Extremum Estimation (EE)

If $\theta = \arg\max_{q\in\Theta\subseteq\square^k} \mathrm{E}\left[h(z,q)\right] \quad \Rightarrow \quad \boxed{\hat\theta = \arg\max_{q\in\Theta} \frac{1}{n}\sum_{i=1}^n h(z_i,q)}$: extremum estimator. $\sqrt{n}(\hat\theta-\theta)\xrightarrow{d} N\left(0;\ Q_2^{-1}Q_{11}Q_2^{-1}\right)$, *where* $Q_2 = \mathrm{E}\left[\frac{\partial h(z,\theta)}{\partial q'\partial q}\right];\ Q_{11} = \mathrm{E}\left[\frac{\partial h(z,\theta)}{\partial q'}\frac{\partial h(z,\theta)}{\partial q}\right]$

Examples: NLLS ($h(x,y,b) = -\left(y-g(x,b)\right)^2$), WNLLS ($h(x,y,b) = -\left(\left(y-g(x,b)\right)/\sigma(x)\right)^2$), MLE ($h = \ln f(z|q)$), LADE ($h = -\left|y-g(x,b)\right|$)

Suitable conditions: Global ID (uniqueness of solution of maximization problem)/ Local ID ($\mathrm{rg}\,Q_2 = k$: full), $h\in C^2(\Theta)$, Θ – compact, $\theta\in\mathrm{int}\,\Theta$.

Maximum Likelihood Estimation (MLE)

If $z \sim f(z|\theta)$, where f is known function, then $\boxed{\hat\theta = \arg\max_{q\in\Theta}\frac{1}{n}\sum_{i=1}^n \log f(z_i|q)}$. $\sqrt{n}(\hat\theta-\theta)\xrightarrow{d} N(0;\ \mathfrak{I}^{-1})$.

Score function: $\underset{k\times1}{s}(z,q) = \frac{\partial\log(z|q)}{\partial q'}$; ZES rule: $\mathrm{E}\left[s(z,q)\right]=0$. Information matrix: $\mathfrak{I} = -\mathrm{E}\left[\frac{\partial s(z,\theta)}{\partial q}\right] = \mathrm{E}\left[s(z,\theta)s(z,\theta)'\right]$.

Suitable conditions: support of z does not depend on θ, dimensionality of θ is fixed, Θ is better to be compact.
MLE is asymptotically efficient in the class of all consistent and asymptotically normal extremum estimators.

Testing $H_0:\theta=\theta_0$. $W = n(\hat\theta-\theta_0)'\mathfrak{I}(\theta_0)(\hat\theta-\theta_0)\to\chi^2_k$, $LR = 2\left(\ell_n(\hat\theta)-\ell_n(\theta_0)\right)\overset{A}{=}W$, $LM = \frac{1}{n}\sum s(z_i,\theta_0)'\hat{\mathfrak{I}}(\theta_0)^{-1}\sum s(z_i,\theta_0)\to\chi^2_k$.

Testing $H_0: g(\theta)=0$, $\underset{r\times k}{G} = \frac{\partial g(\theta)}{\partial q}$. $\left\{W = ng(\hat\theta)'\left[\hat G\hat{\mathfrak{I}}(\hat\theta)^{-1}\hat G'\right]g(\hat\theta),\ LR = 2\left(\ell_n(\hat\theta)-\ell_n(\hat\theta^R)\right),\ LM = \frac{1}{n}\sum s(z_i,\hat\theta^R)'\hat{\mathfrak{I}}(\hat\theta^R)^{-1}\sum s(z_i,\hat\theta^R)\right\}\to\chi^2_r$

Conditional MLE: if $f(z|q) = f(y|x,q_1,q_2)\,f(x|q_2,q_3)$, then $f(y|x,q_1,q_2)$ – conditional MLE, $f(x|q_2,q_3)$ – marginal MLE, $f(z|q)$ – joint MLE.

MLE for time series: $\ell_n^A(q) = \sum_{t=1+p}^T \ln f(z_t|z_{t-1},z_{t-2},\dots,z_{t-p},q)$ – approximate likelihood $\Rightarrow$ approximate MLE $\hat\theta$.

