Asymptotic Efficiency

Robert A. Miller

Structural Econometrics

November 2021

- Suppose two estimators $\theta_N^{(1)}$ and $\theta_N^{(2)}$ have the same rate of convergence, and are both (asymptotically) unbiased for θ_0.
- Denote by $\Sigma^{(i)}$ the covariance matrix of $\theta_N^{(i)}$ for $i \in \{1, 2\}$.
- Then $\theta_N^{(1)}$ is more *efficient* than another estimator $\theta_N^{(2)}$ iff $\forall \delta \in \Theta$:

$$\delta' \Sigma^{(1)} \delta \leq \delta' \Sigma^{(2)} \delta \qquad \text{(abbreviated by } \Sigma^{(1)} \leq \Sigma^{(2)})$$

- This definition is a natural generalization of the efficiency definitions we use in linear estimation. For example consider:
  - GLS versus OLS
  - CLS versus OLS
  - OLS versus LAD

- Limiting the comparison to estimators with the same rate of convergence is not restrictive: estimators with faster convergence rates are always preferred to estimators with slower convergence rates if neither exhibits asymptotic bias
- If $\theta_N^{(1)}$ has a faster rate of convergence, say N^{ς}, to $\theta_N^{(2)}$, say N^{v} where $0 < v < \varsigma < 1$, then we would always prefer $\theta_N^{(1)}$ to $\theta_N^{(2)}$ since:

$$N^{v}\left(\theta_N^{(1)} - \theta_0\right) = N^{v-\varsigma}N^{\varsigma}\left(\theta_N^{(1)} - \theta_0\right) = o\left(1\right)O_p\left(1\right) = o_p\left(1\right)$$

- However insisting on (asymptotic) unbiasedness is problematic.
- We might trade some bias for reducing variance: the MSE does this.

- Recalling $A_0^* \equiv D_0 A_0$:

$$\begin{aligned} & \operatorname{ascov}\left(\sqrt{N}\left(\theta_N - \theta_0\right)\right) \\ \equiv \quad & \left(A_0^* D_0\right)^{-1} A_0^* \Sigma_0 A_0^{*-1\prime} \left(A_0^* D_0\right)^{-1\prime} \\ = \quad & \left(D_0 A_0 D_0\right)^{-1} D_0 A_0 \Sigma_0 \left(D_0 A_0\right)' \left(D_0 A_0 D_0\right)^{-1\prime} \end{aligned}$$

  the first question might be how to choose A_0

Lemma

If Σ_0 is non-singular, then for all A_0

$$\left(A_0^* D_0\right)^{-1} A_0^* \Sigma_0 A_0^{*\prime} \left(A_0^* D_0\right)^{-1\prime} \geq \left(D_0' \Sigma_0^{-1} D_0\right)^{-1}$$

Proof.

Factor $\Sigma_0 = CC'$ where C non-singular and symmetric. Define:

$$G \equiv \left(A_0^* D_0\right)^{-1} A_0^* C - \left(D_0' \Sigma_0^{-1} D_0\right)^{-1} D_0' C^{-1\prime} \tag{1}$$

$$\text{and note that } GC^{-1} D_0 = I - I = 0$$

Since GG' is positive semidefinite, the lemma now follows because:

$$\begin{aligned}
&\left(A_0^* D_0\right)^{-1} A_0^* \Sigma_0 A_0^{*\prime} \left(A_0^* D_0\right)^{-1\prime} \\
&\quad = \left(A_0^* D_0\right)^{-1} A_0^* CC' A_0^{*\prime} \left(A_0^* D_0\right)^{-1\prime} \\
&\quad = \left[\left(A_0^* D_0\right)^{-1} A_0^* C\right] \left[\left(A_0^* D_0\right)^{-1} A_0^* C\right]' \quad \text{(subst. from eq. (1))} \\
&\quad = \left[G + \left(D_0' \Sigma_0^{-1} D_0\right)^{-1} D_0' C^{-1}\right] \left[G + \left(D_0' \Sigma_0^{-1} D_0\right)^{-1} D_0' C^{-1}\right]' \\
&\quad = GG' + \left(D_0' \Sigma_0^{-1} D_0\right)^{-1}
\end{aligned}$$

Weighting Matrix

The optimal weighting matrix

- We call a weighting matrix optimal if the resulting:

$$\operatorname{ascov}\left(N^{1/2}\left(\theta_N - \theta_0\right)\right) = \left(D_0'\Sigma_0^{-1}D_0\right)^{-1}$$

Lemma

Setting $A_0^* = D_0'\Sigma_0^{-1}$ is optimal.

