GENERALIZED METHOD OF MOMENTS

Whitney K. Newey

MIT

October, 2001

Motivation: IV setup from Hausman (1984).

Definitions:

y : $n \times 1$ vector of observations on dependent variable; y_i is i^{th} element;

X : $n \times p$ matrix of observations on right-hand side variables; X_i' is i^{th} row;

Z : $n \times m$ matrix of observations on instruments; Z_i' is i^{th} row.

Model (or moment restriction):

$$E[Z_i \varepsilon_i] = 0, \varepsilon_i = y_i - X_i'\beta_0.$$

Two-Stage Least Squares:

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

Sargan (1958) Interptetation:

$$\hat{\beta} = \arg\min_{\beta}(y - X\beta)'Z(Z'Z)^{-1}Z'(y - X\beta).$$

GMM Interptetation:

$$\begin{aligned} g_i(\beta) &= Z_i(y_i - X_i'\beta) : \text{Vector of moment functions.} \\ \hat{g}(\beta) &= \sum_{i=1}^{n} g_i(\beta)/n : \text{ Sample moments.} \\ \hat{A} &= (Z'Z/n)^{-1} : \text{ Weighting matrix (or distance matrix).} \end{aligned}$$

2SLS sets sample moments to be close to their population counterpart of zero:

$$\hat{\beta} = \arg\min_{\beta} \hat{g}(\beta)'\hat{A}\hat{g}(\beta).$$

Choice of $\hat{A}$: $\hat{A} = (Z'Z/n)^{-1}$ optimal (asymptotic variance minimizing) under homoskedasticity and no autocorrelation. Something else better under heteroskedasticity and/or autocorrelation.

General Assumptions: $\hat{\beta} \xrightarrow{p} \beta_0$, $\hat{g}(\beta)$ differentiable near β_0, and

$$\begin{aligned}
\text{Moment convergence} &: \quad \sqrt{n}\hat{g}(\beta_0) \xrightarrow{d} N(0,\Omega); \Omega \text{ nonsingular;} \\
\text{Jacobian convergence} &: \quad \partial\hat{g}(\bar{\beta})/\partial\beta \xrightarrow{p} G, \text{ for any } \bar{\beta} \xrightarrow{p} \beta_0; \\
\text{Weighting matrix} &: \quad \hat{A} \xrightarrow{p} A; \\
\text{Local identification} &: \quad G'AG \text{ nonsingular.}
\end{aligned}$$

Can check for IV: With independent observations (no autocorrelation) and $E[\varepsilon_i^2|Z_i] = \sigma^2$ (homoskedasticity), by central limit theorem and law of large numbers.

$$\begin{aligned}
\sqrt{n}\hat{g}(\beta_0) &= Z'\varepsilon/\sqrt{n} \xrightarrow{d} N(0, E[\varepsilon_i Z_i Z_i']) = N(0, \sigma^2 Q); Q = E[Z_i Z_i']. \\
\partial\hat{g}(\bar{\beta})/\partial\beta &= -Z'X/n \xrightarrow{p} -E[Z_i X_i'] \stackrel{def}{=} G, \\
& \quad \hat{A} = (Z'Z/n)^{-1} \xrightarrow{p} Q^{-1} \text{ (for } Q \text{ nonsingular),} \\
& \quad G'AG \text{ nonsingular (if } rank(G) \text{ is } p).
\end{aligned}$$

Asymptotic Distribution: Define $\hat{G} = \partial\hat{g}(\hat{\beta})/\partial\beta$;

$$\begin{aligned}
\text{First-order conditions} &: \quad 0 = \hat{G}'\hat{A}\hat{g}(\hat{\beta}). \\
\text{Expansion of } \hat{g}(\beta) &: \quad 0 = \hat{G}'\hat{A}\{\hat{g}(\beta_0) + \partial\hat{g}(\bar{\beta})/\partial\beta[\hat{\beta} - \beta_0]\}. \\
\text{Solve} &: \quad \sqrt{n}(\hat{\beta} - \beta_0) = -[\hat{G}'\hat{A}\partial\hat{g}(\bar{\beta})/\partial\beta]^{-1}\hat{G}'\hat{A}\sqrt{n}\hat{g}(\beta_0). \\
\text{Coefficient convergence} &: \quad -[\hat{G}'\hat{A}\partial\hat{g}(\bar{\beta})/\partial\beta]^{-1}\hat{G}'\hat{A} \xrightarrow{p} -(G'AG)^{-1}G'A. \\
\text{Moment convergence} &: \quad \sqrt{n}\hat{g}(\beta_0) \xrightarrow{d} N(0,\Omega). \\
\text{Slutzky theorem} &: \quad \sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} -(G'AG)^{-1}G'A \cdot N(0,\Omega) = N(0,V). \\
\text{Asymptotic Variance} &: \quad V = (G'AG)^{-1}G'A\Omega AG(G'AG)^{-1}.
\end{aligned}$$

Notes: 1) $\hat{A}$ affects V only through $plim(\hat{A})$; 2) If $m = p$, then $V = G^{-1}\Omega G^{-1\prime}$, and A drops out; 3) If $A = a \cdot \Omega^{-1}$ for a scalar a then $V = (G'\Omega^{-1}G)^{-1}$.