Generalized Method of Moments (GMM)

Moment condition: $\mathrm{E}\left[\underset{l\times1}{m}(z,\theta)\right]=0$, $\theta\in\Theta\subseteq\square^k$, $z\in\square\subseteq\square^p$, $m:\square\times\Theta\to\square^l$. $\quad \underset{l\times k}{Q_{\partial m}} = \mathrm{E}\left[\frac{\partial m(z,\theta)}{\partial q}\right],\ \underset{l\times l}{Q_{mm}} = \mathrm{E}\left[m(z,\theta)m(z,\theta)'\right]$

If $l=k$ (just identification), then CMM estimator: $\boxed{\hat\theta = \arg_{q\in\Theta}\left\{\frac{1}{n}\sum_{i=1}^n m(z_i,q)=0\right\}}$. $\sqrt{n}(\hat\theta-\theta)\xrightarrow{d} N\left(0;\ Q_{\partial m}^{-1}Q_{mm}Q_{\partial m}^{-1}\right)$

If $l>k$ (over identification), $\boxed{\hat\theta^{GMM} = \arg\min_{q\in\Theta}\left(\frac{1}{n}\sum m(z_i,q)\right)'W_n\left(\frac{1}{n}\sum m(z_i,q)\right)}$. $\sqrt{n}(\hat\theta-\theta)\xrightarrow{d} N\left(0;\ (Q_{\partial m}'WQ_{\partial m})^{-1}Q_{\partial m}'WQ_{mm}WQ_{\partial m}(Q_{\partial m}'WQ_{\partial m})^{-1}\right)$

Optimal weighting matrix: $W = Q_{mm}^{-1}$. Feasible GMM: 1) find consistent θ_0 (for example, $\hat\theta^{GMM}$ with $W=I$), 2) find $\hat\theta^{FGMM}$ with $W = \hat Q_{mm}^{-1}$.

Examples: OLS ($m = ex$), GLS ($m = e\frac{x}{\sigma^2(x)}$), NLLS ($m = eg_\beta(x,\beta)$), WNLLS ($m = e\frac{g_\beta(x,\beta)}{\sigma^2(x)}$), MLE ($m = \frac{\partial\log f(y|x,\theta)}{\partial q}$), EE ($m = \frac{\partial h(z,\theta)}{\partial q}$).

Suitable conditions: existence of $Q_{\partial m}, Q_{mm}$, global ID ($\mathrm{E}\left[m(z,q)\right]=0 \Leftrightarrow q=\theta$)/ local ID ($\mathrm{rg}\,Q_{\partial m} = k$ – full rank), Θ is compact, $\theta\in\mathrm{int}\,\Theta$.

Optimal moments: $\tilde m(z,q) = \hat Q_{\partial m}'\hat Q_{mm}^{-1}m(z_i,q) \quad\Rightarrow$ GMM with moment conditions m is equivalent to CMM with moments $\tilde m$.

GMM for time series: $Q_{mm} = LRV\left(m(z,\theta)\right)$, LRV estimator: $\hat V_z = \sum_{j=-m}^m \omega\left(\frac{j}{m+1}\right)\frac{1}{T}\sum_{t=\max(1,1+j)}^{\min(T,T+j)}(z_t-\bar z_T)(z_{t-j}-\bar z_T)'$, $\omega(x)^{N-W} = \left[1-|x|\right]_0$. $\omega^{Andr}(x) = \begin{cases}1-6x^2+6|x|^3, & |x|\le\frac12,\\ 2(1-|x|)^3, & \frac12\le|x|\le1,\\ 0, & 1\le|x|\end{cases}$

Test for over identifying restrictions (J-test), $H_0:\mathrm{E}\left[m(z,\theta)\right]=0,\ H_1:\forall q\ \mathrm{E}\left[m(z,q)\right]\ne0$: $J = n\left(\frac{1}{n}\sum m(z_i,\hat\theta)\right)'\hat Q_{mm}^{-1}\left(\frac{1}{n}\sum m(z_i,\hat\theta)\right)\xrightarrow{d}\chi^2_{l-k}$.

Hausman's test: H_0 : model is correct. Take $\hat\theta_0$ - efficient under H_0, $\hat\theta_1$ – robust. $H = n(\hat\theta_1-\hat\theta_0)'(\hat V_1-\hat V_0)^-(\hat\theta_1-\hat\theta_0)\to\chi^2_{\mathrm{rg}(V_1-V_0)}$, where $VV^-V=V$.