Proof.

Substitute $D_0'\Sigma_0^{-1}$ for A_0^* in:

$$\left(A_0^*D_0\right)^{-1}A_0^*\Sigma_0A_0^{*\prime}\left(A_0^*D_0\right)^{-1\prime}$$

to obtain:

$$\left(D_0'\Sigma_0^{-1}D_0\right)^{-1}\left(D_0'\Sigma_0^{-1}\Sigma_0\Sigma_0^{-1\prime}D_0\right)\left(D_0'\Sigma_0^{-1}D_0\right)^{-1} = \left(D_0'\Sigma_0^{-1}D_0\right)^{-1}$$

Weighting Matrix

Implementation

1. Construct a consistent, asymptotically normal estimator for θ_0 by using an arbitrary weighting matrix A_{N1}, such as the identity matrix, to obtain an estimate θ_{N1}.
2. Form consistent estimates of D_0 and Σ_0 with θ_{N1}, the first round estimate of θ_0.
3. In the second round use the weighting matrix $A_{N2} \equiv D_{N1}\Sigma_{N1}^{-1}$ where D_{N1} and Σ_{N1}^{-1} were formed in the previous step. This yields θ_{N2}, the updated estimate from the updated weighting matrix that efficiently weights the orthogonality conditions.
4. Then both $\left(D_{N1}'\Sigma_{N1}^{-1}D_{N1}\right)^{-1}$ and $\left(D_{N2}'\Sigma_{N2}^{-1}D_{N2}\right)^{-1}$ are consistent estimates of $ascov\sqrt{N}\left(\theta_{N2}-\theta_0\right) \equiv \left(D_0'\Sigma_0^{-1}D_0\right)^{-1}$, where D_{N2} and Σ_{N2}^{-1} are formed using θ_{N2}.

Orthogonality Conditions

The choice of orthogonality conditions

- Which orthogonality conditions should be used in estimation?
- We break the econometric aspects of the answer into two parts:

1. Given a conditional expectation derived from theory:

$$\mathrm{E}\left[h\left(x_n, \theta\right) \mid \mathcal{F}_n\right] = 0$$

what instruments $z_n \in \mathcal{F}_n$ should be used when forming orthogonality conditions of the form:

$$\mathrm{E}\left[z_n h\left(x_n, \theta\right)\right] = 0$$

in estimation?

2. Given the probability distributions underlying the data generating process (not just some conditional expectations), again derived from theory, how do we optimally form orthogonality conditions?

Orthogonality Conditions

A specialization

Consider the following specialization:

1. $(\Omega, \mathcal{F}, P)$ is a probability space, with elements $\omega \in \Omega$, or "world histories", and a sequence of σ-algebras $\mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \ldots \subseteq \mathcal{F}$.
2. $X_n(\omega)$ is stationary, ergodic, and $\mathcal{F}_n$-measurable.
3. $\omega^* \in \Omega$ is the realization of our world history.
4. $x_1, ..., x_N$ is observed, where $x_n \equiv X_n(\omega^*)$.
5. $\theta_0 \in \Theta$ is a parameter to be estimated.
6. Θ is a convex, compact subset of Euclidean space.
7. $h(x, \theta) : X \times \Theta \longrightarrow \mathbb{R}^m$ is known, measurable on Borel sets of X, continuously differentiable in θ.
8. $E[h(X_n(\omega), \theta_0) \mid \mathcal{F}_n] = 0$

- In other words there are m conditional expectation functions derived from the model, known up to a parameterization $\theta \in \Theta$.

Orthogonality Conditions

There are lots of orthogonality conditions when a conditional expectation equation holds

- If $z_n \in \mathcal{F}_n$ then for any measurable function of z_n denoted by $L(z_n)$:

$$\begin{aligned} 0 &= L(z_n) E[h(x_n, \theta_0) \mid \mathcal{F}_n] \\ &= E\{L(z_n) E[h(x_n, \theta_0) \mid \mathcal{F}_n]\} \\ \Rightarrow 0 &= E[L(z_n) h(x_n, \theta_0)] \end{aligned}$$

- Therefore even if there is only one instrument z_n, there is typically an uncountable infinity of orthogonality conditions.
- For example set $L(z_n) = z^{\alpha}$ for any $\alpha \in \mathbb{R}$.

