Panel Econometrics

Paulo M. M. Rodrigues

February 2022

Today's outline

Dynamic models

Outline

Dynamics

Many economic relations are dynamic in nature:

- E.g. state dependence: today's state y_{it} depends on yesterday's $y_{i,t-1}$.
- Then, one of the advantages of panel data is that they allow the researcher to better understand the dynamics of adjustment.

Take a model with adjustment costs. Suppose the optimal quantity y_{it}^* is

$$y_{it}^* = \mathbf{x}_{it}\boldsymbol{\gamma} + \tilde{v}_{it},$$

but due to adjustment costs (governed by α), the realization is

$$y_{it} = \alpha y_{i,t-1} + (1-\alpha) y_{it}^*.$$

Putting the two equations together yields the estimable model:

$$y_{it} = \alpha y_{i,t-1} + \mathbf{x}_{it} \underbrace{(1-\alpha)\boldsymbol{\gamma}}_{\boldsymbol{\beta}} + \underbrace{(1-\alpha)\tilde{v}_{it}}_{v_{it}}.$$

The dynamic panel model

Suppose the structural model is

$$y_{it} = \alpha y_{i,t-1} + \mathbf{x}_{it}\boldsymbol{\beta} + v_{it} \quad i = 1, \ldots, N, \quad t = 2, \ldots, T$$

where α is a scalar, $\mathbf{x}_{it}$ is $1 \times K$ and $\boldsymbol{\beta}$ is $K \times 1$.

We assume there is a one-way error component structure

$$v_{it} = c_i + u_{it}.$$

We will be precise about the assumptions regarding c_i and u_{it} later.

This dynamic panel data model has two sources of persistence over time:

- *state dependency* (y_{it} is a direct function of $y_{i,t-1}$) and
- *individual effects* (like in static models).

Inconsistency of the OLS estimator

- Since y_{it} is a function of c_i, it immediately follows that $y_{i,t-1}$ is also a function of c_i and the strict exogeneity assumption fails.
- Therefore, the right-hand side regressor $y_{i,t-1}$ is correlated with the error term v_{it} through c_i.
- Consequence: the OLS estimator is biased and inconsistent!
- Compare this to pure time series models: here the OLS estimator of the dynamic model is consistent if the disturbance is white noise.
- In time series models, (a) there is no individual effect and (b) T is assumed to be large (hence we use large T asymptotics).
- In panel models, (a) the individual effect introduces autocorrelation in the disturbance and (b) N is assumed to be large (hence we use large N asymptotics).

Inconsistency of the within estimator

The FE estimator wipes out the c_i by the within transformation:

$$\ddot{y}_{i,t-1} = y_{i,t-1} - \bar{y}_{i,-1} = y_{i,t-1} - \frac{1}{T-1}\sum_{t=2}^{T} y_{i,t-1}.$$

and

$$\ddot{u}_{it} = u_{it} - \bar{u}_i = u_{it} - \frac{1}{T-1}\sum_{t=2}^{T} u_{it}.$$

- Hence, the regressor $\ddot{y}_{i,t-1}$ is a function of $y_{i,1}, \ldots, y_{i,T-1}$ while the disturbance $\ddot{u}_{it}$ is a function of $u_{i,2}, \ldots, u_{i,T}$, so there is obvious correlation between the two ("regressor endogeneity").
- This correlation makes the FE estimator biased ("Nickell bias", see Nickell, 1981).
- This bias does not vanish as the number of individuals increases, so the FE estimator is inconsistent for N large and T small.
- Only as T gets large the FE estimator becomes consistent.

