Notes on GMM, MLE and Minimum Distance Estimation Based on Newey-McFadden (1994)

As in previous notes, define an extremum estimator \( \hat{\theta}_n \) as the optimizer of the objective function \( \hat{Q}_n(\theta), \theta \in \Theta \)

(a) MLE
\[
\hat{Q}_n(\theta) = \frac{1}{N} \sum_{i=1}^{N} \ell_n f(y_i \mid \theta)
\]

(b) NLLS
\[
\hat{Q}_n(\theta) = \frac{-1}{N} \sum_{i=1}^{N} (y_i - h(x_i; \theta))^2
\]

(c) GMM

Based on a moment function \( g(y; \theta) \)
\[ E[g(y; \theta)]_{\theta=\theta_0} = 0 \]
\[ \int g(y; \theta) f(y; \theta_0) dy = 0 \]

Sometimes get exact solutions:
\[ \hat{Q}_n(\theta) = \left[ -\frac{1}{N} \sum_{i=1}^{N} g(y_i; \theta) \right] \hat{\mathbf{W}} \left[ \frac{1}{N} \sum g(y_i; \theta) \right] \]
\( \hat{\mathbf{W}} \) is positive semi definite classical

(d) Minimum distance (CMD)
\[ \hat{Q}_n(\theta) = -[\hat{\Pi} - h(\theta)]' \hat{\mathbf{W}} [\hat{\Pi} - h(\theta)] \]
\[ \hat{\Pi} \overset{P}{\rightarrow} \Pi \quad h(\theta) \text{ gives mapping from structure to reduced form.} \]
\( \hat{\mathbf{W}} \) positive semifinite.

More general class: CMD
\[ \hat{Q}_n(\theta) = -\hat{g}_n(\theta)' \hat{W} \hat{g}_n(\theta) \]

GMM when \( \hat{g}_n = \frac{1}{N} \sum g(y_i; \theta) \)

CMD when \( \hat{g}_n(\theta) = \hat{\Pi} - h(\theta) \)

FOC for Extremum: \( \frac{\partial Q_n}{\partial \theta} = 0 \)

\[ \therefore \text{MLE with } \hat{g}_n(\theta) = \frac{\partial Q_n}{\partial \theta}. \]

Observe MLE

\[ \int f(y; \theta_0) dy = 0. \]

If you can differentiate under the integral

sign, you have
\[ 0 = \int \frac{\partial f(y; \theta_0)}{\partial \theta} dy \]

\[ 0 = \int \left[ \frac{\partial \ln f(y; \theta)}{\partial \theta} \right]_{\theta = \theta_0} f(y; \theta_0) dy \]

\[ \uparrow \text{score vector:} \]

:. MLE is GMM with \( g(y; \theta) = \frac{\partial \ln f}{\partial \theta} \)

NLS is GMM with

\[ g(y; \theta) = -2[y - h(x; \theta)]' \frac{\partial h(x; \theta)}{\partial \theta}. \]

Consistency Theorem for Extremum Estimator given before and Need for Uniform Convergence was given before. Asymptotic normality established as before in terms of expansions of the \( Q \) function.
Asymptotically Linear Estimator:

\[ \sqrt{N} (\hat{\theta} - \theta_0) = \sum_{i=1}^{N} \frac{\psi(y_i)}{\sqrt{N}} + o_p(1) \]

\[ E(\psi(y_i)) = 0 \quad E(\psi(y_i)\psi(y'_i)) \text{ exists.} \]

\( \psi \) is called influence function; and we have that \( \psi(y_i) \) is score vector transformed. Thus we have that at Max:

\[ 0 = \frac{\partial Q_N}{\partial \theta} \bigg|_{\theta=\theta_0} + \frac{\partial^2 Q_N}{\partial \theta \partial \theta'}(\hat{\theta}_T - \theta_0) \]

\[ \sqrt{N}(\hat{\theta}_N - \theta_0) = - \left( \frac{\partial^2 Q_N}{\partial \theta \partial \theta'} \right)^{-1} \left( \frac{\partial Q_N}{\partial \theta} \right) \sqrt{N} \]

of this form.