Asymptotically Efficient Distance Matrix: $A = plim(\hat{A}) = a \cdot \Omega^{-1}$.

$$
\begin{aligned}
\text{Definitions} &: \quad L'L = \Omega,\ H = (L')^{-1}G,\ F = (G'AG)^{-1}G'AL'. \\
\text{Intermediate results} &: \quad H'H = G'\Omega^{-1}G, FH = I. \\
\text{Proof of efficiency :} &\quad (G'AG)^{-1}G'A\Omega A'G(G'AG)^{-1} - (G'\Omega^{-1}G)^{-1} \\
&= FF' - (H'H)^{-1} = FF' - FH(H'H)^{-1}H'F' \\
&= F(I - H(H'H)^{-1}H')F'.
\end{aligned}
$$

Notes: 1) Actually implied by Gauss-Markov theorem for linear model $E[Y] = G\beta$, $Var(Y) = \Omega$; 2) Have equality if and only if F' is a linear combination of H, i.e. $LAG(G'AG)^{-1} = (L')^{-1}GC$, i.e. $AG = \Omega^{-1}GD$ for square matrices C, D.

Three IV Cases:

Homoskedasticity, no autocorrelation:

$$
\begin{aligned}
\Omega &= \sigma^2 Q, Q = E[Z_i Z_i']. \\
A &= \sigma^2\Omega^{-1} = Q^{-1} \text{ is optimal, as for 2SLS.}
\end{aligned}
$$

Heteroskedasticity, no autocorrelation: Hansen (1982), White (1982).

$$
\begin{aligned}
\Omega &= E[\varepsilon_i^2 Z_i Z_i']; \\
A &= \Omega^{-1} \text{ is optimal;} \\
\hat{\Omega} &= \sum_{i=1}^{n} \hat{\varepsilon}_i^2 Z_i Z_i/n \text{ is heteroskedasticty consistent for } \Omega; \\
\hat{A} &= \hat{\Omega}^{-1} \text{ is optimal;} \\
\hat{\beta} &= (X'Z\hat{\Omega}^{-1}Z'X)^{-1}X'Z\hat{\Omega}^{-1}Z'y.
\end{aligned}
$$

Heteroskedasticity and/or autocorrelation: Hansen (1982).

$$
\begin{aligned}
\Omega &= E[\varepsilon_i^2 Z_i Z_i'] + \sum_{\ell=1}^{\infty} E[\varepsilon_i\varepsilon_{i-\ell}(Z_i Z_{i-\ell}' + Z_{i-\ell}Z_i')], \\
\hat{\Omega} &= \hat{\Lambda}_0 + \sum_{\ell=1}^{L}[1 - \ell/(L+1)](\hat{\Lambda}_\ell + \hat{\Lambda}_\ell'), \hat{\Lambda}_\ell = \sum_{i=\ell+1}^{n} \hat{\varepsilon}_i\hat{\varepsilon}_{i-\ell}Z_i Z_{i-\ell}'/n \text{ is HAC.} \\
\hat{\beta} &= (X'Z\hat{\Omega}^{-1}Z'X)^{-1}X'Z\hat{\Omega}^{-1}Z'y.
\end{aligned}
$$

Two-Step Efficient GMM: For consistent estimator $\hat{\Omega}$ of Ω,

$$\hat{\beta} = \arg\min_{\beta} \hat{g}(\beta)'\hat{\Omega}^{-1}\hat{g}(\beta).$$

For IV gives $\hat{\beta} = (X'Z\hat{\Omega}^{-1}Z'X)^{-1}X'Z\hat{\Omega}^{-1}Z'y$.

Asymptotic Variance Estimation: For $\hat{G} = \partial\hat{g}(\hat{\beta})/\partial\beta$,

$$\hat{V} = (\hat{G}'\hat{\Omega}^{-1}\hat{G})^{-1},$$

$\hat{V} \xrightarrow{p} V$ where $\sqrt{n}(\hat{\beta} - \beta_0) \xrightarrow{d} N(0, V)$. For IV, $\hat{V} = n^2(X'Z\hat{\Omega}^{-1}Z'X)^{-1}$.

Small Sample Properties: 1) Bias like 2SLS, grows with number of instruments; 2) Inference less accurate with $\hat{\Omega} \neq \hat{\sigma}^2 Z'Z/n$.