Inconsistency of the RE estimator

The RE estimator applies quasi-demeaning to the regressor

$$\tilde{y}_{i,t-1} = y_{i,t-1} - (1-\phi)\bar{y}_{i,-1}$$

and the disturbance

$$\tilde{u}_{it} = u_{it} - (1-\phi)\bar{u}_{i}.$$

- As before, the regressor $\tilde{y}_{i,t-1}$ is a function of $y_{i,1}, \ldots, y_{i,T-1}$ while the disturbance $\tilde{u}_{it}$ is a function of $u_{i,2}, \ldots, u_{i,T}$, so there is obvious correlation between the two.
- Hence, the RE estimator is biased and inconsistent in a dynamic panel data model as well.

Bias correction procedures

Several suggestions to correct for the bias of the popular FE estimator have been proposed in the literature:

- Kiviet (1995): derives an approximation for the bias of the FE estimator in a dynamic panel data model with serially uncorrelated disturbances and strictly exogenous regressors. A bias corrected FE estimator then subtracts a consistent estimator of this bias from the original FE estimator.
- Everaert and Pozzi (2007): bias correction for the FE estimator based on an iterative bootstrap procedure.
- Bun and Carree (2006): derive the asymptotic bias of the FE estimator for finite T and large N in the presence of both time-series and cross-section heteroskedasticity; again, bias correction procedures.

Instead of such correction procedures we will follow the bulk of the literature and use instrumental variables approaches to find consistent estimators (AH, AB, BB; others exist, see textbooks).

Outline

Panel AR(1) model and assumptions

Since $y_{i,t-1}$ is the problematic regressor, let us start with the AR(1) model:

$$y_{it} = \alpha y_{i,t-1} + v_{it}, \quad v_{it} = c_i + u_{it}, \qquad t = 2, \ldots, T.$$

Assumption AR.1 (sequential exogeneity/predeterminedness):

$$\mathrm{E}(u_{it} | y_{i,t-1}, \ldots, y_{i1}, c_i) = 0 \qquad \text{for all } t = 2, \ldots, T.$$

This implies dynamic completeness conditional on c_i:

$$\mathrm{E}(y_{it} | y_{i,t-1}, \ldots, y_{i1}, c_i) = \alpha y_{i,t-1} + c_i.$$

Discussion:

- Sequential exogeneity replaces strict exogeneity assumptions.
- Dynamic completeness means the dynamics of y_{it} is fully specified.
- A consequence is that u_{it} is white noise, $\mathrm{E}(u_{it} u_{is}) = 0 \ \forall \ s \neq t$.

Backward substitution

For further use, rewrite the AR(1) model by recursive backward substitution:

$$\begin{aligned} y_{it} &= \alpha y_{i,t-1} + v_{it} \\ &= \alpha(\alpha y_{i,t-2} + v_{i,t-1}) + v_{it} = \alpha^2 y_{i,t-2} + v_{it} + \alpha v_{i,t-1} \\ &\vdots \\ &= \alpha^{t-1} y_{i1} + v_{it} + \alpha v_{i,t-1} + \ldots + \alpha^{t-2} v_{i2} \end{aligned}$$

Taking first differences yields (since $\Delta v_{it} = \Delta u_{it}$)

$$\Delta y_{it} = \alpha^{t-2}\Delta y_{i2} + \Delta u_{it} + \alpha \Delta u_{i,t-1} + \ldots + \alpha^{t-3}\Delta u_{i3}.$$

Moment conditions

Apply the first difference (FD) transformation to get rid of c_i:

$$y_{it} - y_{i,t-1} = \alpha(y_{i,t-1} - y_{i,t-2}) + u_{it} - u_{i,t-1}.$$

How to consistently estimate this equation?

- The strict exogeneity assumption fails. We can also argue directly: From the previous slide we know that the regressor $\Delta y_{i,t-1}$ is a function of $u_{i,t-1}$ and hence is correlated with the error Δu_{it}.
- Hence, POLS is inconsistent.
- Anderson and Hsiao (1981) suggest to use an instrument for $\Delta y_{i,t-1}$.
- Use either $\Delta y_{i,t-2} = y_{i,t-2} - y_{i,t-3}$ or $y_{i,t-2}$ as instrument.
- These instruments will not be correlated with $\Delta u_{it} = u_{i,t} - u_{i,t-1}$ as long as the u_{it} is white noise.

