Dynamic Panel Data
Ch 2. Dynamic Linear Panel Models

Pr. Philippe Polomé, Université Lumière Lyon 2

M2 EcoFi

2016 – 2017

Overview of Ch. 2

Outline

Introduction

- The previous chapter presented variants of the linear panel data model with
  - a FE or RE (random) intercept and
  - **strongly exogenous** regressors

$$E[\varepsilon_{it}|\alpha_i, x_{i1}, ..., x_{iT}] = 0, \ t = 1, ..., T$$

- Linear models may relax the strong exogeneity assumption via
  - **endogenous** regressors $E[\varepsilon_{it}|x_{ijt}] \neq 0$ for some j
  - **lagged dependent** variables as regressors
    - endogenous in a panel context with autocorrelation
    - or because of estimation – see later
- Dynamic linear panels treat the 2^{nd} case only
  - But use applications of the GMM estimator
    - That may also be used to address the 1^{st} case
  - We will go through the general GMM first

Outline

Analogy Principle

- GMM estimators based on the **analogy principle**
  - population moment conditions
    - lead to sample moment conditions
    - are used to estimate parameters
- Classic example of MM
  - estimation of the population mean when y is iid with mean μ
- In the population $E[y-\mu]=0$ by definition
  - Replacing $E[\cdot]$ for the population by $N^{-1}\sum_{i=1}^{N}(\cdot)$ for the sample yields the corresponding sample moment :

$$\frac{1}{N}\sum_{i=1}^{N}(y_i-\mu)=0$$

- Solving for μ leads to the estimator $\hat{\mu}_{MM}=N^{-1}\sum_{i=1}^{N}y_i=\bar{y}$
  - The MM estimator of the population mean is the sample mean

Linear Cross-Section Regression

- linear regression model $y = x^{'}\beta + u$
  - x and β are $K \times 1$ vectors
- Suppose $E[u|x] = 0$
  - law of iterated expectations
    - K unconditional moment conditions $E[xu] = 0$
  - Thus, when the error has conditional mean zero / is "exogenous" / orthogonal

$$E\left[x\left(y - x^{'}\beta\right)\right] = 0$$

- MM estimator = solution to the corresponding K sample moment conditions

$$\frac{1}{N}\sum_{i=1}^{N} x_i\left(y_i - x_i^{'}\beta\right) = 0$$

  - $\hat{\beta}_{MM} = \left(\sum_i x_i^{'} x_i\right)^{-1} \sum_i x_i^{'} y_i$

Linear IV

- Endogeneity : $E\left[u|x\right] \neq 0$
- Assume we have **instruments** z such that $E\left[u|z\right] = 0$ **and**
  - The instruments are **good**
    - strongly correlated with the regressors
  - $dim\left(z\right) = dim\left(x\right)$: exactly one instrument per regressor
    - exactly-identified model
- Thus, $E\left[u|z\right] = 0$ provides K sample moment conditions
- Then $\hat{\beta}_{MM} = \left(\sum_i z_i' x_i\right)^{-1} \sum_i z_i' y_i$ is consistent
  - while $\hat{\beta}_{OLS} = \left(\sum_i x_i' x_i\right)^{-1} \sum_i x_i' y_i$ is inconsistent
  - $\hat{\beta}_{MM}$ is the Instrumental Variable IV estimator
    - an application of MM estimation

Additional Moment Restrictions

- **additional** moments/instruments can improve **efficiency**
  - since they add information to the estimator
    - that is relevant (correlated to X)
  - but requires adaptation of MM
- Consider that $dim(z) > dim(x)$: more instruments than regressors
  - The model is said over-identified
  - Which ones do we take ? Any selection is **arbitrary**
  - let z_1 and z_2 be subsets of z such that $dim(z_1) = dim(z_2) = dim(x)$
  - Then, $\hat{\beta}_{MM1} = \left(Z_1'X\right)^{-1} Z_1'Y \neq \left(Z_2'X\right)^{-1} Z_2'Y = \hat{\beta}_{MM2}$

Additional Moment Restrictions 2

- If we write $E\left[Z\left(y-X'\beta\right)\right]$
  - the **moments** conditions as above
  - there are **more** conditions than parameters to estimate
- The GMM estimator **chooses** $\hat{\beta}$ to make the vector of sample moment conditions $\frac{1}{N}\sum_i z_i\left(y_i-x_i'\beta\right)$ as **small** as possible in quadratic terms

That is $\hat{\beta}_{GMM}$ **minimizes** :

$$Q_N(\beta)=\left[\frac{1}{N}\sum_i z_i\left(y_i-x_i'\beta\right)\right]'\mathbf{W}_N\left[\frac{1}{N}\sum_i z_i\left(y_i-x_i'\beta\right)\right]$$

where $\mathbf{W}_N$ is a weighting matrix

Additional Moment Restrictions 3

- How to choose $\mathbf{W}_N$?
  - Let $dim(z) = r$; $\mathbf{W}_N$ is $r \times r$, sdp and does not depend on β
  - Different choices of $\mathbf{W}_N$ lead to different estimators that are all consistent
    - but have different variances when $r > k$
  - GMM specifies the **optimal** choice of weighting matrix $\mathbf{W}_N$
    - depending on the case at hand (number of instruments, heteroskedasticity, autocorrelation)
    - such that $\hat{\beta}_{GMM}$ has the **smallest asymptotic variance**
- The exact definition of $\mathbf{W}_N$ depends on the particular case
  - 3 cases for panel