Testing $H_0: h(\theta)=0\ (\theta\in\Theta_0)$. $\left\{W = nh(\hat\theta)'\left[\hat H\left(\hat Q_{\partial m}'\hat Q_{mm}^{-1}\hat Q_{\partial m}'\right)^{-1}\hat H\right]^{-1}h(\hat\theta),\ DD = n\left[Q_n(\hat\theta^R)-Q_n(\hat\theta)\right],\ LM = \frac{n}{4}\lambda(\hat\theta^R)'\left(\hat Q_{\partial m}'\hat Q_{mm}^{-1}\hat Q_{\partial m}'\right)^{-1}\lambda(\hat\theta^R)\right\}\to\chi^2_q$

$$\underset{q\times k}{H} = \frac{\partial h(\theta)}{\partial q},\ Q_n(q) = \left(\frac{1}{n}\sum m(z_i,q)\right)'\hat Q_{mm}^{-1}\left(\frac{1}{n}\sum m(z_i,q)\right),\ \lambda(\theta) = 2\left(\frac{1}{n}\sum\frac{\partial m(z_i,\theta)}{\partial q}\right)'\hat Q_{mm}^{-1}\left(\frac{1}{n}\sum m(z_i,\theta)\right),\ \hat\theta = \arg\min_{q\in\Theta}Q_n(q),\ \hat\theta^R = \arg\min_{q\in\Theta_0}Q_n(q)$$

Instrumental Variables Estimation (IV)

$y = x'\beta+e,\ \mathrm{E}[ex]\ne0$. Instrumental variables: $\underset{l\times1}{z}$, $\mathrm{E}[ez]=0$ (validity), $Q_{zx} = \mathrm{E}[zx']\ne0$ (relevance). Moment condition: $\mathrm{E}\left[(y-x'\beta)z\right]=0$.

$l=k \Rightarrow$ IV estimator: $\boxed{\hat\beta^{IV} = \left(\frac{1}{n}\sum z_ix_i'\right)^{-1}\frac{1}{n}\sum z_iy_i} = (Z'X)^{-1}Z'Y$. $\sqrt{n}(\hat\beta^{IV}-\beta)\xrightarrow{d} N\left(0;\ Q_{zx}^{-1}Q_{zze^2}Q_{zx}'^{-1}\right)$, where $Q_{zze^2} = \mathrm{E}\left[zz'(y-x'\beta)^2\right]$.

$l>k \Rightarrow$ 2SLS estimator: $\boxed{\hat\beta^{2SLS} = \left(\frac{1}{n}\sum x_iz_i'\left(\frac{1}{n}\sum z_iz_i'\right)^{-1}\frac{1}{n}\sum z_ix_i'\right)^{-1}\frac{1}{n}\sum x_iz_i'\left(\frac{1}{n}\sum z_iz_i'\right)^{-1}\frac{1}{n}\sum z_iy_i} = \left(X'Z(Z'Z)^{-1}Z'X\right)^{-1}X'Z(Z'Z)^{-1}Z'Y$

$\sqrt{n}(\hat\beta^{2SLS}-\beta)\xrightarrow{d} N\left(0;\ \left(Q_{xz}Q_{zz}^{-1}Q_{xz}'\right)^{-1}Q_{xz}Q_{zz}^{-1}Q_{zze^2}Q_{zz}^{-1}Q_{xz}'\left(Q_{xz}Q_{zz}^{-1}Q_{xz}'\right)^{-1}\right)$

GMM estimator: $\boxed{\hat\beta^{GMM} = \left(\frac{1}{n}\sum x_iz_i'\left(\frac{1}{n}\sum z_iz_i'\hat e_i^2\right)^{-1}\frac{1}{n}\sum z_ix_i'\right)^{-1}\frac{1}{n}\sum x_iz_i'\left(\frac{1}{n}\sum z_iz_i'\hat e_i^2\right)^{-1}\frac{1}{n}\sum z_iy_i}$, $\sqrt{n}(\hat\beta^{GMM}-\beta)\xrightarrow{d} N\left(0;\ \left(Q_{xz}Q_{zze^2}Q_{xz}'\right)^{-1}\right)$