Anderson-Hsiao details

Note that assumption AR.1 ensures validity of the Anderson and Hsiao moment conditions. Specifically, it implies $\mathrm{E}(y_{i,t-k}u_{it}) = 0$ for $k \geq 1$.

Hence, for all $t = 3, \ldots, T$,

$$\mathrm{E}(y_{i,t-2}\Delta u_{it}) = \mathrm{E}(y_{i,t-2}u_{it}) - \mathrm{E}(y_{i,t-2}u_{i,t-1}) = 0$$

and, for all $t = 4, \ldots, T$,

$$\mathrm{E}(\Delta y_{i,t-2}\Delta u_{it}) = \mathrm{E}(y_{i,t-2}\Delta u_{it}) - \mathrm{E}(y_{i,t-3}\Delta u_{it}) = 0.$$

Which instrument should we use?

- Using $y_{i,t-2}$ leaves us one observation more per individual.
- Arellano (1989) finds that the estimator using $\Delta y_{i,t-2}$ as instruments has a singularity point and very large variances over a significant range of parameter values.
- In contrast, the estimator that uses instruments in levels, i.e. $y_{i,t-2}$, has no singularities and much smaller variances.

Matrix form

FD of structural model

$$y_{it} - y_{i,t-1} = \alpha(y_{i,t-1} - y_{i,t-2}) + u_{it} - u_{i,t-1}.$$

Collect all observations for individual i in the $(T-2) \times 1$ vectors

$$\Delta \mathbf{y}_i = \begin{pmatrix} y_{i,3} - y_{i,2} \\ \vdots \\ y_{i,T} - y_{i,T-1} \end{pmatrix}, \ \Delta \mathbf{y}_{i,-1} = \begin{pmatrix} y_{i,2} - y_{i,1} \\ \vdots \\ y_{i,T-1} - y_{i,T-2} \end{pmatrix}, \ \Delta \mathbf{u}_i = \begin{pmatrix} u_{i,3} - u_{i,2} \\ \vdots \\ u_{i,T} - u_{i,T-1} \end{pmatrix}$$

The equation becomes

$$\Delta \mathbf{y}_i = \alpha \Delta \mathbf{y}_{i,-1} + \Delta \mathbf{u}_i, \qquad i = 1, \ldots, N.$$

Stacking the observations of all individuals yields

$$\Delta \mathbf{y} = \alpha \Delta \mathbf{y}_{-1} + \Delta \mathbf{u},$$

where $\Delta \mathbf{y} = (\Delta \mathbf{y}_1', \ldots, \Delta \mathbf{y}_N')'$, $\Delta \mathbf{y}_{-1} = (\Delta \mathbf{y}_{1,-1}', \ldots, \Delta \mathbf{y}_{N,-1}')'$ etc.

The Anderson and Hsiao instrumental variables estimator

Starting point: orthogonality (moment) conditions

$$\mathrm{E}[y_{i,t-2}\Delta u_{it}] = 0, \qquad t = 3, \ldots, T.$$

Define the $(T-2) \times 1$ vector of instruments

$$\mathbf{y}_{i,-2} = \begin{pmatrix} y_{i,1} \\ \vdots \\ y_{i,T-2} \end{pmatrix}$$

Write the orthogonality condition in stacked form:

$$\mathrm{E}[\mathbf{y}'_{i,-2}\Delta\mathbf{u}_i] = 0.$$

Sample moment conditions

Sample equivalent (normal equation):

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{y}'_{i,-2}\Delta\hat{\mathbf{u}}_i = \frac{1}{N}\sum_{i=1}^{N}\mathbf{y}'_{i,-2}(\Delta\mathbf{y}_i - \hat{\alpha}_{AH}\Delta\mathbf{y}_{i,-1}) \overset{!}{=} \mathbf{0}$$