Outline

The Usual Panel Data Model

- Individual-specific effect α_i model

$$y_{it} = \alpha_i + \mathbf{x}_{it}^{'}\beta + \epsilon_{it} \tag{1}$$

  - x_{it} may include
    - both time-varying and time-invariant components
    - an intercept
- Some components of the regressors x_{it} are now assumed to be **endogenous**
  - $E\left[\mathbf{x}_{it}\left(\alpha_i + \epsilon_{it}\right)\right] \neq 0$
  - This leads to a re-definition of what is FE & RE in the next slide

Random and Fixed Effects Panel with Endogeneity

- We call **RE** if $\exists$ instruments Z_i s.t.

$$E\left[Z_{is}'\left(\alpha_i+\epsilon_{it}\right)\right]=0\,\forall\,period\ s$$

  - Then GMM is applied following the formulas below
- **FE** if
  - it is possible only to find instruments s.t. $E\left[Z_{is}'\epsilon_{it}\right]=0\,\forall s$
  - **but** $E\left[Z_{is}'\alpha_i\right]\neq 0$ for some s at least
  - Then α_i must be eliminated by differencing
    - & only the coefficients of the time-varying regressors are identified

Simplification for presentation

- **no individual-specific effect** α_i
- x_{it} includes **only** current-period variable
  - The linear panel model becomes

$$y_{it} = x_{it}\beta + u_{it} \tag{2}$$

- Stack (**bold**) all T observations for the i^{th} individual

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{u}_i \tag{3}$$

- The estimators presented below
  - may however include individual-specific effects
  - Via a data transformation as in Ch1

Panel GMM Moment Conditions

- Assume a $T \times r$ matrix of instruments $\mathbf{Z}_i$
  - $r \geq K$ is the number of instruments
  - that satisfy the r moment conditions :

$$E\left[\mathbf{Z}_i'\mathbf{u}_i\right] = 0 \tag{4}$$

- The GMM estimator based on these moment conditions **minimizes** a quadratic form as above : $\hat{\beta}_{PGMM} =$

$$\left[\left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_i \mathbf{Z}_i'\mathbf{X}_i\right)\right]^{-1}\left(\sum_i \mathbf{X}_i'\mathbf{Z}_i\right)\mathbf{W}_N\left(\sum_i \mathbf{Z}_i'\mathbf{y}_i\right)$$

- Consistency of this estimator if (4) holds

Case 1. Just-Identified Panel GMM

- Different weighting matrices $\mathbf{W}_N$ lead to **different** GMM estimators
- Except in the just-identified case of $r = K$, that is $dim(z) = dim(x)$
  - Exogenous regressors serve as their own instruments
- In this case $\hat{\beta}_{PGMM}$ **simplifies** to the **IV** estimator for any $\mathbf{W}_N$

$$\hat{\beta}_{IV} = [Z'X]^{-1}Z'y$$

- In the case that all the x are exogenous
  - We can have $X = Z$ and so $\hat{\beta}_{PGMM} = \hat{\beta}_{OLS}$

Case 2 & 3. Overidentified Panel GMM

- If in addition of the r instruments above
  - We can assume that the first lag of each regressor is uncorrelated with the current error
    - then x_{it-1} is available as **additional** instruments for x_{it}
    - **weak exogeneity / predetermined instruments**
- The model is **over-identified**
  - Instead of $\hat{\beta}_{IV}$, more efficient estimation is possible using Panel GMM estimators
  - Generalisation using further lags of x_t comes to mind
    - Clearly, we can have x_t endogenous while x_{t-1} and x_{t-2} pre-determined
    - so that there are more instruments than regressors, without any external instrument
    - We will discuss that in application, it is one of the main strengths of panel

2 optimal Over-identified PGMM Estimators

- Assume no heteroskedasticity and no serial correlation
  - $\hat{\beta}_{2SLS} = \left[\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y}$
    - STATA xtivreg see below
- No such assumption (robust)
  - $\hat{\beta}_{2SGMM} = \left[\mathbf{X}'\mathbf{Z}\hat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{Z}\hat{\mathbf{S}}^{-1}\mathbf{Z}'\mathbf{y}$
    - $\hat{\mathbf{S}} = \frac{1}{N}\sum_i \mathbf{Z}_i'\hat{\mathbf{u}}_i\hat{\mathbf{u}}_i'\mathbf{Z}_i$ is a White-type robust consistent estimate for the $r \times r$ matrix $\mathbf{S} = plim\frac{1}{N}\sum_i \mathbf{Z}_i'\mathbf{u}_i\mathbf{u}_i'\mathbf{Z}_i$
  - Two-step GMM since a first-step consistent estimator of β such as $\hat{\beta}_{2SLS}$ is needed to form the residuals $\hat{\mathbf{u}}_i = \mathbf{y}_i - \mathbf{X}_i\hat{\beta}_{2SLS}$ used to compute $\hat{\mathbf{S}}$
    - STATA We'll see later in applications

2SLS reminder

- What does 2-Stage Least-Squares mean ?
  - A reminder in cross-section context for simplicity
- 2SLS is a particular application of IV
  - That leads to a natural generalization from MM to GMM

Instrumentation equation

- Let the **structural equation** $Y = X\beta + \epsilon$
  - Assume that in X, x_k is endogenous
    - assume we have one instrument z for x_k
  - The instrument matrix is Z,
    - identical to X except for the last column in which x_k is replaced by z
- **The instrumentation equation**

$$x_k = \delta_0 + \delta_1 x_1 + \ldots + \delta_{k-1} x_{k-1} + \delta_k z + \mu = Z\delta + \mu$$

  - Estimated by LS, the fitted values of x_k are
    - $\hat{x}_k = \hat{\delta}_0 + \hat{\delta}_1 x_1 + \ldots + \hat{\delta}_{k-1} x_{k-1} + \hat{\delta}_k z = Z\hat{\delta}$
    - with $\hat{\delta} = \left(Z' Z\right)^{-1} Z' x_k$
  - Thus, $\hat{x}_k$ is a valid instrument for x_k
    - as long as z is a valid instrument for x_k
    - and the other regressors are exogenous

