Chapter 6: More on Linear Panel Data Models

This chapter presents some additional linear panel estimation methods, in particular those involving instrumental variable (IV) estimation and lagged dependent variable:

- IV estimation – Liner Models
  - Basic IV theory,
  - Model setup,
  - Common IV estimators: IV, 2SLS, and GMM,
  - Instrument validity and relevance, etc.

- Panel IV estimation: the `xtivreg` command.

- Hausman-Taylor estimator for FE model.

- Dynamic Panel Data Models: the `xtdpd` command.


6.1. IV Estimation – Linear Models

**OLS estimator.** Consider the multiple linear regression model:

\[ y_i = \alpha + X_i' \beta + u_i = X_i' \beta + u_i, \quad i = 1, \ldots, n, \]

or in matrix form: \( y = X \beta + u \), where \( \text{dim}(\beta) = k \). The ordinary least squares (OLS) estimator of \( \beta \) is

\[ \hat{\beta}_{OLS} = (X'X)^{-1}X'y, \]

which minimizes the sum of squares of errors,

\[ \sum_{i=1}^{n} (y_i - X_i' \beta)^2 = (y - X \beta)'(y - X \beta). \]

The condition for \( \hat{\beta}_{OLS} \) to be valid (unbiased, consistent): (i) \( E(u_i | X_i) = 0 \) (exogeneity of regressors). Under (i), \( E(\hat{\beta}_{OLS} | X) = E((X'X)^{-1}X'y | X) = \beta + E(u | X) = \beta \), implying that \( E(\hat{\beta}_{OLS}) = \beta \) (unbiased).

The conditions for \( \hat{\beta}_{OLS} \) to be efficient: (ii) \( E(u_i^2 | X_i) = \sigma^2 \) (conditional homoskedasticity), and (iii) \( E(u_i u_j | X_i, X_j) = 0, \quad i \neq j \) (conditional zero correlation). Under (ii) and (iii), \( \text{Var}(\hat{\beta}_{OLS}) = \sigma^2 (X'X)^{-1} \) (efficient).
Heteroskedasticity-robust standard errors. If homoskedasticity assumption is violated, i.e., \( \text{E}(u_i^2 | X_i) = \sigma_i^2 \) (heteroskedasticity), the OLS estimator \( \hat{\beta}_{OLS} \) remains valid (unbiased, consistent), but \( \text{Var}(\hat{\beta}_{OLS}) \neq \sigma^2 (X'X)^{-1} \), instead,

\[
\text{Var}(\hat{\beta}_{OLS}) = (X'X)^{-1} \text{diag}(\sigma_i^2) (X'X)^{-1}.
\]

A heteroskedasticity-robust estimator of \( \text{Var}(\hat{\beta}_{OLS}) \) is

\[
\hat{V}_{\text{robust}}(\hat{\beta}_{OLS}) = (X'X)^{-1} \left( \frac{N}{N-k-1} \sum_i u_i^2 X_i X_i' \right) (X'X)^{-1},
\]

where \( u_i \) are OLS residuals, i.e., \( u_i = y_i - X_i' \hat{\beta}_{OLS} \).

Cluster-robust standard errors. If observations possess some cluster or group structure, such that the errors are correlated within a cluster but are uncorrelated across clusters, then a cluster-robust estimator of \( \text{Var}(\hat{\beta}_{OLS}) \) is

\[
\hat{V}_{\text{cluster}}(\hat{\beta}_{OLS}) = (X'X)^{-1} \left( \frac{G}{G-1} \frac{N}{N-k-1} \sum_g X_g \hat{u}_g \hat{u}_g' X_g' \right) (X'X)^{-1},
\]

where \( \hat{u}_g \) is the vector of OLS residuals corresponding to the \( g \)th cluster, and \( X_g \) is the matrix of regressors’ values in the \( g \)th cluster, \( g = 1, \ldots, G, G \to \infty \).
If $E(uu'|X) = \sigma^2 \Omega$, where $\Omega \neq I$, but is a known correlation matrix ((i) and/or (ii) violated), the generalized least-squares (GLS) estimator is:

$$\hat{\beta}_{\text{GLS}} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y,$$

which minimizes the sum of squares: $(y - X\beta)'\Omega^{-1}(y - X\beta)$, and

$$\text{Var}(\hat{\beta}_{\text{GLS}}) = \sigma^2 (X'\Omega^{-1}X)^{-1}.$$

- Both $\hat{\beta}_{\text{OLS}}$ and $\hat{\beta}_{\text{GLS}}$ are unbiased, and consistent.