Locally Linear. (Observe that \( Q_N \) here
includes \( \frac{1}{N} \) factor; other notes do not therefore normalizations shift around).

Asymptotic normality for an extremum estimator has just been established. (See previous notes).

Asymptotic normality for general minimum distance estimators.

**Theorem 1:** Suppose \( \hat{\theta}_n = \arg \max \{ -g_n(\theta)' \hat{W} g_n(\theta) \} \) where \( \hat{W} \overset{P}{\to} W \); \( W \) is positive semidefinite; \( \hat{\theta} \overset{P}{\to} \theta_0 \) and (i) \( \theta_0 \in \text{Interior} (\theta) \) (ii) \( \hat{g}_n(\theta) \) is continuous differentiable in \( \text{Nbd} \)
\[
N \text{ of } \theta_0 \left( \hat{G} = \hat{g}_n(\theta) \right)
\]

(iii) \(\sqrt{N} g_N(\theta_0) \overset{d}{\longrightarrow} N(0, \Omega)\)

(iv) \(\exists G'(\theta)\) continuous at \(\theta_0\) and
\[
\sup_{\theta \in N} \left\| \frac{\partial \hat{g}_n(\theta)}{\partial \theta} - G(\theta) \right\| \overset{P}{\longrightarrow} 0
\]

(v) For \(G = G(\theta_0)\), \(G'WG\) is nonsingular.

Then \(\sqrt{N}(\hat{\theta} - \theta_0) \overset{d}{\longrightarrow} N(0, (G'WG)^{-1}G'W\Omega WG(G'WG)^{-1})\)

**Proof:** Using (i) and (ii), w.p.1 FOC

\[
\hat{G}'(\hat{\theta})' \hat{W} \hat{g}_N(\hat{\theta}) = 0
\]

for \(\hat{G}(\theta) = \frac{\partial g_N(\hat{\theta})}{\partial \theta}\). Expand \(\hat{g}_n(\hat{\theta})\) around \(\theta_0\):

\[
\hat{G}'(\hat{\theta}) \hat{W} \left[ \hat{g}_N(\theta_0) + \frac{\partial g_N(\theta_0)}{\partial \theta}(\hat{\theta} - \theta_0) \right] = 0
\]
\[ \hat{G}'(\hat{\theta}) \hat{W} [\hat{g}_N(\theta_0) + \hat{G}(') (\hat{\theta} - \theta_0)] = 0 \]

\[ \therefore \sqrt{N} (\hat{\theta} - \theta_0) = -\left[ \hat{G}'(\hat{\theta}) \hat{W} \hat{G}(\hat{\theta}) \right]^{-1} \sqrt{N} \hat{G}'(\hat{\theta}) \hat{W} \hat{g}_N(\theta_0) \]

\[ \hat{G}(\hat{\theta}) \xrightarrow{P} G \quad \hat{G}(\bar{\theta}) \longrightarrow G \]

\[ \therefore \text{using Slutsky we get} \]
\[ \sqrt{N} (\hat{\theta} - \theta_0) = -\left[ G'(\theta) W G(\theta) \right]^{-1} \sqrt{N}(G W g_N(\theta_0)). \]

QED. We did this before with the MLE:

**Proof for MLE here.**

Applies to GMM:

\[ \hat{W} \xrightarrow{P} W \text{ and} \]

(i) \(W\) is positive semidefinite

(ii) \(W \ E(g(y; \theta)) = 0 \iff \theta = \theta_0\)
(iii) \( \theta_0 \in \theta \) compact

(iv) \( g(y; \theta) \) continuous in \( \theta \) at each \( \theta \in \theta \) w.p.1.