2-Step Least-Squares

- Write $\hat{X}$, the X mtx in which x_k has been replaced by $\hat{x}_k$
  - The IV estimator using $\hat{X}$ is $\hat{\beta}_{IV} = \left(\hat{X}'X\right)^{-1}\hat{X}'Y$
    - Might be a little better than $(Z'X)^{-1}Z'Y$ since the correlation$(\hat{x}_k, x_k) \geq$ correlation(z, x_k)
    - But the main advantage will come later
- This IV estimator is **equivalent** to an OLS estimation MCO in 2 steps – 2SLS :

1. Estimate by OLS the instrumentation eq. $x_k = Z\delta + \mu$
2. Replace X by $\hat{X}$ in the structural eq. $Y = \pi\hat{X} + \nu$
   - Estimate by OLS
     - $\hat{\pi}_{2SLS} = \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y$
     - Warning: $\left(\hat{X}'\hat{X}\right)^{-1}$ and NOT $\left(\hat{X}'X\right)^{-1}$ as in $\hat{\beta}_{IV}$
   - Below, it is shown that $\hat{\pi}_{2SLS} = \hat{\beta}_{IV} = \left(\hat{X}'X\right)^{-1}\hat{X}'Y$

$\hat{\pi}_{2SLS} = \hat{\beta}_{IV} = \left(\hat{X}'X\right)^{-1}\hat{X}'Y$: Proof

- First we proove that $\hat{X} = Z\left(Z'Z\right)^{-1}Z'X$
  - Remember
    - Z is X in which x_k is replaced by z, the instrument for x_k
    - $\left(Z'Z\right)^{-1}Z'x_k = \hat{\delta}$ the LS estimate of the instrumentation eq.
  - Consider $\hat{X} = Z\left(Z'Z\right)^{-1}Z'\begin{pmatrix} X_{-k} & x_k \end{pmatrix}$
    - So $\hat{X} = \begin{pmatrix} Z\left(Z'Z\right)^{-1}Z'X_{-k} & \hat{x}_k \end{pmatrix}$
    - So, the last col. of $\hat{X}$ is $\hat{x}_k = Z\hat{\delta}$
    - Take any other col of X, x_j
    - Then the jth col of $\hat{X}$ is also the jth col of Z and is $Z\left(Z'Z\right)^{-1}Z'x_j$
  - Consider $\left(Z'Z\right)^{-1}Z'x_j = \hat{\gamma}_j$
    - This is like the OLS estimate of the regression of x_j on itself & the other regressors
    - Then, the adjustment is perfect, the residuals are zero

$\hat{\pi}_{2SLS} = \hat{\beta}_{IV} = \left(\hat{X}'X\right)^{-1}\hat{X}'Y$: Proof

- This shows that $\hat{X} = Z\left(Z'Z\right)^{-1}Z'X$
- Then, it is easy to show that

$$\begin{aligned}
\hat{\pi}_{2SLS} &= \left(\hat{X}'\hat{X}\right)^{-1}\hat{X}'Y \\
&= \left(X'Z\left(Z'Z\right)^{-1}Z'\left(Z\left(Z'Z\right)^{-1}Z'X\right)\right)^{-1}\hat{X}'Y \\
&= \left(X'Z\left(Z'Z\right)^{-1}Z'X\right)^{-1}\hat{X}'Y \\
&= \left(\hat{X}'X\right)^{-1}\hat{X}'Y = \hat{\beta}_{IV}
\end{aligned}$$

2-Step Least-Squares with several instruments

- The last proof also shows the equivalence between
  - Our previous formula for $\hat{\beta}_{2SLS}$
    - $\hat{\beta}_{2SLS} = \left[\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{X}\right]^{-1}\mathbf{X}'\mathbf{Z}\left(\mathbf{Z}'\mathbf{Z}\right)^{-1}\mathbf{Z}'\mathbf{y}$
  - and the present formula $\left(\hat{X}'X\right)^{-1}\hat{X}'Y$
    - (Except that I am not fully consistent with notation)
- Therefore 2SLS is actually IV
  - in which the instrument $\hat{x}_k$ is obtained from an instrumentation equation with p instruments

$$x_k = \delta_0 + \delta_1 x_1 + \ldots + \delta_{k-1}x_{k-1} + \delta_{k1}z_1 + \ldots + \delta_{kp}z_p + \mu$$

To sum up $\hat{\beta}_{PGMM}$ has 3 meanings

1. Just-identified case
   - # instruments = # endogenous regressors
   - $\hat{\beta}_{PGMM}$ is $\hat{\beta}_{IV}$ "one-step"
2. Over-identified & spherical case
   - # instruments > # endogenous regressors
   - No heteroskedasticity, no autocorrelation
   - $\hat{\beta}_{PGMM}$ is $\hat{\beta}_{2SLS}$ "2-step"
     - 2.1 LS estimation of the instrumentation equation(s)
     - 2.2 IV estimation using these results
3. Over-identified general case
   - $\hat{\beta}_{PGMM}$ is $\hat{\beta}_{2SGMM}$ "3-step"
     - Same 2 steps as $\hat{\beta}_{2SLS}$, construct the weighting mtx $\hat{\mathbf{S}}$
     - Use $\hat{\mathbf{S}}^{-1}$ in the general $\hat{\beta}_{PGMM}$ formula

