Lecture 5: First-Differencing IV Methods and GMM

Jesper Riis–Vestergaard Sørensen

University of Copenhagen, Department of Economics

Plan for Panel Data Lectures

Lecture 1: Linear model + OLS in cross section (W.4)

Lecture 2: Fixed effects + First differences (W.10)

Lecture 3: Random effects + Hausman test (W.10)

Lecture 4: Predetermined regressors (W.11)

Lecture 5: First-Differencing IV Methods + GMM (W.11).

Next: Estimation methods for *non*linear models.

Linear Panel Model under Sequential Exogeneity

Still consider model:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + u_{it},$$
$$E(\underbrace{u_{it}|\mathbf{x}_{it}, \mathbf{x}_{it-1}, \ldots, \mathbf{x}_{i1}, c_i}_{\text{seq. exogeneity}}) = 0, \quad t = 1, 2, \ldots, T.$$

In differences:

$$\Delta y_{it} = \Delta \mathbf{x}_{it}\boldsymbol{\beta} + \Delta u_{it}, \quad t = 2, 3, \ldots, T.$$

Have shown:

- $\Delta \mathbf{x}_{it}$ endogenous, $E(\Delta \mathbf{x}'_{it}\Delta u_{it}) = -E\left(\mathbf{x}'_{it}u_{it-1}\right) \neq \mathbf{0}$.
- Valid instruments at time t,

$$\mathbf{x}^o_{it-1} = \left(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{it-1}\right). \qquad \left(1 \times (t-1)K\right)$$

Example: Instruments in AR(1) Model

AR(1) model:

$$y_{it} = \rho y_{it-1} + c_i + u_{it},$$
$$E(u_{it} | y_{it-1}, y_{it-2}, \ldots, y_{i0}, c_i) = 0, \quad t = 1, 2, \ldots, T.$$

''

x_{it}

Implies:

$$\boxed{\Delta y_{it} = \rho \Delta y_{it-1} + \Delta u_{it}, \quad t = 2, 3, \ldots, T.}$$

Here:

- $\mathbf{x}_{it} = y_{it-1}$.
- Valid instruments for $\Delta \mathbf{x}_{it} = \Delta y_{it-1}$ at time t:

$$\mathbf{x}^o_{it-1} = (y_{i0}, y_{i1}, \ldots, y_{it-2}) = \mathbf{y}^o_{it-2}.$$

Pooled IV Estimation of AR(1) Model

- Pooled IV (PIV) with instrument z_{it}:

$$\boxed{\widehat{\rho}_{PIV} := \frac{\sum_{i=1}^{N} \sum_t z_{it} \Delta y_{it}}{\sum_{i=1}^{N} \sum_t z_{it} \Delta y_{it-1}}.}$$

- With $z_{it} = y_{it-2}$, is in y_{it-2}^{o}

$$\widehat{\rho}_{PIV} = \frac{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-2} \Delta y_{it}}{\sum_{i=1}^{N} \sum_{t=2}^{T} y_{it-2} \Delta y_{it-1}}.$$

- Alternatively, with $z_{it} = \Delta y_{it-2}$, $= y_{it-2} - y_{it-3}$

$$\widehat{\rho}_{PIV} = \frac{\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta y_{it-2} \Delta y_{it}}{\sum_{i=1}^{N} \sum_{t=3}^{T} \Delta y_{it-2} \Delta y_{it-1}}.$$

- Using more IVs? Efficiency?

Outline

First-Differencing IV Methods and GMM

First-Differenced Equation as System

Write equation in first differences

$$\Delta y_{it} = \Delta \mathbf{x}_{it} \boldsymbol{\beta} + \Delta u_{it}, \quad t = 2, 3, \ldots, T,$$

as *linear system*

$$\Delta \mathbf{y}_i = \Delta \mathbf{X}_i \boldsymbol{\beta} + \Delta \mathbf{u}_i.$$

- Here $\Delta \mathbf{y}_i$ $[(T-1) \times 1]$, $\Delta \mathbf{X}_i$ $[(T-1) \times K]$.