(v) \( E[\sup_{\theta \in \theta} || g(y; \theta) ||] < \infty \), then \( \theta \xrightarrow{P} \theta_0 \).

Follow from previous results.

**Note: Influence Function for GMM**

\[
\psi(y) = -(G'WG)^{-1}G'W g_n(y; \theta_0).
\]

**One Step Theorem:**

Let \( \bar{\theta} \) be an initial estimator. Let \( \bar{H} \) be a consistent estimator of

\[
H = \text{plim} \left[ \frac{\partial^2}{\partial \theta \partial \theta'} \hat{Q}_N(\theta_0) \right]
\]

\[
\tilde{\theta} = \bar{\theta} - \bar{H}^{-1} \frac{\partial \hat{Q}_N(\bar{\theta})}{\partial \theta}.
\]
For $\bar{H} = \frac{\partial^2 Q_N(\tilde{\theta})}{\partial \theta \partial \theta'}$ : Newton Raphson Iteration.
Theorem 2: Suppose $\sqrt{N}(\bar{\theta} - \theta_0)$ is bounded in probability and $\tilde{\theta}$ is as defined above and asymptotic normality theorem satisfied (for extremum estimator) and

A. $\bar{H} = \frac{\partial^2 Q_N(\bar{\theta})}{\partial \theta \partial \theta'}$

or

B. $\bar{H} \xrightarrow{P} H$.

Then if

$\sqrt{N}(\bar{\theta} - \theta_0) \xrightarrow{d} N\left(0, \left(E \frac{\partial^2 Q}{\partial \theta \partial \theta'}\right)^{-1} E \left(\frac{\partial Q}{\partial \theta} \frac{\partial Q}{\partial \theta'}\right) E \left(\frac{\partial^2 Q}{\partial \theta \partial \theta'}\right)^{-1}\right)$.

If
\[ \tilde{\theta} = \bar{\theta} - (\tilde{G}'WG')^{-1}\tilde{G}'\hat{W}\hat{g}_N(\bar{\theta}) \]

and GMM asymmetric normality satisfied and

either

A. \[ \bar{G} = \frac{\partial}{\partial \theta}\hat{g}_N(\bar{\theta}) \]

or

B. \[ \bar{G} \xrightarrow{P} G \]

Then

\[ \sqrt{N}(\tilde{\theta} - \theta_0) \xrightarrow{d} N(0, (G'WG)^{-1}G'W\OmegaWG(GWG)^{-1}). \]

**Proof:** Expand \( \frac{\partial Q}{\partial \theta}(\bar{\theta}) \) in nbd of \( \theta_0 \):

\[ \tilde{\theta} = \bar{\theta} - \bar{H}^{-1} \left[ \frac{\partial Q_N(\theta_0)}{\partial \theta} + \frac{\partial^2 Q(\bar{\theta})}{\partial \theta \partial \theta'}(\bar{\theta} - \theta_0) \right] \]

\[ \sqrt{N}(\tilde{\theta} - \theta_0) = \sqrt{N}(\bar{\theta} - \theta_0) \]
\[-\bar{H}^{-1} \frac{\partial^2 Q(\dot{\theta})}{\partial \theta \partial \theta'} \sqrt{N}(\bar{\theta} - \theta_0) - \bar{H}^{-1} \frac{\partial \hat{Q}_N(\theta_0)}{\partial \theta}\]

(\dot{\theta}) is mean value. \(\bar{H} \rightarrow H\)

Second term converges to

\[N \left(0, \left( E \frac{\partial^2 Q}{\partial \theta \partial \theta} \right)^{-1} E \left[ \frac{\partial Q}{\partial \theta} \frac{\partial Q'}{\partial \theta} \right] \cdot \left( E \frac{\partial^2 Q}{\partial \theta \partial \theta'} \right)^{-1} \right).\]