Panel-Robust Statistical Inference

- $\hat{\beta}_{PGMM}$ is asymptotically normal
  - with a complicated asymptotic variance matrix
  - A consistent estimate of that matrix exists
    - conditionnally on a choice for $\mathbf{W}_N$
    - and if independence over i is assumed
- A White-type robust estimate exists
  - It yields panel-robust standard errors allowing for both heteroskedasticity and correlation over time
  - That may not be implemented in many packages
    - Not in Stata for the general case
    - but for some special cases
- Alternatively, the panel bootstrap could be used

Outline

Model

- Consider again the individual-specific effect α_i model (1)

$$y_{it} = \alpha_i + \mathbf{x}_{it}^{'}\beta + \epsilon_{it}$$

  where some components of x_{it} are assumed to be **endogenous**
  - $E\left[\mathbf{x}_{it}\left(\alpha_i + \epsilon_{it}\right)\right] \neq 0$
  - To simplify, take only one regressor $\mathbf{x}_{it} \rightarrow x_{it}$
- The case when the x_{it} regressor is a lag of the dependant $x_{it} \rightarrow y_{i,t-1}$
  - is discussed in details in the next section
  - So in this section, we focus on $x_{it} \neq y_{i,t-1}$

IV estimation

- Consider using $x_{i,t-1}$ as instrument for x_{it}
  - $x_{i,t-1}$ "good" instrument since correlated with x_{it}
  - Generally, we are willing to assume some autocorrelation of the x_i
    - Dangerous if x_i is (per individual) I(1)
- $x_{i,t-1}$ valid instrument only when it is not correlated with the error $\alpha_i + \epsilon_{it}$
  - We may often assume that $x_{i,t-1}$ is uncorrelated to ϵ_{it}
    - in the sense that $x_{i,t-1}$ is **predetermined** wrt ϵ_{it}
  - When $x_{i,t-1}$ is not correlated to α_i, this is a RE model
  - Else, we are in a FE model
    - and then α_i must be eliminated by differencing as in Ch. 1

Anderson–Hsiao (AH) estimator

- The AH estimator essentially applies the logic of Ch. 1
  - using an IV estimator instead of a LS one
  - In all cases, it is a 2SLS estimator
- The $x_{i,t-1}$ instrument is called internal
  - $x_{i,t-2}$ could also be used
  - Generally, lags of a regressor are called internal instruments
    - but external instruments, and their lags, may also be used
    - we do not consider them here
- Anderson-Hsiao is generally IV regression,
  - but Stata has a Panel implementation that is more convenient

AH estimator in STATA

- Requires a list of instruments
  - we will only give lags endogenous regressors
- Menu stat→Longitudinal→Endogenous covariates→Instrumental
- Command xtivreg : 5 different estimators
  - detailed below, except be, that is dropped

Within AH estimator: xtivreg, fe

- Within : transform model (1)

$$y_{it} = \alpha_i + \mathbf{x}_{it}^{'}\beta + \epsilon_{it}$$

  using within :

$$y_{it} - \bar{y}_i = \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)^{'}\beta + \epsilon_{it} - \bar{\epsilon}_i$$

  - We submit y_{it} and $\mathbf{x}_{it}$ to STATA
    - The transform is automatic
- Estimate by 2SLS
  - We supply $\mathbf{x}_{i,t-1}$ as instrument
  - But STATA will transform everything
  - So that the actual instrument is $x_{i,t-1} - \bar{x}_i$ for the endogenous regressors
    - So time invariant regressors cannot be analysed
  - $x_{it} - \bar{x}_i$ for the exogenous ones (implicitly)

First-Differences AH estimator: xtivreg, fd

- First-Differences : transform model (1)

$$y_{it} = \alpha_i + \mathbf{x}_{it}^{'}\beta + \epsilon_{it}$$

  using 1st diff (D1 or fd):

$$y_{it} - y_{i,t-1} = (\mathbf{x}_{it} - \mathbf{x}_{i,t-1})^{'}\beta + \epsilon_{it} - \epsilon_{i,t-1}$$

- 2SLS FD estimator
  - Detailed in the next section w/ $x_{it} = y_{i,t-1}$
  - but might be used in this section w/ $x_{it} \neq y_{i,t-1}$
- If the i are FE without serious correlation issues
  - FD 2SLS not as efficient as within 2SLS
- If no endogenous variable is a lagged dependent variable
  - and the i are iid RE
  - then RE 2SLS is more efficient than FD 2SLS
- However, when these conditions fails, the FD 2SLS may be the only consistent estimator
  - e.g. with a lagged dependent variable

Random Effects AH estimator : xtivreg, re

- Random effects: transform model (1)

$$y_{it} = \alpha_i + \mathbf{x}_{it}^{'}\beta + \epsilon_{it}$$

using the RE transformation of Ch. 1 :

$$y_{it} - \hat{\lambda}\bar{y}_i = \left(1 - \hat{\lambda}\right)\mu + \left(x_{it} - \hat{\lambda}\bar{x}_i\right)^{'}\beta + \nu_{it}$$

- 2SLS RE, two implementations
  - G2SLS from Balestra and Varadharajan-Krishnakumar (default)
  - EC2SLS from Baltagi
  - I do not detail, we use the default
- Application is postponed to next section

Outline

Lagged dependent variable

- Regressors include the dependent variable **lagged** once
  - That is the **dynamic panel data** model DPD

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}^{'} \beta + \alpha_i + \epsilon_{it} \tag{5}$$

- Assume $|\gamma| < 1$
  - In applications, that can be tested using (panel) unit-root tests
  - $\neg R$ unit root, then y_{it} is a random walk
  - Therefore, inference is invalid
    - Estimation may be inconsistent as LS/IV assumptions are not satisfied
    - We do not deal with this case in this course