Will instrument (rows of) $\Delta \mathbf{X}_i$.

Instruments under Sequential Exogeneity

Gather valid IVs into instrument matrix

$$
\mathbf{Z}_i := \begin{array}{cc}
\begin{bmatrix}
\mathbf{x}_{i1}^{o} & \mathbf{0}_{1\times 2K} & \mathbf{0}_{1\times 3K} & \cdots & \mathbf{0}_{1\times (T-1)K} \\
\mathbf{0}_{1\times K} & \mathbf{x}_{i2}^{o} & \mathbf{0}_{1\times 3K} & \cdots & \mathbf{0}_{1\times (T-1)K} \\
\mathbf{0}_{1\times K} & \mathbf{0}_{1\times 2K} & \mathbf{x}_{i3}^{o} & & \\
\vdots & \vdots & & \ddots & \vdots \\
\mathbf{0}_{1\times K} & \mathbf{0}_{1\times 2K} & & \cdots & \mathbf{x}_{iT-1}^{o}
\end{bmatrix}_{((T-1)\times L)} . &
\begin{array}{c} t=2 \\ t=3 \\ \vdots \\ \\ t=T \end{array}
\end{array}
$$

where $L := KT(T-1)/2$.

- $\mathbf{Z}_i$ has $T-1$ rows corresponding to $T-1$ rows of system.
- $L = KT(T-1)/2$ comes from

$$K + 2K + 3K + \cdots + (T-1)K = K\sum_{t=1}^{T-1} t.$$

$t=2$ $t=3$ .... $t=T$

$\sum_{m=1}^{M} m = \frac{M(M+1)}{2}$

Here $M = T-1$

Instrument Matrix in AR(1) Model

- Instrument matrix with y_{it-1} only regressor $(K = 1)$:

$$\mathbf{Z}_i = \begin{bmatrix} y_{i0} & 0 & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & y_{i0} & y_{i1} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & y_{i0} & y_{i1} & y_{i2} & & 0 \\ \vdots & \vdots & & & & & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & 0 & \cdots & \mathbf{y}_{iT-2}^{o} \end{bmatrix} \begin{pmatrix} t = 2 \\ t = 3 \\ t = 4 \\ \vdots \\ t = T \end{pmatrix}.$$

- Here $L = T\left(T - 1\right)/2$.
- Treats y_{i0} as observed.
- Otherwise, redefine $T + 1 \leftarrow T$ and relabel.

Generalized Method of Moments

Method of Moments

- Instrument matrix yields

(under seq. exog.)

$$E\left(\overset{L\times 1}{\mathbf{Z}_i'\Delta\mathbf{u}_i}\right) = \mathbf{0}_{L\times 1}.$$

- Under a rank condition (check!), $\boldsymbol{\beta}$ unique solution to

$$E\left[\mathbf{Z}_i'\left(\Delta\mathbf{y}_i - \Delta\mathbf{X}_i\mathbf{b}\right)\right] = \mathbf{0}_{L\times 1}. \qquad \text{(in } \mathbf{b}\text{)}$$

- Analogy principle suggests estimator as solution to

$$\frac{1}{N}\sum_{i=1}^{N}\mathbf{Z}_i'\left(\Delta\mathbf{y}_i - \Delta\mathbf{X}_i\mathbf{b}\right) = \mathbf{0}_{L\times 1}.$$

- Method of moments (MM).

Generalized Method of Moments

- **Issue:** $\mathbf{b} \mapsto N^{-1} \sum_{i=1}^{N} \mathbf{Z}_i' (\Delta \mathbf{y}_i - \Delta \mathbf{X}_i \mathbf{b})$ may have no root.
- **Idea:** Choose $\mathbf{b}$ to minimize distance to origin.
- Choose $\widehat{\mathbf{W}}$ $(L \times L)$ symmetric positive definite.
  - Weighting matrix. Possibly random.