But first term is

\[\left[ I - \bar{H}^{-1} \frac{\partial^2 Q(\dot{\theta})}{\partial \theta \partial \theta'} \right] \sqrt{N}\]

\((\bar{\theta} - \theta_0) \rightarrow 0.\)

This term is Bounded in Probability

\[\therefore\] asymptotically efficient.
GMM Form:
\[
\sqrt{N}(\hat{\theta} - \theta_0) = 
\left[ I - (\bar{G}' \hat{W} \bar{G})^{-1} \bar{G}' \hat{W} \frac{\partial g_N(\theta)}{\partial \theta} \right] \cdot \sqrt{N}(\hat{\theta} - \theta_0) 
- (\bar{G}' \hat{W} \bar{G})^{-1} \bar{G}' \hat{W} \sqrt{N}(\hat{g}_N(\theta_0))
\]

and first term goes as before.

Asymmetric Efficiency of MLE:

Cramer Rao Lower Bound:
\[
\sqrt{N} (\hat{\theta} - \theta_0)_{\text{MLE}} \sim 
N \left( 0, \left[ E \left( \frac{\partial \ln f}{\partial \theta} \frac{\partial \ln f}{\partial \theta} \right) \right]^{-1} \right)
\]

\[
S = \frac{\partial \ln f}{\partial \theta}
\]
is score vector (for a single observation).
Asymmetric Variance GMM:

\[(E[M_\theta])^{-1} E(M M')(E M'_\theta)^{-1}\]

\[M_\theta = E \left( \frac{\partial g(y; \theta_0)}{\partial \theta} \right) W \frac{\partial g(y; \theta_0)}{\partial \theta}\]

\[M = E \left( \frac{\partial g(y; \theta_0)}{\partial \theta} \right) W g(y; \theta_0)\]

GMM Moment Condition:

\[
\int g(y; \theta) f(y; \theta) dy = 0.
\]

Differentiate under the integral sign:

\[
\int \left[ \frac{\partial g(y; \theta)}{\partial \theta} \right] f(y; \theta) dy + \int g(y; \theta) \frac{\partial f(y; \theta)}{\partial \theta} dy = 0
\]

\[E \left[ \frac{\partial g(y; \theta)}{\partial \theta} \right]_{\theta_0} + 15\]
\[
\text{COV}\left(g(y; \theta), \frac{\partial \ln f(y; \theta)}{\partial \theta}\right)_{\theta_0} = 0
\]

when \(g(y; \theta) = \frac{\partial \ln f(y; \theta)}{\partial \theta}\) produces

\[
E\left[\frac{\partial^2 \ln f(y; \theta)}{\partial \theta \partial \theta'}\right]_{\theta = \theta_0} = -E\left[\frac{\partial \ln f}{\partial n} \frac{\partial \ln f}{\partial \theta'}\right]_{\theta = \theta_0}.
\]

Thus

\[
E(M_\theta) + E(MS') = 0
\]

(just multiply through by \(E(\nabla_\theta g(y; \theta))'W\)).

Thus difference of GMM and MLE asymptotic variances
\[(E[M_\theta])^{-1}E(MM')[E(M_\theta)]^{-1} - (E[SS'])^{-1}\]

\[
= [E(MS')]^{-1}E(MM')[E(SM')]^{-1} - (E(SS'))^{-1} \\
= [E(MS')]^{-1}[E(MM') - E(MS')] \\
= [E(SS')]^{-1}E(SM')[E(SM')]^{-1} \\
= [E(MS')]^{-1}[E UU']E(SM') \\
U = M - E(MS')[E(SS')]^{-1}S
\]

\[\therefore \text{MLE is efficient within the class of GMM.}\]

Now what is the efficient weight matrix
within OMD framework? Let $Z$ be any random vector such that

$\Omega = E[ZZ']$ define
\[ M = G'WZ \]
\[ \tilde{M} = G'\Omega^{-1}Z \]