Optimal instrument: $\zeta = Q_{xz}Q_{zze^2}^{-1}z$, $\hat\zeta_i = \sum x_jz_j'\left(\sum z_jz_j'\hat e_j^2\right)^{-1}z_i$. Then $\hat\beta^{GMM} = \left(\sum\hat\zeta_ix_i'\right)^{-1}\sum\hat\zeta_iy_i = \hat\beta^{IV(\zeta)}$

Panel Data Analysis

Projectors: $P = X(X'X)^{-1}X'$, $Q = I - P$. Cronecker's product: $(A \otimes B)' = A' \otimes B'$, $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$, $(A \otimes B)(C \otimes D) = (AC \otimes BD)$.

Theorem (Frisch-Waugh-Lovell): $\left(y = x_1'\beta_1 + x_2'\beta_2 + e \to \hat{\beta}_2, \hat{e}\right) \Leftrightarrow \left(y = \gamma_1 x_1 + u \to \hat{u},\ x_2 = \gamma_2 x_1 + v \to \hat{v},\ \hat{u} = \hat{v}'\beta_2 + e \to \hat{\beta}_2, \hat{e}\right)$.

Panel data structure: $\{y_{it}, x_{it}\}$, $i = 1...n, t = 1,...,T$. Asymptotic: $n \to \infty$ but T stays fixed. Model: $y_{it} = \alpha + x_{it}'\beta + u_{it}$, $\beta \in \mathbb{R}^k$.

One-way ECM: $u_{it} = \mu_i + v_{it}$, Two-way ECM: $u_{it} = \mu_i + \lambda_t + v_{it}$. μ_i – individual effects, λ_t – time effects, v_{it} – idiosyncratic effects.

One-way ECM fixed effects: $y_{it} = \mu_i + x_{it}'\beta + v_{it}$, $i = 1...n, t = 1...T$, $v_{it} \sim iid(0, \sigma_v^2)$, v_{it} are independent from x_{it}'s.

Vector form: $\underset{nT\times 1}{y} = \underset{nT\times n}{D}\underset{n\times 1}{\mu} + \underset{nT\times k}{X}\underset{k\times 1}{\beta} + \underset{nT\times 1}{v}$, where $D = I_n \otimes i_T$, $Q = I_{nT} - D'(D'D)^{-1}D = I_{nT} - \frac{1}{T}I_n \otimes J_T$. Then $\hat{\beta}^{LSDV} = (X'QX)^{-1}X'Qy$.

Within transformation: $\boxed{y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + v_{it} - \bar{v}_i} \to \hat{\beta}^W, \hat{\mu}_i = \bar{y}_i - \bar{x}_i\hat{\beta}^W$. $\text{Var}\left[\hat{\beta}^W \middle| X\right] = \sigma_v^2 (X'QX)^{-1}$, $\hat{\sigma}_v^2 = RSS/(nT - n - k)$.

Testing for individual effects: $H_0: \mu_1 = \mu_2 = ... = \mu_n$. F-test: $F = \dfrac{(RSS^R - RSS^U)/(n-1)}{RSS^U/(nT - n - k)} \overset{H_0, v_{it}\sim N}{\sim} F(n-1, nT - n - k)$.

Two-way ECM fixed effects: $y_{it} = \alpha + \mu_i + \lambda_t + x_{it}'\beta + v_{it}$, $i = 1...n, t = 1...T$, $v_{it} \sim iid(0, \sigma_v^2)$, $\mu_n = \lambda_T = 0$, v_{it} are independent from x_{it}'s.

Vector form: $\underset{nT\times 1}{y} = \underset{nT\times 1}{i_{nT}} \cdot \alpha + \underset{nT\times(n-1)}{D_\mu} \cdot \underset{(n-1)\times 1}{\mu} + \underset{nT\times(T-1)}{D_\lambda} \cdot \underset{(T-1)\times 1}{\lambda} + \underset{nT\times k}{X} \cdot \underset{k\times 1}{\beta} + \underset{nT\times 1}{v} = D\gamma + X\beta + v$, where $D_\mu = \begin{pmatrix} I_{n-1} \\ \underset{1\times n-1}{\mathbf{0}} \end{pmatrix} \otimes i_T$, $D_\lambda = i_n \otimes \begin{pmatrix} I_{T-1} \\ \underset{1\times T-1}{\mathbf{0}} \end{pmatrix}$

$\underset{nT\times(n+T-1)}{D} = \begin{pmatrix} i_{nT} & D_\mu & D_\lambda \end{pmatrix}$, $\gamma' = \begin{pmatrix} \alpha & \mu' & \lambda' \end{pmatrix}$. Then $Q = I_{nT} - \frac{1}{T}I_n \otimes J_T - \frac{1}{n}J_n \otimes I_T + \frac{1}{nT}J_{nT}$.

