Panel GMM

$$y_{it} = x_{it}'\beta + u_{it}, \quad i = 1, \ldots, n, \; t = 1, \ldots, T$$

x_{it} contains endogenous variables and lagged dependent variables.
IV: $Z_i(T \times L)$

population moment:

$$E[Z_i' u_i] = 0$$

sample moment:

$$g_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} Z_i'(y_i - X_i\beta)$$

GMM estimator:

$$\hat{\beta}_{\text{PGMM}}(W) = (S_{xz}' W S_{xz})^{-1} S_{xz}' W s_{zy}$$

asymptotic normality of $\hat{\beta}_{\text{PGMM}}$:

$$\sqrt{n}\left(\hat{\beta}_{\text{PGMM}} - \beta\right) \xrightarrow{d} N(0, \, \text{Avar}(\hat{\beta}_{\text{PGMM}}))$$

$$\text{Avar}(\hat{\beta}_{\text{PGMM}}) = \left(\Sigma_{xz}' W \Sigma_{xz}\right)^{-1} \Sigma_{xz}' W S W \Sigma_{xz} \left(\Sigma_{xz}' W \Sigma_{xz}\right)^{-1}$$

$$\widehat{\text{Avar}}(\hat{\beta}_{\text{PGMM}}) = (S_{xz}' W S_{xz})^{-1} S_{xz}' \hat{W} \hat{S} \hat{W} S_{xz} (S_{xz}' W S_{xz})^{-1}$$

Panel-robust standard errors allowing for both heteroskedasticity and auto-correlation.

Set $W = S_{zz}^{-1}$, **One-Step GMM** or panel-2SLS estimator:

$$\hat{\beta}_{\text{2SLS}} = \left[X'Z(Z'Z)^{-1}Z'X\right]^{-1} X'Z(Z'Z)^{-1}Z'y = (\hat{X}'\hat{X})^{-1}\hat{X}'y$$

$$\widehat{\text{Avar}}(\hat{\beta}_{\text{2SLS}}) = \hat{\sigma}^2 (\hat{X}'\hat{X})^{-1}, \quad \hat{\sigma}^2 = \frac{1}{nT}\sum_{i,t} \hat{u}_{it}^2$$

Set $W = \hat{S}^{-1}, \;\; \hat{S} \to S$, **Two-step GMM** estimator (efficient):

$$\hat{\beta}_{\text{2SGMM}} = \left[X'Z\hat{S}^{-1}Z'X\right]^{-1} X'Z\hat{S}^{-1}Z'y$$

$$\widehat{\text{Avar}}(\hat{\beta}_{\text{2SGMM}}) = \left[X'Z\hat{S}^{-1}Z'X\right]^{-1}$$

Tests of Over-identifying Restrictions (OIR, Hansen's J):

$$OIR = J = n \cdot g(\hat{\beta}_{\text{2SGMM}})'\hat{S}^{-1} g(\hat{\beta}_{\text{2SGMM}}) \xrightarrow{d} \chi^2(L-K)$$

IV Selection

Exogeneity Assumption	Moment Condition	Instrument Vector[a]
Summation	$\mathrm{E}\left[\sum_t \mathbf{z}_{it}u_{it}\right] = \mathbf{0}$	$[\mathbf{z}_{it}]$
Contemporaneous	$\mathrm{E}[\mathbf{z}_{it}u_{it}] = \mathbf{0}$, all t	$[\mathbf{0}'_{r_1} \cdots \mathbf{0}'_{r_{t-1}}\ \mathbf{z}'_{it}\ \mathbf{0}'_{r_{t+1}} \cdots \mathbf{0}'_{r_T}]$
Weak	$\mathrm{E}[\mathbf{z}_{is}u_{it}] = \mathbf{0}$, $s \le t$, all t	$[\mathbf{0}'_{r_1} \cdots \mathbf{0}'_{r_{t-1}}\ (\mathbf{z}^t_{it})'\ \mathbf{0}'_{r_{t+1}} \cdots \mathbf{0}'_{r_T}]$
Strong	$\mathrm{E}[\mathbf{z}_{is}u_{it}] = \mathbf{0}$, all s and t	$[\mathbf{0}'_{r_1} \cdots \mathbf{0}'_{r_{t-1}}\ (\mathbf{z}^T_{it})'\ \mathbf{0}'_{r_{t+1}} \cdots \mathbf{0}'_{r_T}]$

[a] The instrument vector is the tth row of $\mathbf{Z}_i$ in (22.11); $(\mathbf{z}^t_{it})' = [\mathbf{z}'_{i1} \dots \mathbf{z}'_{it}]$, $(\mathbf{z}^T_{it})' = [\mathbf{z}'_{i1} \dots \mathbf{z}'_{iT}]$; and $r_s = \dim[\mathbf{z}'_{is}]$ or $\dim[\mathbf{z}^s_{is}]$ or $\dim[\mathbf{z}^T_{is}]$.