Then
\[ G'WG = E(M\tilde{M}') \]
\[ G'\Omega^{-1}G = E(\tilde{M}\tilde{M}') \]

Then we have that
\[
(G'WG)^{-1}G'W\OmegaWG(G'WG)^{-1} - (G'\Omega^{-1}G)^{-1} = (G'WG)^{-1}E(UU')(GWG)^{-1}
\]

where
\[ U = m - E(M\tilde{M}')[E(\tilde{M}\tilde{M}')]^{-1}\tilde{M} \]

since \((G'\Omega^{-1}G)^{-1}\) is the minimum distance estimator when \(W = \Omega^{-1}\), this is minimal.
**Proof:**

\[ UU' = MM' - M\tilde{M}'[E(\tilde{M}\tilde{M}')]^{-1}E(\tilde{M}M') \]

\[ -E(M\tilde{M}')[E(\tilde{M}\tilde{M}')]^{-1}\tilde{M}M' \]

\[ + E(M\tilde{M}')[E(\tilde{M}\tilde{M}')]^{-1}(\tilde{M}\tilde{M}') \cdot \]

\[ (E(\tilde{M}\tilde{M}'))^{-1}E(\tilde{M}'M) \]

\[ E(UU') = \]

\[ (G'W'\Omega WG) - (G'WG)(G'\Omega^{-1}G)^{-1}G'WG. \]

Rest follow

\[ U \text{ is residual of } G'WZ \text{ on a regression of } G'WZ \text{ on } G'\Omega^{-1}Z. \]

Applications: \( \hat{g}(\theta) = \hat{\pi} - \pi(\theta) \) efficient
weighting matrix is inverse of asymptotic variance of $\pi$.

**GMM:**

$$\hat{g}(\theta) = \frac{1}{N} \sum_{i=1}^{N} g(y_i; \theta)$$ efficient weighting matrix:

$$W^{-1} = \Omega$$

$$\Omega = E(g(y; \theta)g'(y; \theta)).$$

**Applications:**

**Direct Argument: GMM**

$$\frac{1}{\sqrt{N}} \sum g(y; \theta) \xrightarrow{d} N(0, \Omega_0)$$

weighting matrix $\{a_N\} \xrightarrow{P} a$

$g$ has continuous first derivative.

$$G = E \left[ \frac{\partial g(y; \theta)}{\partial \theta} \right]$$ is well defined.
Taylor Expansions:

\[
0 = \frac{1}{\sqrt{N}} \sum g(y; \hat{\theta}) = \frac{1}{\sqrt{N}} \sum_{n=1}^{N} g(y; \theta_0) \\
+ \frac{1}{N} \sum_{n=1}^{N} \frac{\partial g(y; \bar{\theta})}{\partial \theta}[(\theta - \theta_0) \sqrt{N}].
\]

Multiply both sides by \(a_N\) (nonsingular)

\[
a_N \frac{1}{\sqrt{N}} \sum_{n=1}^{N} g(y; \hat{\theta}) = 0
\]

call \(G_N = \left( \frac{1}{N} \sum_{n=1}^{N} \frac{\partial g(y; \bar{\theta})}{\partial \theta} \right)\)

\[
\therefore 0 = a_N
\]

\[
\frac{1}{\sqrt{N}} \sum g(y; \theta_0) + a_N G_N (\hat{\theta} - \theta_0) \sqrt{N}
\]

\[
\sqrt{N}(\hat{\theta} - \theta_0) a_N = -(a_N G_N)^{-1} a_N
\]

\[
\frac{1}{\sqrt{N}} \sum_{n=1}^{N} g(y; \theta_0)
\]

22
\[ \sqrt{N}(\hat{\theta} - \theta_0)_{aN} \xrightarrow{d} N(0; (aG')^{-1}a\Omega_0a'(aG')^{-1}). \]

We seek to minimize the asymptotic variance

\[ \text{Min} [(a'G)^{-1}]'a \Omega_0a'(aG)^{-1}. \]