Solving for $\hat{\alpha}_{AH}$ yields the estimator:

$$\hat{\alpha}_{AH} = \frac{\mathbf{y}'_{-2}\Delta\mathbf{y}}{\mathbf{y}'_{-2}\Delta\mathbf{y}_{-1}} = \frac{\sum_{i=1}^{N}\mathbf{y}'_{i,-2}\Delta\mathbf{y}_i}{\sum_{i=1}^{N}\mathbf{y}'_{i,-2}\Delta\mathbf{y}_{i,-1}} = \frac{\sum_{i=1}^{N}\sum_{t=3}^{T}y_{i,t-2}\Delta y_{i,t}}{\sum_{i=1}^{N}\sum_{t=3}^{T}y_{i,t-2}\Delta y_{i,t-1}}$$

(which is not surprising, though).

Consistency

To see that the estimator is consistent substitute the model

$$\hat{\alpha}_{AH} = \frac{\sum_{i=1}^{N} \mathbf{y}'_{i,-2}\Delta\mathbf{y}_i}{\sum_{i=1}^{N} \mathbf{y}'_{i,-2}\Delta\mathbf{y}_{i,-1}} = \alpha + \frac{\sum_{i=1}^{N} \mathbf{y}'_{i,-2}\Delta\mathbf{u}_i}{\sum_{i=1}^{N} \mathbf{y}'_{i,-2}\Delta\mathbf{y}_{i,-1}}.$$

Under standard conditions, as $N \to \infty$,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{y}'_{i,-2}\Delta\mathbf{y}_{i,-1} \xrightarrow{\text{p}} \mathbf{B} \equiv \mathrm{E}(\mathbf{y}'_{i,-2}\Delta\mathbf{y}_{i,-1}) \neq 0.$$

By assumption,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{y}'_{i,-2}\Delta\mathbf{u}_i \xrightarrow{\text{p}} \mathrm{E}[\mathbf{y}'_{i,-2}\Delta\mathbf{u}_i] = 0$$

(a CLT also typically holds). Taken together

$$\hat{\alpha}_{AH} \xrightarrow{\text{p}} \alpha + \frac{0}{\mathbf{B}} = \alpha.$$

Not all is fine

The instrumental variable (IV) estimation method leads to consistent but not necessarily efficient estimates of the parameters in the model because

(a) it does not make use of all the available moment conditions and

(b) it does not take into account the differenced structure on the disturbances (Δu_{it}) which are NOT white noise.

To see (a), note that the stacked moment condition $\mathrm{E}[\mathbf{y}'_{i,-2}\Delta\mathbf{u}_i] = 0$ means:

$$\mathrm{E}[y_{i,1}\Delta u_{i3}] + \ldots + \mathrm{E}[y_{i,T-2}\Delta u_{iT}] = 0$$

which is much weaker than what we originally assumed:

$$\mathrm{E}[y_{i,1}\Delta u_{i3}] = \ldots = \mathrm{E}[y_{i,T-2}\Delta u_{iT}] = 0.$$

Outline

Model

Start again from

$$y_{it} = \alpha y_{i,t-1} + v_{it}, \quad v_{it} = c_i + u_{it}, \qquad t = 2, \ldots, T.$$

Take first differences to eliminate the individual effect:

$$y_{it} - y_{i,t-1} = \alpha \left(y_{i,t-1} - y_{i,t-2} \right) + \left(u_{it} - u_{i,t-1} \right), \quad t = 3, \ldots, T$$

(not much new, so far).

Intuition

- Consider period $t = 3$, the first period we observe the relationship:

$$y_{i3} - y_{i2} = \alpha \left(y_{i2} - y_{i1} \right) + \left(u_{i3} - u_{i2} \right).$$

In this case, y_{i1} is a likely valid instrument: correlated with $(y_{i2} - y_{i1})$, not correlated with $(u_{i3} - u_{i2})$ if u_{it} is white noise.