Summation Assumption

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}'_{i1} \\ \mathbf{z}'_{i2} \\ \vdots \\ \mathbf{z}'_{i\mathrm{T}} \end{bmatrix}, \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix}$$

$$\mathrm{E}\left[\sum_{t=1}^{T} \mathbf{z}_{it}u_{it}\right] = \mathbf{0}$$

The sum of the expectations between the instruments and the error terms over the entire sample period is zero.

Contemporaneous Exogeneity Assumption(Stronger)

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}'_{i1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{z}'_{i2} & \cdots & \vdots \\ \vdots & \vdots & \ddots & \mathbf{0} \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{z}'_{iT} \end{bmatrix}, \quad \mathbf{u}_i = \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix}$$

$$E[\mathbf{z}_{it}u_{it}] = \mathbf{0}, \quad t = 1, \cdots, T$$

The instrument in each period is uncorrelated with the error term in the same period. If x includes time dummies then there are only $TK - (T-1)$ moment conditions available.

- How to construct IV: For period t, we use only the current-period instrument and set instruments from other periods to zero.

Weak Exogeneity Assumption

$$E[z_{is}u_{it}] = 0, \quad s \le t$$

The current and lagged values of the instrument are uncorrelated with the current error term.

- How to construct IV: For period t, we use all instruments up to and including period t.

Strong Exogeneity Assumption

$$E[z_{is}u_{it}] = 0, \quad s, t = 1, \ldots, T$$

All values of the instrument (past, present, and future) are uncorrelated with the error term in any period.

- How to construct IV: For any period t, we can use instruments from all T periods.

Redundant Instruments

Time-invariant instruments can be used only once.
Instruments that are the product of time dummies and a time-invariant regressor should be included only once.

Chamberlain's Optimal Distance Estimator

$$y_{it} = \alpha_i + x_{it}'\beta + u_{it}$$

Assume $E[x_{is}u_{it}] = 0$, for all s, t.

> To eliminate α_i, we typically apply **within transformation** or **first-differencing**, obtaining a transformed model ($\tilde{y}_{it} = \tilde{x}_{it}'\beta + \tilde{u}_{it}$).
>
> However, this transformation **loses information**, and if x_{it} is strongly exogenous, then pre-transformation values of $x_{is}(s \neq t)$ from other periods can also serve as valid instruments.
>
> Chamberlain's method aims to **more fully utilize these cross-period moment conditions**, thereby obtaining a more efficient estimator than the standard fixed effects estimator.

Chamberlain's model setting:

$$y_i = e\alpha_i + (I_T \otimes \beta')x_i + u_i$$

$e = [1, 1, \ldots, 1]'$

Chamberlain assumes the fixed effect α_i can be linearly predicted by the explanatory variables x_i (values from all periods):

$$E^*[\alpha_i | x_i] = \mu + \sum_{t=1}^{T} \lambda_t' x_{it} = \mu + \lambda' x_i$$

Assume $E^*[u_i | \alpha_i, x_i] = 0$

$$
\begin{aligned}
E^*[y_i|x_i] &= eE^*[\alpha_i|x_i] + (I_T \otimes \beta')x_i \\
&= e\mu + (I_T \otimes \beta' + e\lambda')x_i \\
&= \Pi_0 + \Pi_1 x_i
\end{aligned}
$$

Step 1: Estimate the Reduced-Form Parameters

- run a multivariate OLS regression to get the reduced-form parameter $\hat{\Pi}$.
- estimate the variance-covariance matrix of $\hat{\Pi}$: $\hat{V}[\mathrm{Vec}(\hat{\Pi})]$ (Vec for Vectorization).