Within-transformation: $\boxed{y_{it} - \bar{y}_i - \bar{y}_t + \bar{y} = (x_{it} - \bar{x}_i - \bar{x}_t + \bar{x})'\beta + (v_{it} - \bar{v}_i - \bar{v}_t + \bar{v})} \to \hat{\beta}^W$. $\text{Var}\left[\hat{\beta}^W \middle| X\right] = \sigma_v^2(X'QX)^{-1}$, $\hat{\sigma}_v^2 = \frac{RSS}{nT - n - T - k + 1}$.

Test for individual effects: $H_0: \mu_1 = \mu_2 = ... = \mu_n$, $F = \dfrac{(RSS^R - RSS^U)/(n-1)}{RSS^U/(nT - n - T - k + 1)} \underset{v_{it}\sim N}{\overset{H_0}{\sim}} F(n-1, nT - n - T - k + 1)$.

Test for time effects: $H_0: \lambda_1 = \lambda_2 = ... = \lambda_T$, $F = \dfrac{(RSS^R - RSS^U)/(T-1)}{RSS^U/(nT - n - T - k + 1)} \underset{v_{it}\sim N}{\overset{H_0}{\sim}} F(T-1, nT - n - T - k + 1)$; $(T-1)F \xrightarrow[n\to\infty, T\ fixed]{d} \chi^2_{(T-1)}$.

One-way ECM random effects: $y_{it} = \mu_i + x_{it}'\beta + v_{it}$, $i = 1...n, t = 1...T$, $v_{it} \sim iid(0, \sigma_v^2)$, $\mu_i \sim iid(0, \sigma_\mu^2)$ – independent from x_{it}'s and mutually.

Spectral decomposition: $\Omega = \text{Var}\,U = \sigma_v^2 I_{nT} + \sigma_\mu^2 I_n \otimes J_T = \sigma_v^2 Q + (T\sigma_\mu^2 + \sigma_v^2)P$. $\Omega^{-1} = \frac{1}{\sigma_v^2}(Q + \theta P)$; $\Omega^{-1/2} = \frac{1}{\sigma_v}(Q + \sqrt{\theta}P)$; $\theta = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\mu^2}$.

OLS estimator ($\theta = 1$): $\boxed{y_{it} = x_{it}'\beta + u_{it}} \to \hat{\beta}^{OLS} = (X'X)^{-1}X'Y$. $\text{Var}\left(\hat{\beta}\middle| X\right) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$

Within transformation ($\theta = 0$): $\boxed{y_{it} - \bar{y}_i = (x_{it} - \bar{x}_i)'\beta + v_{it} - \bar{v}_i} \to \hat{\beta}^W = (X'QX)^{-1}X'QY$. $\text{Var}\left(\hat{\beta}^W\middle| X\right) = (X'QX)^{-1}X'Q\Omega QX(X'QX)^{-1}$

Between-transformation ($\theta = \infty$): $\boxed{\bar{y}_i = \bar{x}_i'\beta + \mu_i + \bar{v}_i} \to \hat{\beta}^B = (X'PX)^{-1}X'PY$. $\text{Var}\left(\hat{\beta}^B\middle| X\right) = (X'PX)^{-1}X'P\Omega PX(X'PX)^{-1}$.

FGLS-transformation ($\theta = (1 - \hat{\pi})^2$): $\boxed{y_{it} - \hat{\pi}\bar{y}_i = (x_{it} - \hat{\pi}\bar{x}_i)'\beta + (1 - \hat{\pi})\mu_i + (v_{it} - \hat{\pi}\bar{v}_i)} \to \hat{\beta}^{GLS}$. $\text{Var}\left(\hat{\beta}^{GLS}\middle| X\right) = (X'\Omega^{-1}X)^{-1}$.