Generalized method of moments (GMM) minimizes

$$\left[\sum_{i=1}^{N} \mathbf{Z}_i' (\Delta \mathbf{y}_i - \Delta \mathbf{X}_i \mathbf{b}) \right]' \widehat{\mathbf{W}} \left[\sum_{i=1}^{N} \mathbf{Z}_i' (\Delta \mathbf{y}_i - \Delta \mathbf{X}_i \mathbf{b}) \right].$$

Closed-Form Solution

Stack over i:

- $\mathbf{Z}$ $[N(T-1) \times L]$,
- $\Delta\mathbf{y}$ $[N(T-1) \times 1]$,
- $\Delta\mathbf{X}$ $[N(T-1) \times K]$.

$\widehat{\beta}_{GMM}$ minimizes *quadratic form*

$$[\mathbf{Z}'(\Delta\mathbf{y} - \Delta\mathbf{X}\mathbf{b})]' \widehat{\mathbf{W}} [\mathbf{Z}'(\Delta\mathbf{y} - \Delta\mathbf{X}\mathbf{b})].$$

Multivariate calculus/direct substitution shows

$$\widehat{\beta}_{GMM} = \left(\Delta\mathbf{X}'\mathbf{Z}\widehat{\mathbf{W}}\mathbf{Z}'\Delta\mathbf{X}\right)^{-1}\left(\Delta\mathbf{X}'\mathbf{Z}\widehat{\mathbf{W}}\mathbf{Z}'\Delta\mathbf{y}\right).$$

Choice of Weighting Matrix

Choice of Weighting Matrix

- E.g. $\widehat{\mathbf{W}} = \mathbf{I}_L \Rightarrow$ equal weighting
- E.g. $\widehat{\mathbf{W}} = (\mathbf{Z}'\mathbf{Z})^{-1} \Rightarrow$ less weight to "noisy" instruments.
- Provided also $\widehat{\mathbf{W}} \to_p \mathbf{W}$ nonrandom symmetric p.d.,

as $N \to \infty$

$$\widehat{\boldsymbol{\beta}}_{GMM} \xrightarrow{p} \boldsymbol{\beta}.$$

- In fact, $\widehat{\boldsymbol{\beta}}_{GMM}$ $\sqrt{N}$-asymptotically normal.
- Asymptotic variance reveals optimal weighting matrix:

$$\mathbf{W}_{\text{opt}} := \left[E\left(\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i\right)\right]^{-1}, \quad \mathbf{e}_i := \Delta\mathbf{u}_i,$$

$= \mathrm{var}(\mathbf{Z}_i'\mathbf{e}_i)$

- Then

$$\text{Avar}(\widehat{\boldsymbol{\beta}}^{\text{opt}}_{GMM}) = \left\{E\left(\Delta\mathbf{X}_i'\mathbf{Z}_i\right)\left[E\left(\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i\right)\right]^{-1}E\left(\mathbf{Z}_i'\Delta\mathbf{X}_i\right)\right\}^{-1}/N.$$

Optimally Weighted GMM

Issue: $\mathbf{W}_{\text{opt}} = \left[E\left(\mathbf{Z}_i'\mathbf{e}_i\mathbf{e}_i'\mathbf{Z}_i\right)\right]^{-1}$ depends on (unknown) $\boldsymbol{\beta}$.

Solution: Two-Step Procedure

1. Estimate $\boldsymbol{\beta}$ using weights $\check{\mathbf{W}} := (\mathbf{Z}'\mathbf{Z})^{-1}$ to get

$$\check{\boldsymbol{\beta}} = \left(\Delta\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\Delta\mathbf{X}\right)^{-1}\left(\Delta\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\Delta\mathbf{y}\right),$$

   and store $\check{\mathbf{e}}_i := \Delta\mathbf{y}_i - \Delta\mathbf{X}_i\check{\boldsymbol{\beta}}$.

