weighting matrix $\hat{W}$. Suppose further that the assumptions in proposition 7.10 (conditions for asymptotic normality of the GMM estimator) are satisfied. Then, the asymptotic variance of the GMM estimator is the inverse of the information matrix. That is:

$$Asy.\mathbb{V}[\hat{\theta}_{GMM}] = (G'WG)^{-1}G'WSWG(G'WG)^{-1} = \mathbb{E}[s(w_i;\theta_0)s(w_i;\theta_0)']^{-1}$$

where $G = \mathbb{E}[D_\theta g(w_i;\theta_0)]$ and S is the asymptotic variance of the sample average of $g(w_i;\theta_0)$.

Proof. Notice that in the case where the g function is the score, the number of orthogonality conditions is equal to the dimension of the parameter vector. Hence, the expected Hessian, G, is square. So, the asymptotic variance of the GMM estimator simplifies to:

$$\text{Asy.}\mathbb{V}[\hat{\theta}_{GMM}] = G^{-1}SG^{-1}$$

Moreover, by the information matrix equality, $G = \mathbb{E}[s(w_i;\theta_0)s(w_i;\theta_0)'] = S$. Thus, $G^{-1}SG^{-1} = G$, so the asymptotic variance is:

$$\text{Asy.}\mathbb{V}[\hat{\theta}_{GMM}] = \mathbb{E}[s(w_i;\theta_0)s(w_i;\theta_0)']^{-1}$$

□

5. (Taylor expansion of the sampling error for GMM)

Claim: Consider the GMM framework. Let $\hat{\theta}$ denote the GMM estimator corresponding with the sample orthogonality condition $g_n(\theta) = n^{-1}\sum_{i=1}^n g(w_i;\theta)$ and weighting matrix $\hat{W}$. Let θ_0 denote the true parameter value. Then, the mean value expansion of the sampling error, given by:

$$n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = -\Big[G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})\Big]^{-1}G_n(\hat{\theta})'\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^n g(w_i;\theta_0)\right]$$

can be written as:

$$n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = -\Psi^{-1}\Big[n^{\frac{1}{2}}D_\theta Q_n(\theta_0)\Big] + o_p(1)$$

Proof. Begin with the mean value expansion:

$$-\Big[G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})\Big]^{-1}G_n(\hat{\theta})'\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^n g(w_i;\theta_0)\right]$$

Write the inverse as:

$$\begin{aligned}
-\Big[G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})\Big]^{-1} &= -\Big[G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})\Big]^{-1} + [G'WG]^{-1} - [G'WG]^{-1} \\
&= -[G'WG]^{-1} + \left([G'WG]^{-1} - \Big[G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})\Big]^{-1}\right)
\end{aligned}$$

where G and W are the probability limits of $G_n(\theta_0)$ and W, respectively. Since $\hat{\theta}$ is consistent, G is also the probability limit of $G_n(\hat{\theta})$. Moreover, since $\bar{\theta}$ lies on the line segment between θ_0 and $\hat{\theta}$, G is the probability limit of $G_n(\bar{\theta})$. So, by the continuous mapping theorem:

$$[G'WG]^{-1} - \left[G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})\right]^{-1} \xrightarrow{p} 0$$

Hence, the inverse term can be written as:

$$-\left[G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})\right]^{-1} = -[G'WG]^{-1} + o_p(1)$$

Now, consider the remaining terms:

$$\begin{aligned}
&G_n(\hat{\theta})'\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)\right] \\
&= G_n(\hat{\theta})'\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)\right] + G_n(\theta_0)\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)\right] - G_n(\theta_0)\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)\right] \\
&= G_n(\theta_0)\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)\right] + \left[G_n(\hat{\theta}) - G_n(\theta_0)\right]G_n(\theta_0)\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)\right]
\end{aligned}$$

As argued previously, G is the common probability limit of $G_n(\theta_0)$ and $G_n(\hat{\theta})$. Moreover, by assumption, $n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)$ converges in distribution. Thus, applying Slutsky's theorem and noting that $D_\theta(\theta) = -G_n(\theta)\hat{W}n^{-1}\sum_{i=1}^{n} g(w_i;\theta_0)$, we have:

$$G_n(\hat{\theta})'\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)\right] = -n^{\frac{1}{2}}D_\theta Q_n(\theta_0) + o_p(1)$$

Combining these two expressions (for the inverse and the remaining term), we have:

$$\begin{aligned}
n^{\frac{1}{2}}(\hat{\theta} - \theta_0) &= \left[-[G'WG]^{-1} + \left([G'WG]^{-1} - \left[G_n(\hat{\theta})'\hat{W}G_n(\bar{\theta})\right]^{-1}\right)\right] \\
&\quad \cdot \left(-n^{\frac{1}{2}}D_\theta Q_n(\theta_0) + \left[G_n(\hat{\theta}) - G_n(\theta_0)\right]G_n(\theta_0)\hat{W}\left[n^{-\frac{1}{2}}\sum_{i=1}^{n} g(w_i;\theta_0)\right]\right) \\
&= [G'WG]^{-1}n^{\frac{1}{2}}D_\theta Q_n(\theta_0) + o_p(1)
\end{aligned}$$

Thus, letting $\Psi = -G'WG$, we have:

$$n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = -\Psi^{-1}n^{\frac{1}{2}}D_\theta Q_n(\theta_0) + o_p(1)$$

□