Outline

Correlation Between y_{it} and $y_{i,t-1}$

- Time-series correlation in y_{it}
  - is induced directly by $y_{i,t-1}$
  - in addition to the indirect effect via α_i
  - These two causes lead to **different interpretations** of correlation over time
- From (5) with $\beta = 0$
  - so that $y_{it} = \gamma y_{i,t-1} + \alpha_i + \epsilon_{it}$

$$\begin{aligned} Cor\left[y_{it}, y_{i,t-1}\right] &= Cor\left[\gamma y_{i,t-1} + \alpha_i + \epsilon_{it}, y_{i,t-1}\right] \\ &= \gamma Cor\left[y_{i,t-1}, y_{i,t-1}\right] + Cor\left[\alpha_i, y_{i,t-1}\right] \\ &= \gamma + \frac{1-\gamma}{1 + (1-\gamma)\,\sigma_\epsilon^2 / (1+\gamma)\,\sigma_\alpha^2} \end{aligned}$$

- The second equality assumes $Cor\left[\epsilon_{it}, y_{i,t-1}\right] = 0$
  - $y_{i,t-1}$ is predetermined wrt ϵ_{it}
- The third equality is obtained after some algebra
  - for the case of RE with $\epsilon_{it} \sim iid\left[0, \sigma_\epsilon^2\right]$ & $\alpha_{it} \sim iid\left[0, \sigma_\alpha^2\right]$

2 possible reasons for correlation between y_{it} and $y_{i,t-1}$

- **True state dependence**
  - When correlation over time is due to the causal mechanism that $y_{i,t-1}$ last period determines y_{it} this period
  - This dependence is relatively large
    - if the individual effect $\alpha_i \simeq 0$
    - or more generally when σ_α^2 is small relative to σ_ϵ^2 as then $Cor\left[y_{it}, y_{i,t-1}\right] \simeq \gamma$
- **Spurious** correlation between y_{it} and $y_{i,t-1}$ arises even if there is no causal mechanism
  - due to **unobserved heterogeneity**
    - so $\gamma = 0$
    - but nonetheless $\hat{\gamma}_{OLS} \neq 0$ because $Cor\left[y_{it}, y_{i,t-1}\right] = \sigma_\alpha^2 / \left(\sigma_\alpha^2 + \sigma_\epsilon^2\right)$ as in Ch. 1
  - also due to time-series issues: y_{it} is I(1)
    - Not covered in this Ch.

True State Dependence & Unobserved Heterogeneity

- Both extremes permit the correlation to be close to 1
  - because either $\gamma \to 1$ or $\sigma_\alpha^2 / \sigma_\epsilon^2 \to 0$
  - However, these 2 explanations give different policy implications
- **True state dependence** explanation
  - Earnings y_{it} are continuously high (or low) over time even after controlling for regressors x_{it} because future earnings are determined by past earnings and γ is large
- **Unobserved heterogeneity** explanation
  - Actually γ is small, but important variables have been **omitted** from x_{it}, leading to a high α_i in each time period
    - So that $\hat{\gamma}_{LS}$ appears high
- That is, are people poor (or rich) because
  - They were poor (or rich)?
    - In that case, poverty may be addressed by transfering money
  - Or they have individual characteristics that make them poor?
    - In that case, poverty might be better addressed by education

Inconsistency of Ch. 1 Estimators

- When the regressors include lagged dependent variables
  - **All** the estimators from Ch.1 are **inconsistent**
- **OLS** of y_{it} on $y_{i,t-1}$ and $\mathbf{x}_{it}$
  - Error term $(\alpha_i + \epsilon_{it})$, correlated $y_{i,t-1}$ through α_i
- **Within** estimator : $y_{it} - \bar{y}_i$ on $(y_{i,t-1} - \bar{y}_i)$ and $(\mathbf{x}_{it} - \bar{\mathbf{x}}_i)$ with error $(\epsilon_{it} - \bar{\epsilon}_i)$
  - $y_{i,t-1}$ is correlated with $\epsilon_{i,t-1}$ and hence $\bar{\epsilon}_i$
  - Thus, within **creates** a correlation between regressor and error
- Inconsistency also for the **RE** estimator from Ch 1
  - since RE is based on the transformation $y_{it} - \hat{\lambda}\bar{y}_i$

Nickel bias

- It has been shown that the bias $\hat{\gamma} - \gamma$
  - such correlation generates
  - on the within estimator of γ
    - in the DPD model $y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}^{'} \beta + \alpha_i + \epsilon_{it}$
  - is approximately $-(1+\gamma)/(1+T)$ as $n \to \infty$
    - It disappears with long series
    - But can be important when T is small
    - e.g. with $T = 10$ and $\gamma = 0.5$, the bias is -0.167, about 1/3 of the parameter
- Additional regressors
  - or additional obs.
  - do not remove the bias
  - If the additional regressors are correlated with y_{t-1}
    - Their estimated coef will also be inconsistent

Outline

First-differences Model

- First-difference the dynamic model (5), $t = 2, \ldots, T$

$$y_{it} - y_{i,t-1} = \gamma \left(y_{i,t-1} - y_{i,t-2}\right) + \left(\mathbf{x}_{it} - \mathbf{x}_{i,t-1}\right)^{\prime} \beta + \left(\epsilon_{it} - \epsilon_{i,t-1}\right) \quad (6)$$

- OLS inconsistent because $y_{i,t-1}$ is correlated with $\epsilon_{i,t-1}$
  - so regressor $(y_{i,t-1} - y_{i,t-2})$ correlated with error $(\epsilon_{it} - \epsilon_{i,t-1})$