2. Optimally weighted GMM ($\widehat{\boldsymbol{\beta}}^{\text{opt}}_{GMM}$) arises from GMM with

$$\widehat{\mathbf{W}} = \left(\frac{1}{N}\sum_i \mathbf{Z}_i'\check{\mathbf{e}}_i\check{\mathbf{e}}_i'\mathbf{Z}_i\right)^{-1}. \qquad \left(= \widehat{\mathbf{W}}_{\text{opt}}\right)$$

Optimally Weighted GMM

- First-step akin to 2SLS applied to differenced equation

$$\check{\boldsymbol{\beta}} = \left(\Delta\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\Delta\mathbf{X}\right)^{-1}\left(\Delta\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\Delta\mathbf{y}\right).$$

- Dealing with system $\Rightarrow$ system 2SLS (S2SLS).
- Yield consistent but inefficient (GMM) estimator.
- Efficient GMM arises from optimal weighting (asymptotically).

Inference

Asymptotic Variance Estimation

W_{opt}

$$\text{Avar}(\hat\beta_{GMM}^{opt}) = \left\{ E(\Delta X_i' z_i)[E(z_i e_i e_i' z_i)]^{-1} E(z_i' \Delta X_i) \right\}^{-1} / N$$

Consistent asymptotic variance estimator of efficient GMM:

$$\widehat{\text{Avar}(\widehat{\beta}_{GMM}^{\text{opt}})} := \left[(\mathbf{X}'\mathbf{Z}) \left(\sum_{i=1}^{N} \mathbf{Z}_i' \widehat{\mathbf{e}}_i \widehat{\mathbf{e}}_i' \mathbf{Z}_i \right)^{-1} (\mathbf{Z}'\mathbf{X}) \right]^{-1},$$

$\Delta X' Z$ $\quad$ $= Z' \Delta X$

where $\widehat{\mathbf{e}}_i := \Delta \mathbf{y}_i - \Delta \mathbf{X}_i \widehat{\beta}_{GMM}^{\text{opt}}$.

Q: Use 1st or 2nd stage residuals?

A: Asymptotically irrelevant. May matter in finite sample.

Testing Overidentification Restrictions

- With K elements of $\Delta \mathbf{x}_{it}$, K parameters to be estimated.
- L orthogonality conditions (# columns of $\mathbf{Z}$).
- $L = K$: "Exact identification".
- $L > K$: "Overidentification."
- Idea: K orthogonality conditions used for estimation.
- $L - K$ degrees of freedom left for testing.
- Overidentification allows test of

$$\mathrm{H}_0 : \quad E\left(\mathbf{Z}_i' \Delta \mathbf{u}_i\right) = \mathbf{0}_{L \times 1}$$

Testing Overidentification Restrictions

- Stack GMM residuals into $\widehat{\mathbf{e}}$ $[N(T-1)\times 1]$.
- Let $\widehat{\mathbf{W}}_{\mathrm{opt}}$ estimate optimal weighting matrix.

Overidentification test statistic (Sargan's J):

$$J := \widehat{\mathbf{e}}'\mathbf{Z}\widehat{\mathbf{W}}_{\mathrm{opt}}\mathbf{Z}'\widehat{\mathbf{e}}/N.$$

May show: Under H_0, $J \to_d \chi^2_{L-K}$.

Reject null ("validity of instruments") at level $\alpha \Leftrightarrow$

$$J > (1-\alpha)\text{-quantile of } \chi^2_{L-K}.$$

A More General Dynamic Model

More General Dynamic Model

Starting point

$$y_{it} = \overbrace{\rho y_{it-1} + \mathbf{z}_{it}\gamma}^{= x_{it}\beta} + c_i + u_{it}, \quad t = 1, 2, \ldots, T.$$

Strict exogeneity (still) ruled out by presence of LDV.