We can always choose \( a \) so that \( aG = I \) (this is a nonsingular matrix)

:. reformulate problem multiply by a nonsingular matrix

\[ \text{Min} a\Omega_0 a' + \lambda(I - G'a') \]

**FOC:**

\[ a\Omega_0 = \lambda G'' \]

\[ a = \lambda G''\Omega_0^{-1} \]

\[ I = aG = \lambda G''\Omega_0^{-1}G \]

23
\[ a^* = (G'\Omega_0^{-1}G)^{-1}G'\Omega_0^{-1} \]

so with this choice of weight matrix, we obtain
\[ \sqrt{N}(\hat{\theta} - \theta_0)_{a^*} \overset{d}{\rightarrow} N(0, (G'\Omega_0^{-1}G)^{-1}). \]

Now take any other weight matrix with
\[ aG = I \]
\[ \sqrt{N}(\hat{\theta} - \theta_0)_a = N(0, a\Omega_0 a') \]
\[ (a\Omega_0 a') - (G'\Omega_0^{-1}G)^{-1} = a\Omega_0. \]

Verification is a form of Gauss Markov Theorem.

Observe that
\[ \sqrt{N} (\hat{\theta}_a - \theta_0) = a \sqrt{N} G \]
\[ \sqrt{N} (\hat{\theta}_{a^*} - \theta_0) = a^* \sqrt{N} G \]
\[
\sqrt{N} (\hat{\theta}_a - \hat{\theta}_{a^*}) = (a - a^*) \sqrt{N} G \\
\text{Var}(\sqrt{N} (\hat{\theta}_a - \hat{\theta}_{a^*})) = (a - a^*) v_0 (a - a^*)'
\]

Observe that cross terms have special property:

\[
a \Omega_0 (a^*)' = a \Omega_0 (\Omega_0^{-1})' G (G' v_0^{-1} G)^{-1} \\
= (aG) = (aG) (G' \Omega_0^{-1} G)^{-1} \\
= (G' \Omega_0^{-1} G) = a^* \Omega_0 a^* \\
\therefore (a - a^*) \Omega_0 (a - a^*)' = a \Omega_0 a' - a^* \Omega_0 (a^*)'
\]

But this is now negative. \therefore variance is bigger for anything but \(a^*\).

**IV Example:**

\[
Y_t = X_t' \beta_0 + U_t \\
1 \times K \\
K \times 1 \\
25
\]
\[ E(Z_t'U_t) = 0 \]

**IV Condition:**

\[ g_t = Z_t'(y_t - X_t\beta) = 0 \quad \int \iff \beta = \beta_0 \]

\[ G_t = -E(Z_t'X_t') \]

\[ G = \frac{-1}{T} \sum_{t=1}^{T} Z_t'X_t' \]

\[ \Omega = \text{Var} \left( \frac{\sum Z_tU_t}{\sqrt{T}} \right) = \frac{1}{T} \sum \sigma_t^2 Z_tZ_t' \]

if \( U_t \) is i.i.d mean zero we have that

\[ W_T = \frac{1}{T} \sum_{t=1}^{T} \sigma_t^2 Z_tZ_t'. \]

\[ a_T = \left( \frac{1}{T} \sum Z_tX_t' \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} Z_tZ_t'\sigma_t^2 \right)^{-1}. \]

IV estimator defined from estimation
equation:

\[
\frac{1}{T} \left( \sum_{t=1}^{T} Z_t X'_t \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^{T} Z_t Z'_t \sigma_t^2 \right)^{-1} \left( \sum_{t=1}^{T} Z_t (y_t - X_t \hat{\beta}_{IV}) \right) = 0
\]

just ordinary IV Equation. Nonlinear IV

g_T(\beta)(W_T)^{-1}g_T(\beta).