Orthogonality Conditions

There are only as many optimal instruments as there are parameters

- However in the quadratic formulation of GMM, the FOC solves for the estimator by setting s equations to zero.
- Similarly when we focus on m orthogonality conditions, and q instruments we use a $s \times qm$ weighting matrix to reduce the number of equations we solve to s, the number of parameters to estimate.
- Rather than separately choosing q instruments and a $s \times qm$ weighting matrix, we combine the operations by choosing an $s \times m$ instrumental variable matrix Z_n to obtain an estimator for $\theta_N \in \mathbb{R}^s$ by solving:

$$N^{-1} \sum_{n=1}^{N} Z_n h\left(x_n, \theta_N\right) = 0$$

Orthogonality Conditions

The asymptotic covariance matrix when there are as many instruments as parameters

- When there are only s orthogonality conditions, the equation system defining θ_N with any A_0^* is solved by setting $A_0^* = I$, simplifying the covariance matrix to:

$$\operatorname{ascov}\left[N^{1/2}\left(\theta_N - \theta_0\right)\right] = D_0^{-1}\Sigma_0 D_0^{-1'}$$

where:

$$\begin{aligned} D_0 &= \mathrm{E}\left[\frac{\partial Z_n h\left(x_n, \theta_0\right)}{\partial \theta}\right] = \mathrm{E}\left[Z_n \frac{\partial h\left(x_n, \theta_0\right)}{\partial \theta}\right] \\ \Sigma_0 &= \mathrm{E}\left[Z_n h\left(x_n, \theta_0\right) h\left(x_n, \theta_0\right)' Z_n'\right] \end{aligned}$$

- We call the estimator defined by the conditional expectations functions formed off $h\left(x_n, \theta_0\right)$ optimal if the choice of Z_n minimizes:

$$\operatorname{ascov}\left[N^{1/2}\left(\theta_N - \theta_0\right)\right]$$

Lemma

$Z_n^* \equiv \Psi_n' \Phi_n^{-1}$ is the optimal instrument matrix, where:

$$\begin{aligned} \Phi_n &= E\left[h\left(x_n, \theta_0\right) h\left(x_n, \theta_0\right)' \mid \mathcal{F}_n\right] \\ \Psi_n &= E\left[\frac{\partial h\left(x_n, \theta_0\right)}{\partial \theta} \mid \mathcal{F}_n\right] \end{aligned}$$

and:

$$\Sigma_0 = E\left[\Psi_n' \Phi_n^{-1} \Psi_n\right] \equiv D_0^*$$

$$\operatorname{ascov}\left[N^{1/2}\left(\theta_N - \theta_0\right)\right] = E\left[\Psi_n' \Phi_n^{-1} \Psi_n\right]^{-1}$$

Proof.

[Proof (1 of 3)] Upon substituting $Z_n^* = \Psi_n' \Phi_n^{-1}$ for Z_n in Σ_0, we obtain:

$$\begin{aligned}
\Sigma_0 &\equiv E\left\{\Psi_n' \Phi_n^{-1} h\left(x_n, \theta_0\right) h\left(x_n, \theta_0\right)' \Phi_n^{-1\prime} \Psi_n\right\} \\
&= E\left\{\Psi_n' \Phi_n^{-1} E\left[h\left(x_n, \theta_0\right) h\left(x_n, \theta_0\right)' \mid \mathcal{F}_n\right] \Phi_n^{-1\prime} \Psi_n\right\} \\
&= E\left[\Psi_n' \Phi_n^{-1} \Phi_n \Phi_n^{-1} \Psi_n\right] \\
&= E\left[\Psi_n' \Phi_n^{-1} \Psi_n\right]
\end{aligned}$$