First-difference estimator

- First-difference estimator of this model is OLS on

$$
\begin{aligned}
(y_{it} - y_{i,t-1}) - (y_{i,t-1} - y_{i,t-2}) &= \gamma\left[(y_{i,t-1} - y_{i,t-2}) - (y_{i,t-2} - y_{i,t-3})\right] \\
&+ (\epsilon_{it} - \epsilon_{i,t-1}) - (\epsilon_{i,t-1} - \epsilon_{i,t-2})
\end{aligned}
$$

- The x regressors have been omitted for simplicity
    - This D1 estimator is **still inconsistent**
        - as the regressor is correlated with error

Outline

IV estimation

- We have seen the Anderson–Hsiao (AH) IV estimator in the previous section
  - Essentially, use the linear panel transformations
    - and apply a 2SLS estimator using the lag of the regressors as intruments
  - Here, we apply it to the DPD model (5)

$$y_{it} = \gamma y_{i,t-1} + \mathbf{x}_{it}^{'}\beta + \alpha_i + \epsilon_{it}$$

- Consider using $y_{i,t-2}$ as instrument for $y_{i,t-1}$
  - $y_{i,t-2}$ "good" instrument since correlated with $y_{i,t-1}$
    - because of unobserved heterogeneity or true state dependance or both
  - $y_{i,t-2}$ valid instrument only when it is not correlated with the error $\alpha_i + \epsilon_{it}$
    - That is not the case since $y_{i,t-2}$ always contains α_i
    - That is the difference with the previous section
- Thus the AH estimator must be one that removes α_i

AH (IV) estimation of the within model

- Within : transform the DPD model (5)

$$y_{it} = \alpha_i + y_{i,t-1}\gamma + \mathbf{x}_{it}^{'}\beta + \epsilon_{it}$$

using within (I remove the $\mathbf{x}_{it}$ for simplicity):

$$y_{it} - \bar{y}_i = \left(y_{i,t-1} - \bar{y}_i\right)\gamma + \epsilon_{it} - \bar{\epsilon}_i$$

- Clearly, $y_{i,t-2}$, or $y_{i,t-2} - \bar{y}_i$
  - is an invalid instrument in such a model
    - since $\epsilon_{i,t-2}$ from the instrument is present in $\bar{\epsilon}_i$ from the error term

AH (IV) estimation of the first-differences model

- Reconsider the first-difference dynamic model (6)

$$y_{it} - y_{i,t-1} = \gamma \left(y_{i,t-1} - y_{i,t-2} \right) + \left(\mathbf{x}_{it} - \mathbf{x}_{i,t-1} \right)' \beta + \left(\epsilon_{it} - \epsilon_{i,t-1} \right)$$

  - Use $y_{i,t-2}$ as instrument for $(y_{i,t-1} - y_{i,t-2})$
    - $y_{i,t-2}$ valid instrument since not correlated with $(\epsilon_{it} - \epsilon_{i,t-1})$
    - assuming the errors $(\epsilon_{it} - \epsilon_{i,t-1})$ are not (too much) serially correlated
  - $y_{i,t-2}$ "good" instrument since correlated with $(y_{i,t-1} - y_{i,t-2})$
    - $(y_{i,t-2} - y_{i,t-3})$ would also be a good instrument
- Even if the errors $(\epsilon_{it} - \epsilon_{i,t-1})$ are serially correlated (AR(1))
  - using the third & fourth lags may still be a valid strategy

STATA application: command xtivreg, fd

- fd for first difference (D1)
  - The only AH estimator applicable in DPD model
- Arellano-Bond data set
  - loaded from Ch.1 & saved as abdata.dta
    - Year = t, n = log of employment, w = log of real wage, k = log of gross capital, ys = log of industry output, unit = firm index (i)
    - Panel structure : xtset unit year, yearly
- Commands in the ab.do file

Anderson–Hsiao estimator

- xtivreg n l2.n l(0/1).w l(0/2).(k ys) yr1981-yr1984 (l.n = l3.n), fd
  - l2.n (or L2.n) is a direct way to write lag 2 of y, that is $y_{i,t-2}$
  - l(0/1).w "creates" 2 variables
    - w itself (lag "0")
    - lag 1 of w
  - l(0/2).(k ys) "creates" 6 variables
  - yr1981-yr1984 includes 4 time dummies
  - (l.n = l3.n) indicates the list of instruments
    - l.n is an endog. regressor
    - l3.n is its instrument, but is not a regressor
    - Additional endog regr. / instruments are inserted after a space
    - All the regressors not specified here are treated as exogenous/predetermined
- In the output however, the D1 appear
  - that is the first diff Δ caused by the fd option

Estimation results : xtivreg, fd

Variable	Coefficient	(Std. Err.)
LD.n	1.423	(1.583)
L2D.n	-0.165	(0.165)
D.w	-0.752	(0.177)
LD.w	0.963	(1.087)
D.k	0.322	(0.147)
LD.k	-0.325	(0.580)
L2D.k	-0.095	(0.196)
D.ys	0.766	(0.370)
LD.ys	-1.362	(1.157)
L2D.ys	0.321	(0.544)
D.yr1981	-0.057	(0.043)
D.yr1982	-0.088	(0.071)
D.yr1983	-0.106	(0.109)
D.yr1984	-0.117	(0.152)
Intercept	0.016	(0.034)

Outline

More Efficient Estimation of First-differences Model

- Using the IV estimator with only the $y_{i,t-2}$ instrument is a **just-identified** estimator
  - It requires availability of **3** periods of data for each individual
- Using **additional** lags of the dependent variable as instruments
  - is more efficient
  - **Overidentified**, estimation by 2SLS (AH)
  - or possibly 2SGMM to account for non spherical disturbances

Which additional lags ?