Estimated $\hat{\theta}$: $\hat{\theta} = \dfrac{RSS_W}{RSS_B}\dfrac{n-k}{nT - n - k + 1} \xrightarrow{p} \theta$.

Hausman's specification test: $H_0: \text{E}[x_i u_{it}] = 0$, $H_a: \text{E}[x_i u_{it}] \neq 0$. Take $\theta_0 = \theta^{GLS}$ and $\theta_1 = \theta^{Within}$.

Dynamic Panel Regression

Model: $y_{it} = \delta y_{i,t-1} + x_{it}'\beta + u_{it}$, $t = 2,...,T$, $i = 1,...,n$. One-way ECM fixed effects (not random!): $u_{it} = \mu_i + v_{it}$, $v_{it} \sim iid(0, \sigma_v^2)$.

FD-transformation: $\boxed{y_{it} - y_{i,t-1} = \delta(y_{i,t-1} - y_{i,t-2}) + (x_{it}' - x_{i,t-1}')\beta + (v_{it} - v_{i,t-1})}$, $t = 3,...,T$.

Model without regular regressors: $\underset{n(T-2)\times 1}{\Delta Y} = \delta\Delta Y_{-1} + \Delta v$.

Instrument set: $W = \begin{pmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \end{pmatrix}$, $W_i = \begin{pmatrix} y_{i1} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & y_{i1} & y_{i2} & 0 & 0 & \cdots & 0 \\ 0 & 0 & y_{i1} & y_{i2} & y_{i3} & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{pmatrix}$, $G = \begin{pmatrix} 2 & -1 & 0 & \cdots \\ -1 & 2 & -1 & \cdots \\ 0 & -1 & 2 & \cdots \\ \ldots & \ldots & \ldots & \ldots \end{pmatrix}$

Estimator: $\hat{\delta}^{GMM} = \left(\Delta Y_{-1}'W\left(W'(I_n \otimes G)W\right)^{-1}W'\Delta Y_{-1}\right)^{-1}\Delta Y_{-1}'W\left(W'(I_n \otimes G)W\right)^{-1}W'\Delta Y$. $\text{AVar}\left(\hat{\delta}^{GMM}\right) = \hat{\sigma}_v^2\left(\Delta Y_{-1}'W\left(W'(I_n \otimes G)W\right)^{-1}W'\Delta Y_{-1}\right)^{-1}$.

Model with regular regressors: $\underset{n(T-2)\times 1}{\Delta Y} = \delta \underset{n(T-2)\times 1}{\Delta Y_{-1}} + \underset{n(T-2)\times k}{\Delta X}\underset{k\times 1}{\beta} + \underset{n(T-2)\times 1}{\Delta v}$.

Instruments: $W_i = \begin{pmatrix} y_{i1} & x_{i1}' & \cdots & x_{iT}' & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & y_{i1} & y_{i2} & x_{i1}' & \cdots & x_{iT}' & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \cdots & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} & x_{i1}' & \cdots & x_{iT}' \end{pmatrix}$, $W_i = \begin{pmatrix} y_{i1} & x_{i1}' & x_{i2}' & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & y_{i1} & y_{i2} & x_{i1}' & x_{i2}' & x_{i3}' & 0 & 0 & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & y_{i1} & y_{i2} & \cdots & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \cdots & y_{i1} & \cdots & y_{i,T-2} & x_{i1}' & \cdots & x_{i,T-1}' \end{pmatrix}$

x_i's are strictly exogenous — x_i's are weakly exogenous

$$\begin{pmatrix} \hat{\delta}^{GMM} \\ \hat{\beta}^{GMM} \end{pmatrix} = \left(\Delta X'\Delta Y_{-1}'W\left(W'(I_n \otimes G)W\right)^{-1}W'\Delta Y_{-1}\Delta X\right)^{-1}\Delta X'\Delta Y_{-1}'W\left(W'(I_n \otimes G)W\right)^{-1}W'\Delta Y$$

$$\text{AVar}\begin{pmatrix} \hat{\delta}^{GMM} \\ \hat{\beta}^{GMM} \end{pmatrix} = \hat{\sigma}_v^2\left(\Delta X'\Delta Y_{-1}'W\left(W'(I_n \otimes G)W\right)^{-1}W'\Delta Y_{-1}\Delta X\right)^{-1}$$