$$\begin{aligned}
\Rightarrow \quad & \operatorname{ascov}\left[N^{1/2}\left(\theta_N - \theta_0\right)\right] \\
= \quad & E\left[\Psi_n' \Phi_n^{-1} \frac{\partial h\left(x_n, \theta_0\right)}{\partial \theta}\right]^{-1} D_0^* E\left[\Psi_n' \Phi_n^{-1} \frac{\partial h\left(x_n, \theta_0\right)}{\partial \theta}\right]^{-1\prime} \\
= \quad & E\left\{\Psi_n' \Phi_n^{-1} E\left[\frac{\partial h\left(x_n, \theta_0\right)}{\partial \theta} \mid \mathcal{F}_n\right]\right\}^{-1} D_0^* E\left\{\Psi_n' \Phi_n^{-1} \frac{\partial h\left(x_n, \theta_0\right)}{\partial \theta}\right\}^{-1\prime} \\
= \quad & E\left[\Psi_n' \Phi_n^{-1} \Psi_n\right]^{-1} D_0^* E\left[\Psi_n' \Phi_n^{-1} \Psi_n\right]^{-1\prime} \\
= \quad & D_0^{*-1}
\end{aligned}$$

Proof.

[Proof (2 of 3)] For any other IV matrix choice Z_n, we now define:

$$\begin{aligned} G_n &\equiv D_0^{-1} Z_n h(x_n, \theta_0) - D_0^{*-1} Z_n^* h(x_n, \theta_0) \\ &= D_0^{-1} Z_n h(x_n, \theta_0) - D_0^{*-1} \Psi_n' \Phi_n^{-1} h(x_n, \theta_0) \end{aligned}$$

Abbreviating $h(x_n, \theta_0)$ by h_n, note that:

$$\begin{aligned} \mathrm{E}\left[G_n h_n' \Phi_n^{-1} \Psi_n\right] &= D_0^{-1} E\left[Z_n h_n h_n' \Phi_n^{-1} \Psi_n\right] \\ &\quad - D_0^{*-1} E\left[\Psi_n' \Phi_n^{-1\prime} h_n h_n' \Phi_n^{-1} \Psi_n\right] \\ &= D_0^{-1} E\left[Z_n \Psi_n\right] - D_0^{*-1} E\left[\Psi_n' \Phi_n^{-1} \Psi_n\right] \\ &= D_0^{-1} D_0 - I = 0 \end{aligned}$$

Proof.

[Proof (3 of 3)] Therefore:

$$\begin{aligned}
D_0^{-1}\Sigma_0 D_0^{-1\prime} &= D_0^{-1} E\left[Z_n h_n h_n' Z_n'\right] D_0^{-1\prime} \\
&= E\left[\left(G_n + D_0^{*-1}\Psi_n^{-1}\Phi_n^{-1} h_n\right)\left(G_n + D_0^{*-1}\Psi_n^{-1}\Phi_n^{-1} h_n\right)'\right] \\
&= E\left[G_n G_n'\right] + E\left[D_0^{*-1}\Psi_n'\Phi_n^{-1} h_n h_n'\Phi_n^{-1}\Psi_n D_0^{*-1\prime}\right] \\
&= E\left[G_n G_n'\right] + D_0^{*-1} E\left[\Psi_n'\Phi_n^{-1}\Psi_n\right] D_0^{*-1\prime} \\
&= E\left[G_n G_n'\right] + D_0^{*-1}
\end{aligned}$$

Asymptotic Efficiency

Fisher Information Matrix

- We now analyze the optimal orthogonality conditions when the DGP is known up to θ.
- First we revisit the case in which an unbiased estimator exists.
- Let $\mathcal{L}$ denote the likelihood, and define the Information matrix as the variance of the score:

$$-\left[E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial\theta\partial\theta'}\right)\right]$$

- The outer product of the derivatives equates to the Information matrix in expectation:

Lemma

For all $\theta \in \Theta$:

$$-\left[\mathrm{E}\left(\frac{\partial^2 \ln \mathcal{L}}{\partial\theta\partial\theta'}\right)\right] = \left[\mathrm{E}\left(\frac{\partial \ln \mathcal{L}}{\partial\theta}\frac{\partial \ln \mathcal{L}}{\partial\theta'}\right)\right]$$

Proof.