- With the AH estimator,
  - the number of instruments is "decided upon"
    - e.g. we use the first 2 lags
- Reconsider the first-difference dynamic model (6)

$$y_{it} - y_{i,t-1} = \gamma \left(y_{i,t-1} - y_{i,t-2} \right) + \left(\mathbf{x}_{it} - \mathbf{x}_{i,t-1} \right)' \beta + \left(\epsilon_{it} - \epsilon_{i,t-1} \right)$$

rewritten as $\Delta y_{it} = \gamma \Delta y_{i,t-1} + \Delta \mathbf{x}_{it}' \beta + \Delta \epsilon_{it}$

- The AH D1 estimator uses $y_{i,t-2}$ as instrument for $\Delta y_{i,t-1}$
  - $y_{i,t-3}$ could also be used, then we **loose** a third period of data
  - The matrix of instruments Z is then

$$\begin{pmatrix} y_{i,t-2} & y_{i,t-3} & \Delta \mathbf{x}_{it} \end{pmatrix}$$

  - There is then a trade-off between the number of instruments and the number of periods

Arellano-Bond view of the instruments

- In period 3 only y_{i1} is available as an instrument,
  - but in period 4 both y_{i1} and y_{i2} would be available,
    - and so on
- Arellano & Bond suggest using **one set of instruments per period**
- To avoid loosing periods of data by introducing further periods
  - Arellano & Bond use zeros outside the period
    - The zeros are not data, but the resulting instruments still satisfy the conditions (uncorrelated to errors, correlated to regressor)
- Each instrument is then relatively weak
  - As the zeros damage the correlation
  - But overall, we use more information in this way
    - So we increase efficiency

Arellano-Bond matrix of instruments

- The resulting $\mathbf{Z}_i$ matrix of instruments :

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_{i3}^{'} & 0 & \cdots & 0 \\ 0 & \mathbf{z}_{i4}^{'} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & 0 & \mathbf{z}_{iT}^{'} \end{bmatrix}$$

where $\mathbf{z}_{it}' = \left[y_{i,t-2}, y_{i,t-3}, \ldots, y_{i1}, \Delta\mathbf{x}_{it}^{'} \right]$

  - Following xtabond help file :
    - Lagged levels of the dependent variable, the predetermined variables, and the endogenous variables are used to form GMM-type instruments.
    - First differences of the strictly exogenous variables are used as standard instruments.
- Lags of $\mathbf{x}_{it}$ or $\Delta\mathbf{x}_{it}$ can additionally be used as instruments
  - The number of instruments increases rapidly with T
    - $(1 \text{ in } t = 3) + (2 \text{ in } t = 4)$ ... 1, 3, 6, 10...

Arellano–Bond Estimator

- The panel GMM **Arellano–Bond AB** estimator is $\hat{\beta}_{AB} =$

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

  where
  - $\tilde{\mathbf{X}}_i$ is a $(T-2)\times(K+1)$ matrix with t^{th} row $\left(\Delta y_{i,t-1}, \Delta\mathbf{x}_{it}'\right), \ T = 3, \ldots, T$
  - $\tilde{\mathbf{y}}_i$ is a $(T-2)\times 1$ vector with t^{th}row Δy_{it}
  - $\mathbf{Z}_i$ is the $(T-2)\times r$ matrix of instruments defined previously
  - Note the use of levels as instruments for Δ
- Arellano and Bond (AB) (Rev. Ec. Stud., 1991)
  - Built upon Holtz-Eakin, Newey and Rosen (Econometrica, 1988)

Arellano–Bond Estimator

- 2SLS and 2SGMM correspond to different weighting matrices W_N
  - depending error cov mtx, see previous Section
- Arellano-Bond is primarily concerned with lagged y
  - To simplify, we presented x as exogenous
  - But some x could be endogenous
  - Then, they would be treated just like the lag of y
    - The Arellano-Bond can accomodate them
- The Arellano–Bond estimator can be interpreted as specifying a system of equations
  - One eq. per time period following the above $\mathbf{Z}$ matrix
    - the instruments applicable to each equation differ
  - It is sometimes called System GMM
    - but that is confusing as other estimators are called that as well

Arellano–Bond estimator in Stata : xtabond

- stat→longitud→DPD→Arellano-Bond
- Uses only the level instruments
  - that is e.g. $y_{i,t-2}$ as instrument for $\Delta y_{i,t-1}$
- requires no autocorrelation of the level errors ϵ_{it}
  - There is a so-called "two-step estimator"
    - corresponding to 2SGMM based on residuals from the 2SLS estimation (cfr earlier formulas)
    - So this should be more efficient than 2SLS
  - Stata is not very explicit on how it calculates this
    - It is in principle designed to address heteroskedasticity and autocorrelation of the RE type
    - It is accessed with the option twostep

Application: Arellano-Bond estimator

- xtabond n w k ys, lags(1) vce(robust) artests(2)
  - We indicate only the exogenous regressors
    - option lags(1) indicates that one lag of the dependant is used as regr.
    - xtabond automatically includes L1.n in the regressors even if lags(1) is not indicated
    - That is, the lags option allows for the specification of more lags
- Output is + informative than xtivreg
  - # instruments, computation of std.err.

the option vce(xxx)

- specifies the type of standard error reported
  - vce(gmm), default, conventionally derived variance estimator for GMM
    - this is based on iid ϵ_{it}
    - when this is not the case, the information content of the data is lower than iid
    - likely resulting in over-estimated asympt. t-stat (the reported z)
  - vce(robust) uses a robust estimator
    - For the one-step estimator, that is 2SLS, this is the Arellano–Bond robust VCE estimator
    - For the two-step estimator, this is Windmeijer's (2005)
    - The vce(gmm) is biased for the 2-step
- Large difference when using robust vce
  - But not much differences between 1-step & 2-step

xtabond : Output

Table: Estimation results : 1-step xtabond

Var.	Coef.	Robust Std.Err.	Default Std.Err.
L.n	0.341	0.125	0.055
w	-0.504	0.157	0.047
k	0.295	0.053	0.028
ys	0.606	0.087	0.050
Intercept	-0.421	0.735	0.304