Overidentification Statistic:

$$T = n\hat{g}(\hat{\beta})'\hat{\Omega}^{-1}\hat{g}(\hat{\beta}) \xrightarrow{d} \chi^2(m-p).$$

For IV, $T = \hat{\varepsilon}'Z\hat{\Omega}^{-1}Z'\hat{\varepsilon}/n, \hat{\varepsilon} = y - X\hat{\beta}$. Two special cases:

Homoskedasticity, no autocorrelation: For $\hat{\sigma}^2 = \hat{\varepsilon}'\hat{\varepsilon}/n$, $\hat{\Omega} = \hat{\sigma}^2 Z'Z/n$,

$$T = \hat{\varepsilon}'Z(Z'Z)^{-1}Z'\hat{\varepsilon}/\hat{\sigma}^2 = n \cdot \hat{\varepsilon}'Z(Z'Z)^{-1}Z'\hat{\varepsilon}/\hat{\varepsilon}'\hat{\varepsilon} = nR^2 \text{ from regression of } \hat{\varepsilon} \text{ on } Z.$$

Heteroskedasticity, no autocorrelation: For $\hat{r} = [\hat{\varepsilon}_1 Z_1, ..., \hat{\varepsilon}_n Z_n]'$, $\hat{\Omega} = \hat{r}'\hat{r}/n$, $e = (1, ..., 1)'$,

$$e'\hat{r}(\hat{r}'\hat{r})^{-1}\hat{r}'e = n \cdot e'\hat{r}(\hat{r}'\hat{r})^{-1}\hat{r}'e/e'e = nR^2 \text{ from regression of } e \text{ on } \hat{r}.$$

Asymptotic Distribution:

$$\begin{aligned}
\text{Definitions} &: L'L = \Omega,\ H = (L')^{-1}G, D = I - H(H'H)^{-1}H', \hat{U} = (L')^{-1}\sqrt{n}\hat{g}(\beta_0); \\
\text{Expansions} &: \sqrt{n}\hat{g}(\hat{\beta}) = \sqrt{n}\hat{g} + [\partial\hat{g}(\bar{\beta})/\partial\beta]\sqrt{n}(\hat{\beta} - \beta_0) \\
&= L'D\hat{U} + o_p(1) = O_p(1), \hat{\Omega}^{-1} = \Omega^{-1} + o_p(1) = L^{-1}(L')^{-1} + o_p(1); \\
\text{Test statistic} &: T = \sqrt{n}\hat{g}(\hat{\beta})'L^{-1}(L')^{-1}\sqrt{n}\hat{g}(\hat{\beta}) + O_p(1)o_p(1)O_p(1) \\
&= \hat{U}'D'D\hat{U} + o_p(1) = \hat{U}'D\hat{U} + o_p(1) \\
\text{Cont map thm} &: T = \hat{U}'D\hat{U} + o_p(1) \xrightarrow{d} U'DU, U \sim N(0, I_m) \\
\text{Gaussian distn} &: T \xrightarrow{d} U'DU \sim \chi^2(tr(D)) = \chi^2(m-p).
\end{aligned}$$

Nonlinear models: Known function $\rho(y, X, \beta)$ with

$$E[Z_i\varepsilon_i] = 0,\ \varepsilon_i = \rho(y_i, X_i, \beta_0).$$

Moment Functions: $g_i(\beta) = Z_i\rho(y_i, X_i, \beta)$.

GMM Estimator: For consistent estimator $\hat{\Omega}$ of Ω,

$$\hat{\beta} = \arg\min_{\beta} \hat{g}(\beta)'\hat{\Omega}^{-1}\hat{g}(\beta).$$

Homoskedasticity: $\hat{\Omega} = \hat{\sigma}^2 Z'Z/n$, estimator is nonlinear 2SLS.

Heteroskedasticity: $\hat{\Omega} = \sum_{i=1}^{n} \hat{\varepsilon}_i^2 Z_i Z_i/n$, $\hat{\varepsilon}_i = \rho(y_i, X_i, \tilde{\beta})$.

Notes: 1) Intuition for consistency like linear case, i.e. choosing $\hat{\beta}$ so moments close to zero population value; 2) Argument for consistency given in 14.385; 3) Asymptotic normality follows from general case described above.

Intertemporal CAPM: c_i consumption at time i, R_i is asset return between i and $i+1$, α_0 is time discount factor, $u(c, \gamma_0)$ utility function, Z_i observations on variables available at time i;

$$\begin{aligned}
(y_i, X_i) &= (c_{i+1}, R_i),\ \beta = (\gamma, \alpha), \\
\rho(y_i, X_i, \beta) &= R_i \cdot \alpha \cdot u_c(c_{i+1}, \gamma)/u_c(c_i, \gamma) - 1, \\
E[\rho(y_i, X_i, \beta_0)|Z_i] &= 0; \text{First-order conditions for utility maximization,} \\
& \quad \text{No autocorrelation.}
\end{aligned}$$