$$
\begin{aligned}
1 &= \int \mathcal{L}(x,\theta)\,\mathrm{d}x \\
0 &= \int \frac{\partial \mathcal{L}}{\partial \theta}(x,\theta)\,\mathrm{d}x = \int \mathcal{L}(x,\theta)\,\frac{\partial}{\partial \theta}\ln \mathcal{L}(x,\theta)\,\mathrm{d}x \qquad (2)
\end{aligned}
$$

$$
\begin{aligned}
\Rightarrow 0 &= \frac{\partial}{\partial \theta}\int \mathcal{L}\frac{\partial}{\partial \theta'}\ln \mathcal{L}\mathrm{d}x \\
&= \int \frac{\partial \mathcal{L}}{\partial \theta}\frac{\partial}{\partial \theta'}\ln \mathcal{L}\mathrm{d}x + \int \mathcal{L}\frac{\partial^2 \ln \mathcal{L}}{\partial \theta \partial \theta'}\mathrm{d}x \\
&= \int \frac{\partial \ln \mathcal{L}}{\partial \theta}\frac{\partial \ln \mathcal{L}}{\partial \theta'}\mathcal{L}\mathrm{d}x + E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta \partial \theta'}\right) \\
&= E\left(\frac{\partial \ln \mathcal{L}}{\partial \theta}\frac{\partial \ln \mathcal{L}}{\partial \theta'}\right) + E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial \theta \partial \theta'}\right) \qquad (3)
\end{aligned}
$$

Asymptotic Efficiency

Cramer-Rao Lower Bound

- The Cramer-Rao lower bound (CRLB) states that the variance of unbiased estimators is at least as big as the Information matrix.

Lemma (Cramer-Rao Lower Bound)

If $\theta_N(\omega)$ is unbiased for θ_0 with finite covariance matrix P, then:

$$P \geq -\left[E\left(\frac{\partial^2 \ln \mathcal{L}}{\partial\theta\partial\theta'}\right)\right]^{-1}$$

- Intuitively, the second derivative defines how quickly $\mathcal{L}$ declines from a stationary point, or how easy it is to pinpoint the maximum.
- The following lemma is useful in proving the CRLB.

Lemma

$$E\left[(\theta_N - \theta_0)\frac{\partial \ln \mathcal{L}}{\partial\theta'}\right] = I_{s\times s} \tag{4}$$

Proof.

[Proof of Lemma] If $\theta_N(\omega)$ is an unbiased estimator with finite covariance:

$$\begin{aligned} \theta_0 &= \int \mathcal{L}_N(\omega, \theta_0)\, \theta_N(\omega)\, d\omega \\ I_{s\times s} &= \frac{\partial}{\partial \theta_0} \int \mathcal{L}_N(\omega, \theta_0)\, \theta_N(\omega)\, d\omega \\ &= \int \frac{\partial \mathcal{L}_N(\omega, \theta_0)}{\partial \theta_0} \theta_N(\omega)\, d\omega \end{aligned} \tag{5}$$

$$\begin{aligned} \Rightarrow E\left[(\theta_N - \theta_0)\frac{\partial \ln \mathcal{L}}{\partial \theta'}\right] &= E\left[\theta_N \frac{\partial \ln \mathcal{L}}{\partial \theta'}\right] \quad \text{using (2)} \\ &= E\left[\theta_N \mathcal{L}^{-1} \frac{\partial \mathcal{L}}{\partial \theta'}\right] \\ &= \int \theta_N \frac{\partial \mathcal{L}}{\partial \theta} d\omega \\ &= I_{s\times s} \qquad \text{by (5)} \end{aligned}$$

Proof.

[Proof of CRLB]

$$\operatorname{cov}\begin{pmatrix}\theta_N \\ \frac{\partial \ln \mathcal{L}}{\partial \theta}\end{pmatrix} \equiv \begin{bmatrix} P & I_{s\times s} \\ I_{s\times s} & R \end{bmatrix}$$

where the off diagonal elements exploit (4) and:

$$P \equiv \mathrm{E}\left[(\theta_N - \theta_0)(\theta_N - \theta_0)'\right] \text{ and } R \equiv \mathrm{E}\left[\frac{\partial \ln \mathcal{L}}{\partial \theta}\frac{\partial \ln \mathcal{L}}{\partial \theta'}\right]$$

Assuming $R > 0$ (that is $\frac{\partial \ln \mathcal{L}}{\partial \theta}$ is not degenerate), R^{-1} exists and since covariance matrices are positive definite:

$$\begin{aligned} 0 &< \begin{bmatrix} I_{s\times s} & -R^{-1} \end{bmatrix} \begin{bmatrix} P & I_{s\times s} \\ I_{s\times s} & R \end{bmatrix} \begin{bmatrix} I_{s\times s} \\ -R^{-1} \end{bmatrix} \\ &= P - R^{-1} - R^{-1} + R^{-1}RR^{-1} \\ &= P - R^{-1} \\ &\Rightarrow P \geq R^{-1} \end{aligned}$$

Asymptotic Efficiency

CRLB for nonlinear models

- These results for unbiased estimators have a large sample analogue in consistent estimators.
- A consistent estimator is called asymptotically efficient if it's covariance matrix is the information matrix (and obtains the CRLB).
- Except on Lebesque measure zero, the information matrix bounds from below all consistent and asymptotically normal estimators (Le Cam, 1953).