- Consider period $t = 4$, the second period we observe the relation:

$$y_{i4} - y_{i3} = \alpha \left(y_{i3} - y_{i2} \right) + \left(u_{i4} - u_{i3} \right).$$

In this case, y_{i2} and y_{i1} are valid instruments for $(y_{i3} - y_{i2})$.

- Consider period $t = 5$, the third period we observe the relation:

$$y_{i5} - y_{i4} = \alpha \left(y_{i4} - y_{i3} \right) + \left(u_{i5} - u_{i4} \right).$$

In this case, y_{i1}, y_{i2} and y_{i3} are valid instruments for $(y_{i4} - y_{i3})$.

- And so on until period T, for which the set of valid instruments is $y_{i1}, \dots, y_{i,T-2}$.

Assumption and moment conditions

Assumption AR.1 (sequential exogeneity):

$$\mathrm{E}(u_{it}|y_{i,t-1},\ldots,y_{i1},c_i)=0 \qquad t=2,\ldots,T.$$

The assumption implies that

- u_{it} is uncorrelated with $y_{i,t-1},\ldots,y_{i1}$,
- $u_{i,t-1}$ is uncorrelated with $y_{i,t-2},\ldots,y_{i1}$, and thus
- Δu_{it} is uncorrelated with $y_{i,t-2},\ldots,y_{i1}$.

The Arellano-Bond estimator uses *all* these moment conditions:

$$\mathrm{E}[y_{i1}\Delta u_{it}]=\cdots=\mathrm{E}[y_{i,t-2}\Delta u_{it}]=0, \qquad t=3,\ldots,T.$$

The number of moment conditions

To count the moment conditions, write them out:

$$
\begin{array}{ll}
\mathrm{E}[y_{i1}\Delta u_{i3}] = 0, & \text{1 condition for } t = 3, \\
\mathrm{E}[y_{i1}\Delta u_{i4}] = \mathrm{E}[y_{i2}\Delta u_{i4}] = 0, & \text{2 conditions for } t = 4, \\
\vdots & \vdots \\
\mathrm{E}[y_{i1}\Delta u_{iT}] = \ldots = \mathrm{E}[y_{i,T-2}\Delta u_{iT}] = 0 & \text{T-2 conditions for } t = T.
\end{array}
$$

Altogether, the number of moment conditions grows quadratically with T:

$$L = 1 + 2 + \cdots + T - 2 = (T-2)(T-1)/2.$$

For $T = 3$ we have $L = 1$, for $T = 4$ we have $L = 3$, for $T = 10$ we have $L = 36$ moment conditions.

This implies that for $T \geq 4$, we have overidentifying moment conditions. This is why Arellano and Bond use GMM estimation.

Instrument matrix

The instrument matrix for individual i thus is

$$\mathbf{W}_i = \begin{bmatrix} [y_{i1}] & & & 0 \\ & [y_{i1}, y_{i2}] & & \\ & & \ddots & \\ 0 & & & [y_{i1}, \ldots, y_{i,T-2}] \end{bmatrix}$$

and the set of moment conditions is written (differently from PIV/P2SLS) as

$$\mathrm{E}(\mathbf{W}_i' \Delta \mathbf{u}_i) = 0.$$

Stacking all observations $i = 1, \ldots, N$, the matrix of instruments is

$$\mathbf{W} = \begin{pmatrix} \mathbf{W}_1 \\ \vdots \\ \mathbf{W}_N \end{pmatrix}.$$

GMM estimation

Since for $T \geq 4$, the conditions $\mathrm{E}(\mathbf{W}_i' \Delta \mathbf{u}_i) = 0$ are overidentifying, it is not possible to find an estimator of α that equates the sample equivalent

$$N^{-1} \sum_{i=1}^{N} \mathbf{W}_i' \Delta \hat{\mathbf{u}}_i$$

exactly to zero. We need a symmetric, positive definite $L \times L$ weighting matrix $\boldsymbol{\Xi}$ (ideally the GMM optimal weighting scheme).