**Numerical Solutions: 2 Stage Estimation:**

**GMM Estimator:**

\[ (*) \frac{1}{N} \sum_{i=1}^{N} g(y_i, \theta, \hat{\gamma}) = 0 \]

\( \hat{\gamma} \) is a first step estimator. Obtained from
\[ (** \) \] \[ \frac{1}{N} \sum_{i=1}^{N_1} M(y_i; \gamma) = 0 \]  

(assume same number of moments as regressor)  

\[ \tilde{g} = [M(y, \gamma)'g(y, \theta, \gamma)']' \]  

Stack: \[ \frac{1}{N} \sum \tilde{g}(y_i, \hat{\theta}, \hat{\gamma}) = 0. \]  

Let  

\[ G_\theta = E \left[ \frac{\partial g(y, \theta, \gamma)}{\partial \theta} \right]_{\theta_0, \gamma_0} = 0 \]  

\[ G_\gamma = E \left[ \frac{\partial g(y, \theta, \gamma)}{\partial \gamma} \right]_{\theta_0, \gamma_0} \]  

\[ g(y) = g(y, \theta_0, \gamma_0) \]  

\[ \bar{M} = E \left[ \frac{\partial M(y, \gamma)}{\partial \gamma} \right]_{\gamma=\gamma_0} \]  

\[ \psi(y) = \bar{M}^{-1}M(y, \theta_0). \]  

28
Theorem 3: If (*) and (**) are satisfied with probability approaching 1, $\hat{\theta} \xrightarrow{P} \gamma_0$, $\hat{\gamma} \xrightarrow{P} \gamma_0$ and $\tilde{g}$ satisfies conditions stated above for consistency of GMM, then $\hat{\theta}$ and $\hat{\gamma}$ are asymptotically normal. Let $N = N$, then

$$\sqrt{N}(\hat{\theta} - \theta_0) \xrightarrow{d} N(0, V)$$

$$V = G_0^{-1}E\{[g(y) + G_\gamma \psi(y)]'(G_0^{-1})'\}.$$

**Proof:** $\hat{\theta}, \hat{\gamma}$ is a GMM estimator with moment functions:

$$\tilde{g}(y, \theta, \gamma) = [M(y, \gamma)'g(y, \theta, \gamma)']\hat{W} = I.$$

$$(\tilde{G}'I\tilde{G})^{-1}\tilde{G}' = \tilde{G}^{-1}.$$
Asymmetric variance of estimator

\[
(\tilde{G}'I\tilde{G})^{-1}\tilde{G}'IE[\tilde{g}(y, \theta_0, \gamma_0)\tilde{g}(y, \theta_0, w)](I\tilde{G})(\tilde{G}'I\tilde{G})^{-1}
= \tilde{G}^{-1} E[\tilde{g}(y, \theta_0, \gamma_0)\tilde{g}(y, \theta_0, \gamma_0)'(\tilde{G}^{-1})']
\]

where

\[
\tilde{G} = E \left[ \frac{\partial \tilde{g}(y, \gamma_0, \theta_0)}{\partial \theta' \partial \gamma'} \right] = \begin{bmatrix} G_\theta & G_\gamma \\ 0 & \bar{M} \end{bmatrix}
\]

\[
\tilde{G}^{-1} = \begin{bmatrix} G_\theta^{-1} & -G_\gamma^{-1} G_\gamma \bar{M}^{-1} \\ 0 & \bar{M}^{-1} \end{bmatrix}.
\]

Observe that the first row of \(\tilde{G}\) is

\[
G_\theta^{-1}[I, -G_\gamma \bar{M}^{-1}] \text{ now}
\]

\[
E[\tilde{g}(y, \theta_0, \gamma_0)\tilde{g}(y, \theta_0, \gamma_0)'] = [I, -G_\gamma \bar{M}^{-1}][\tilde{g}]
\]

\[
= g(y) + G_\gamma \psi(y)
\]

from partitional increase.