Table: Estimation results : 2-step xtabond

Variable	Coef.	Robust Std.Err.	Default Std.Err.
L.n	0.304	0.107	0.042
w	-0.450	0.112	0.035
k	0.267	0.056	0.036
ys	0.637	0.084	0.060
Intercept	-0.719	0.557	0.328

Default Std.Err. are biased w/ 2-step GMM

option artests(2)

- estat abond : Arellano–Bond test for serial correlation in the first-diff. residuals
  - (post-estimation) diagnostic tool
  - H_0 : ϵ_{it} are iid
- By construction, $\hat{\Delta}\epsilon_{it}$ **should** be AR(1), so R no AR(1)
  - but if H_0 is true
    - $\hat{\Delta}\epsilon_{it}$ should not exhibit significant AR(2) behavior
    - so not R no AR(2)
  - If a significant AR(2) statistic is encountered (p.val $<$ .05)
    - the second lags of endogenous variables will not be valid instruments for their current values
- To use w/ the one-step estimator
  - Not computed for the 1-step w/ vce(gmm)
  - W/ the 2-step model, the residuals are transformed
- W/ the previous model, p-val AR(1) is .023, AR(2) .575
  - But that is no guarantee that there is no serial correlation

Blundell and Bond: The lagged levels may be poor instruments for first diff variables

- Blundell and Bond (1998) show that the lagged-level instruments in the AB estimator become weak
  - as the autoregressive process becomes too persistent
    - or the ratio of the variance of the panel-level effects α_i to the variance of the idiosyncratic ϵ_{it} becomes too large.
- They proposed a "system-GMM" estimator that uses moment conditions in which
  - lagged differences are used as instruments for the level equation
    - in addition to the moment conditions of lagged levels as instruments for the differenced equation.
  - valid only if the initial condition $E[\alpha_i \Delta y_{i2}] = 0$ holds $\forall i$
  - It's like there are 2 eqs, so "system GMM"
    - xtabond often called difference GMM

Generalizations of xtabond in Stata : xtdpdsys

- The Blundell and Bond is available as xtdpdsys
  - stat→longitud→DPD→Arellano-Bover / Blundell-Bond estimation
- Still requires no autocorrelation in the $\Delta\epsilon_{it}$
  - that is : a 2SLS estimator
    - However, Stata supplies a two-step option as for xtabond
  - This estimator may be quite popular

xtdpdsys : Application

- See the AB.do file
  - Compare the xtabond & xtdpdsys estimators
    - On the same model : n L(0/2).(w k) yr1980-yr1984 year
    - Corresponding to the model in Blundell and Bond (1998)
    - automatically includes L1.n in the regressors as w/ xtabond
- The system estimator
  - produces a higher estimate of the coef. on lagged employment
    - this agrees with the results in Blundell and Bond (1998)
    - who show that the system estimator does not have the downward bias that the Arellano–Bond estimator has when the true value is high.
  - xtdpdsys has 7 more instruments than xtabond
    - Since xtdpdsys includes lagged differences of n as instruments for the level equation

xtdpdsys : Output

Variable	xtabond Coef.	Std.Err.	xtdpdsys Coef.	Std.Err.
L.n	0.629	(0.116)	0.822	(0.093)
w	-0.510	(0.190)	-0.543	(0.188)
L.w	0.289	(0.141)	0.370	(0.166)
L2.w	-0.044	(0.077)	-0.073	(0.091)
k	0.356	(0.060)	0.364	(0.066)
L.k	-0.046	(0.070)	-0.122	(0.070)
L2.k	-0.062	(0.033)	-0.090	(0.034)
yr1980	-0.028	(0.017)	-0.031	(0.017)
yr1981	-0.069	(0.029)	-0.072	(0.029)
yr1982	-0.052	(0.042)	-0.038	(0.037)
yr1983	-0.026	(0.053)	-0.012	(0.050)
yr1984	-0.009	(0.070)	-0.005	(0.066)
year	0.002	(0.012)	0.006	(0.012)
Intercept	-2.543	(23.979)	-10.592	(23.921)

Generalizations of xtabond in STATA : xtdpd

- The xtabond and xtdpdsys estimators still assume uncorrelated ϵ_{it} errors
  - Then the $\Delta\epsilon_{it}$ errors are autocorrelated
  - The one-step default of xtabond and xtdpdsys
    - are 2SLS estimators
    - so, essentially disregards this
  - But allow to estimate a "robust" VCE
    - That might be biased in some cases
  - Their 2SGMM two-step option
    - Allows for some autocorrelation of the $\Delta\epsilon_{it}$ errors
- The xtdpd estimator is a similar estimator
  - that allows for some moving-average (auto)correlation in ϵ_{it}
    - It also has a 2-step option
  - Meant to be used when the "estat abond" rejects absence of autocorrelation of order 2
    - because in this case, the ϵ_{it} errors are not iid
  - stat→longitud→DPD→Linear DPD
  - I do not detail that

Generalization of xtabond in STATA : xtabond2

- David Roodman's contribution
  - install it using findit xtabond2
    - Not available in a menu
  - paper on the course webpage
- Features :
  - ability to specify on how many lags are to be included for the instruments
    - If T is more than 7–8, an unrestricted set of lags will introduce a large number of instruments, with a possible loss of efficiency
  - does not support factor variables
- I also do not detail

This is the end