Step 2: Minimum Distance Estimation

$$
J(\beta, \lambda) = \left(\mathrm{Vec}(\hat{\Pi} - I_T \otimes \beta' - e\lambda')\right)' W \left(\mathrm{Vec}(\hat{\Pi} - I_T \otimes \beta' - e\lambda')\right)
$$

no homoskedasticity assumption is needed.

Example

Table 21.2. *Hours and Wages: Standard Linear Panel Model Estimators*[a]

	POLS	Between	Within	First Diff	RE–GLS	RE–MLE
α	7.442	7.483	7.220	.001	7.346	7.346
β	.083	.067	.168	.109	.119	.120
Robust se[b]	(.030)	(.024)	(.085)	(.084)	(.051)	(.052)
Boot se	[.030]	[.019]	[.084]	[.083]	[.056]	[.058]
Default se	{.009}	{.020}	{.019}	{.021}	{.014}	{.014}
R^2	.015	.021	.016	.008	.014	.014
RMSE	.283	.177	.233	.296	.233	.233
RSS	427.225	0.363	259.398	417.944	288.860	288.612
TSS	433.831	17.015	263.677	420.223	293.023	292.773
σ_α	.000		.181		.161	.162
σ_ε	.283		.232		.233	.233
λ	0.000	–	1.000	–	.585	.586
N	5320	532	5320	4788	5320	5320

[a] Shown are pooled OLS (POLS), between, within, first-differences, random effects (RE) GLS and MLE linear panel regression of lnhrs on lnwg. Standard errors for the slope coefficients are panel robust in parentheses, panel bootstrap in square brackets, and default estimates that assume iid errors in curly braces. The R^2, root mean square error (RMSE), residual sum of squares (RSS), total sum of squares (TSS), and sample size come from the appropriate regression given in Section 21.2. The parameter λ is defined after (21.11).

[b] se, standard error.

		Base Case		Stacked	
	OLS	2SLS	2SGMM	2SLS	2SGMM
β_1	0.112	0.209	0.547	0.543	0.330
Panel se	(.096)	(.374)	(.327)	(.209)	(.110)
Het se	[.079]	[.423]	[–]	[.226]	[–]
Default se	{.023}	{.389}	{ – }	{.169}	{ – }
RMSE	.283	.296	.307	.307	.298
Instruments	5	9	9	72	72
OIR Test	–	–	5.45	–	69.51
dof	–	–	4	–	67
p-value	–	–	.244	–	.393
N	4256	4256	4256	4256	4256

[a] Differenced regression uses annual data from 1981–1988 for 532 men. Reported are β_1, the coefficient of Δ lnwg, and three estimated standard errors: panel robust in parentheses, heteroskedastic robust in square brackets, and usual default estimates that assume iid errors in curly braces. All regressions additionally include Δkids, Δage, Δagesq, and Δdisab as regressors but their coefficient estimates are not reported. The instruments are lnwg lagged twice and kids, age, agesq, and disab lagged both once and twice. For the base case there are 9 instruments and for stacked instruments there are $8 \times 9 = 72$ instruments. RMSE is the root mean square error of the residual. OIR is the over identifying restictions test statistic, dof is the degrees of freedom, and p-value is the p-value for this test.

$$\Delta \ln \mathrm{hrs}_{it} = \beta_1 \Delta \ln \mathrm{wg}_{it} + \beta_2 \Delta \mathrm{kids}_{it} + \beta_3 \Delta \mathrm{age}_{it} + \beta_4 \Delta \mathrm{agesq}_{it} + \beta_5 \Delta \mathrm{disab}_{it} + \Delta u_{it}$$

2SLS with Base-Case Instruments: 9 IV
$lnwg_{i,t-2}, kids_{i,t-1}, age_{i,t-1}, agesq_{i,t-1}, disab_{i,t-1}, kids_{i,t-2}, age_{i,t-2}, agesq_{i,t-2}, disab_{i,t-2}$

2SLS with Stacked Instruments: 72(=8×9) IV

The **two-step GMM estimator** is more efficient than 2SLS.

Test of Overidentifying Restrictions: see OIR Test and its p-value
H_0: All overidentifying moment conditions are valid. In other words: all instruments are exogenous.

Test of Weak Instruments:
H_0: The correlation between the instruments and the endogenous explanatory variable is weak./The coefficients of the instruments in the first-stage regression are jointly zero (or close to zero).