An asymptotically optimal weighting matrix satisfies

$$\boldsymbol{\Xi} = \boldsymbol{\Lambda}^{-1}, \qquad \boldsymbol{\Lambda} = \mathrm{E}(\mathbf{W}_i' \Delta \mathbf{u}_i \Delta \mathbf{u}_i' \mathbf{W}_i).$$

While we can estimate this matrix in a second step based on first-step residuals $\Delta \hat{\mathbf{u}}_i$, we need a good first-step choice.

Arellano and Bond suggest to use one that is based on homoskedasticity assumptions.

First step weighting matrix

We make the conditional homoskedasticity assumption

$$\mathrm{E}(\Delta\mathbf{u}_i\Delta\mathbf{u}_i'|\mathbf{W}_i) = \mathrm{E}(\Delta\mathbf{u}_i\Delta\mathbf{u}_i') = \mathbf{\Sigma}_{\Delta\mathbf{u}}.$$

By the LIE, it implies that $\mathbf{\Xi}^{-1}$ simplifies to

$$\mathbf{\Xi}_1^{-1} = \mathbf{\Lambda}_1 = \mathrm{E}(\mathbf{W}_i'\Delta\mathbf{u}_i\Delta\mathbf{u}_i'\mathbf{W}_i) = \mathrm{E}(\mathbf{W}_i'\mathbf{\Sigma}_{\Delta\mathbf{u}}\mathbf{W}_i).$$

So we need to find the (unconditional) variance matrix $\mathbf{\Sigma}_{\Delta\mathbf{u}}$ of the differenced error term $\Delta\mathbf{u}_i$.

Since by assumption u_{it} is white noise, Δu_{it} is MA(1) with unit coefficient and zero mean.

In addition, let us assume that second moments are time-invariant (homoskedastic over time, stationary).

Variance matrix of $\Delta \mathbf{u}_i$ under homoskedasticity $\star$

Under the white noise and homoskedasticity assumptions, we have:

$$\mathrm{E}[(\Delta u_{it})^2] = \mathrm{E}[(u_{it} - u_{i,t-1})^2] = \mathrm{E}(u_{it}^2) + \mathrm{E}(u_{i,t-1}^2) = 2\sigma_u^2,$$

$$\mathrm{E}[\Delta u_{it} \Delta u_{i,t-1}] = \mathrm{E}[(u_{it} - u_{i,t-1})(u_{i,t-1} - u_{i,t-2})] = -\mathrm{E}(u_{i,t-1}^2) = -\sigma_u^2,$$

$$\mathrm{E}[\Delta u_{it} \Delta u_{i,t-k}] = 0 \qquad \forall k \geq 2.$$

Hence,

$$\mathbf{\Sigma}_{\Delta \mathbf{u}} = \mathrm{E}\left(\Delta \mathbf{u}_i \Delta \mathbf{u}_i'\right) = \sigma_u^2 \begin{pmatrix} 2 & -1 & 0 & \dots & 0 & 0 \\ -1 & 2 & -1 & \dots & 0 & 0 \\ 0 & -1 & 2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 2 & -1 \\ 0 & 0 & 0 & \dots & -1 & 2 \end{pmatrix} = \sigma_u^2 \mathbf{G}$$

with the $(T-2) \times (T-2)$ matrix $\mathbf{G}$ defined implicitly.