Lemma

Suppose $x_n(\omega)$ is an independent process. Then ML is asymptotically efficient.

Proof.

The $s \times 1$ FOC is:

$$0 = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial \ln \mathcal{L}(x_n, \theta)}{\partial \theta} \qquad \text{(after multiplying through by } 1/N\text{)}$$

Recall that ML is a GMM estimator, and the asymptotic distribution of GMM estimators is

$$N^{1/2}\left(\theta_{ML}^{(N)} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, D_0^{-1} \Sigma_0 D_0^{-1\prime}\right)$$

where

$$D_0 = \mathrm{E}\left[\frac{\partial \ln L^2}{\partial \theta \partial \theta'}\right] \text{ and } \Sigma_0 = \mathrm{E}\left[\frac{\partial \ln L}{\partial \theta} \frac{\partial \ln L'}{\partial \theta}\right]$$

But we already know that $D_0^{-1} = -\Sigma_0$, so result follows:

$$N^{1/2}\left(\theta_{ML}^{(N)} - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, -D_0^{-1}\right)$$

Asymptotic Efficiency in Two Steps

The Newton-Raphson Algorithm

- Suppose $Q_N(\theta)$ is the criterion function for an M estimator satisfying standard regularity conditions.
- The Newton-Raphson algorithm is defined by its steps:

$$\theta^{(i+1)} = \theta^{(i)} - \left[\partial^2 Q_N\left(\theta^{(i)}\right)/\partial\theta\partial\theta'\right]^{-1}\left[\partial Q_N\left(\theta^{(i)}\right)/\partial\theta\right]$$

  where N indicates the sample, θ is the parameter value,
- This algorithm converges to the maximand if the criterion function is strictly concave, and/or if $\theta^{(i)}$ is close enough to the maximum.
- It is based on the quadratic approximation:

$$\begin{aligned} Q_N(\theta) \simeq\ & Q_N\left(\theta^{(i)}\right) + \left[\partial Q_N\left(\theta^{(i)}\right)/\partial\theta\right]'\left(\theta-\theta^{(i)}\right) \\ & + \frac{1}{2}\left(\theta-\theta^{(i)}\right)'\left[\partial^2 Q_N\left(\theta^{(i)}\right)/\partial\theta\partial\theta'\right]\left(\theta-\theta^{(i)}\right) \end{aligned}$$

Asymptotic Efficiency in Two Steps

Iterating an extra step (Amemiya, 1985, pp 137 - 139)

- Suppose $\theta^{(1)}$ is a $\sqrt{N}$ consistent estimator for the (interior) true value $\theta_0 \in \Theta$, the parameter space.
- Then it is well known that $\theta^{(2)}$ has the same asymptotic properties as $\theta^{(\infty)}$, the limit of the sequence, namely:

$$\sqrt{N}\left(\theta^{(2)} - \theta_0\right) \xrightarrow{d} \left[p\lim \frac{1}{N}\frac{\partial^2 Q_N\left(\theta^{(1)}\right)}{\partial\theta\partial\theta'}\right]^{-1}\left[p\lim \frac{1}{\sqrt{N}}\frac{\partial Q_N\left(\theta^{(1)}\right)}{\partial\theta}\right]$$

Specializing $Q_N(\theta) = \log L_N(\theta)$, the log likelihood, $\theta^{(2)}$ is asymptotically efficient, and $\sqrt{N}\left(\theta^{(2)} - \theta_0\right)$ is asymptotically normal with mean zero and covariance:

$$-\left\{p\lim \frac{1}{N}\frac{\partial^2 \log L_N(\theta_0)}{\partial\theta\partial\theta'}\right\}^{-1} = -\left\{E\left[\frac{\partial^2 \log L_N(\theta_0)}{\partial\theta\partial\theta'}\right]\right\}^{-1}$$