In differences

$$\boxed{\Delta y_{it} = \rho \Delta y_{it-1} + \Delta \mathbf{z}_{it}\gamma + \Delta u_{it}, \quad t = 2, 3, \ldots, T.}$$

Instruments for Δy_{it-1}: $\mathbf{y}^o_{it-2}$.

Q: For $\Delta \mathbf{z}_{it}$?

Instrumenting Other Regressors

$$\Delta y_{it} = \rho \Delta y_{it-1} + \Delta \mathbf{z}_{it}\gamma + \Delta u_{it}, \quad t = 2, 3, \ldots, T.$$

Instruments available for $\Delta \mathbf{z}_{it}$?

- Depends on what we're willing to assume about $\{\mathbf{z}_{it}\}_{t=1}^{T}$.

$\{\mathbf{z}_{it}\}_{t=1}^{T}$ strictly exogenous:

- May use $\mathbf{z}_i = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{iT})$, which $= \mathbf{z}^o_{iT}$.
- Or just $\Delta \mathbf{z}_{it}$ (instrumenting itself).

$\{\mathbf{z}_{it}\}_{t=1}^{T}$ predetermined:

- May use $\mathbf{z}^o_{it-1} = (\mathbf{z}_{i1}, \mathbf{z}_{i2}, \ldots, \mathbf{z}_{it-1})$, -1 due to FD.

Contemporaneous Correlation

Issue: Contemporanous correlation $E(\mathbf{z}'_{it} u_{it}) \neq \mathbf{0}$.

- Rules out predeterminedness.
- Persists in differences.

May arise for various reasons...

1. Omitted (time-varying) variable,
2. Simultaneity.
3. Measurement error.

We need to find an external instrument!

- Where? Case-specific.
- Sometimes we can exploit panel structure.

Problem Set 4

Models female labor force participation $y_{it} = \mathbf{1}\,(\text{works in } t)$.

$\mathbf{z}_{it}$ contains

- $k_{it} = \mathbf{1}\,(\text{kids aged } [2, 6])$,
- $f_{it} = \mathbf{1}\,(\text{gives birth in year } t)$.

Both can be argued correlated with $c_i \Rightarrow$ FD.

Fertility f_{it} likely

- Contemporaneously endogenous.
- Requires **external instrument**.

Having kids aged 2–6 is likely

- Predetermined (as u_{it} affects f_{it+1} and thus k_{it+3}).
- Use $\mathbf{k}^o_{it-1} := (k_{i1}, \ldots, k_{it-1})$ (upon differencing).

Summary of Linear Panel Data Model Lectures

Summary: Linear Panel Data Models

Appropriate estimation method depends on assumptions.

If $E\left(u_{it} | \mathbf{x}_i, c_i\right) = 0$ and $E\left(c_i | \mathbf{x}_i\right) = 0$:

- RE estimator appropriate.
- Exploits both within i variation over time...
- ... and between is variation.

If $E\left(u_{it} | \mathbf{x}_i, c_i\right) = 0$ but $E\left(c_i | \mathbf{x}_i\right) \neq 0$:

- FE/D appropriate.
- Exploits within i variation.
- Choice FE v. FD then a matter of efficiency.

Summary: Linear Panel Data Models

If $E\left(u_{it} \mid \mathbf{x}_{i1}, \ldots, \mathbf{x}_{it}, c_i\right) = 0$ but $E\left(u_{it} \mid \mathbf{x}_{it+1}, \ldots, \mathbf{x}_{iT}, c_i\right) \neq 0$ and $E\left(c_i \mid \mathbf{x}_i\right) \neq 0$:

- IV/GMM estimators appropriate.
- FD equation (system) of interest.
- May use lagged levels (or differences) as IVs.

$E\left(\mathbf{x}_{it}' u_{it}\right) \neq \mathbf{0}$:

- Typically need external instrument.
- Uncorrelated with (composite) error.
- Solution case specific.