One-step GMM estimator ⋆

Under the white noise and homoskedasticity assumptions, we thus find

$$\mathrm{E}(\mathbf{W}_i'\Delta\mathbf{u}_i\Delta\mathbf{u}_i'\mathbf{W}_i) = \sigma_u^2\,\mathrm{E}(\mathbf{W}_i'\mathbf{G}\mathbf{W}_i).$$

Since σ_u^2 is a scalar that does not change the relative weighting, we may use

$$\boldsymbol{\Xi}_1^{-1} = \boldsymbol{\Lambda}_1 = \mathrm{E}(\mathbf{W}_i'\mathbf{G}\mathbf{W}_i),$$

which can be consistently estimated as

$$\hat{\boldsymbol{\Xi}}_1^{-1} = \hat{\boldsymbol{\Lambda}}_1 = N^{-1}\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{G}\mathbf{W}_i = N^{-1}\mathbf{W}'(\mathbf{I}_N\otimes\mathbf{G})\mathbf{W}.$$

Note that all we need for this estimator is the data $\mathbf{W}$.

This leads to... ⋆

The one-step GMM estimator minimizes the objective function

$$
\begin{aligned}
Q_N(\alpha) &= \left[N^{-1}\sum_{i=1}^{N}\mathbf{W}_i'\Delta\mathbf{u}_i\right]'\left[\sum_{i=1}^{N}\mathbf{W}_i'\mathbf{G}\mathbf{W}_i\right]^{-1}\left[N^{-1}\sum_{i=1}^{N}\mathbf{W}_i'\Delta\mathbf{u}_i\right] \\
&= \Delta\mathbf{u}'\mathbf{W}\left[\mathbf{W}'(\mathbf{I}_N\otimes\mathbf{G})\mathbf{W}\right]^{-1}\mathbf{W}'\Delta\mathbf{u}
\end{aligned}
$$

This yields the Arellano and Bond (1991) one-step estimator

$$
\begin{aligned}
\hat{\alpha}_{AB,1} \quad = \quad & \left[\Delta\mathbf{y}_{-1}'\mathbf{W}\left(\mathbf{W}'(\mathbf{I}_N\otimes\mathbf{G})\mathbf{W}\right)^{-1}\mathbf{W}'\Delta\mathbf{y}_{-1}\right]^{-1} \\
& \times\left[\Delta\mathbf{y}_{-1}'\mathbf{W}\left(\mathbf{W}'(\mathbf{I}_N\otimes\mathbf{G})\mathbf{W}\right)^{-1}\mathbf{W}'\Delta\mathbf{y}\right]
\end{aligned}
$$

Two-step GMM estimator ⋆

The homoskedasticity assumptions are often deemed too restrictive. Therefore, Arellano and Bond (1991) suggest to use a nonparametric (robust) estimator of the optimal weighting matrix in a second step.

Based on first-step residuals this is straightforward:

$$\hat{\mathbf{\Xi}}_2 = \hat{\mathbf{\Lambda}}_2^{-1} = \left[N^{-1} \sum_{i=1}^{N} \mathbf{W}_i' \Delta \hat{\mathbf{u}}_i \Delta \hat{\mathbf{u}}_i' \mathbf{W}_i \right]^{-1} .$$

The resulting estimator is the two-step GMM estimator:

$$\hat{\alpha}_{AB,2} = \left[\Delta \mathbf{y}_{-1}' \mathbf{W} \hat{\mathbf{\Xi}}_2 \mathbf{W}' \Delta \mathbf{y}_{-1} \right]^{-1} \left[\Delta \mathbf{y}_{-1}' \mathbf{W} \hat{\mathbf{\Xi}}_2 \mathbf{W}' \Delta \mathbf{y} \right].$$

A consistent estimator of the asymptotic variance is given by

$$\mathrm{Avar} \left[\sqrt{N} (\hat{\alpha}_{AB,2} - \alpha) \right] = \left[\Delta \mathbf{y}_{-1}' \mathbf{W} \hat{\mathbf{\Xi}}_2 \mathbf{W}' \Delta \mathbf{y}_{-1} \right]^{-1} .$$