- But $\hat{\beta}_{\text{GLS}}$ is more efficient than $\hat{\beta}_{\text{OLS}}$, because $\text{Var}(\hat{\beta}_{\text{GLS}}) = (X'\Omega^{-1}X)^{-1}$ is “less than” $\text{Var}(\hat{\beta}_{\text{OLS}}) = \sigma^2 (X'X)^{-1}$.

In case where $\Omega$ is known up to a finite number of parameters $\gamma$, i.e., $\Omega = \Omega(\gamma)$, and if a consistent estimator of $\gamma$ is available, say $\hat{\gamma}$, then a feasible GLS (FGLS) estimator of $\beta$ and its variance are:

$$\hat{\beta}_{\text{FGLS}} = (X'\hat{\Omega}^{-1}X)^{-1}X'\hat{\Omega}^{-1}y, \text{ where } \hat{\Omega} = \Omega(\hat{\gamma});$$

$$\text{Var}(\hat{\beta}_{\text{FGLS}}) = \hat{\sigma}^2 (X'\hat{\Omega}^{-1}X)^{-1}, \text{ where } \hat{\sigma}^2 \text{ is a consistent estimator of } \sigma^2.$$
**IV Estimation: Basic Idea**

The most critical assumption for the validity of the usual linear regression analysis is the exogeneity assumption, $E(u|X) = 0$. Violation of this assumption renders OLS and GLS inconsistent.

**Instrumental variables (IV)** provides a consistent estimator under a strong assumption that valid instruments exists, where the instruments $Z$ are the variables that:

(i) are correlated with the regressors $X$;
(ii) satisfy $E(u|Z) = 0$.

Consider the simplest linear regression model without a intercept:

$$y = x\beta + u,$$

where $y$ measures earnings, $x$ measures years of schooling, and $u$ is the error term.

If this simplest model assumes $x$ is unrelated with $u$, then the only effect of $x$ on $y$ is a direct effect, via the term $x\beta$, as shown below:
In the path diagram, the absence of a direct arrow from \( u \) to \( x \) means that there is no association between \( x \) and \( u \). Then, the OLS estimator \( \hat{\beta} = \frac{\sum_i x_i y_i}{\sum_i x_i^2} \) is consistent for \( \beta \).

The errors \( u \) embodies all factors other than schooling that determine earnings. One such factor is ability, which is likely correlated to \( x \), as high ability tends to lead to high schooling. The OLS estimator \( \hat{\beta} \) is then inconsistent for \( \beta \), as \( \hat{\beta} \) combines the desired direct effect of schooling on earnings (\( \beta \)) with the indirect effect of ability: high \( x \) \( \Rightarrow \) high \( u \) \( \Rightarrow \) high \( y \).

A regressor \( X \) is said to be endogenous, if it arises within a system that influences \( u \). As a consequence \( E(u|X) \neq 0 \);

By contrast, an exogenous regressor arises outside the system and is unrelated to \( u \).
An obvious solution to the endogeneity problem is to include as regressors controls for ability, called *control function approach*. But such regressors may not be available, and even if they do (e.g., IQ scores), there are questions about the extent to which they measure inherent ability.

IV approach provides an alternative solution. Let $z$ be an IV such that the changes in $z$ is associated with the changes in $x$ but do not lead to changes in $y$ (except indirectly via $x$). This leads to the path diagram:

For example, proximity to college ($z$) may determine college attendance ($x$) but not directly determines earnings ($y$).

The IV estimator for this simple example is $\hat{\beta}_{IV} = \sum_i z_i y_i / \sum_i z_i x_i$.

The IV estimator $\hat{\beta}_{IV}$ is consistent for $\beta$ provided the instrument $z$ is unrelated with the error $u$ and correlated with the regressor $x$. 
We now consider the more general regression model with a scalar dependent variable $y_1$ which depends on $m$ endogenous regressors $Y_2$, and $K_1$ exogenous regressors $X_1$ (including an intercept). This model is called a structural equation, with

$$y_{1i} = X_{1i}' \beta_1 + Y_{2i}' \beta_2 + u_i, \quad i = 1, \ldots, n. \quad (6.1)$$

- The $u_i$ are assumed to be uncorrelated with $X_{1i}$, but are correlated with $Y_{2i}$, rendering OLS estimator of $\beta = (\beta_1', \beta_2')'$ inconsistent.
- To obtain a consistent estimator, we assume the existence of at least $m$ IVs, $X_2$, for $Y_2$ that satisfy the condition $E(u_i|X_{2i}) = 0$.
- The instruments $X_2$ need to be correlated with $Y_2$, so that they provide some information on the variables being instrumented. One way to see this is, for each component $y_{2j}$ of $Y_2$,

$$y_{2ji} = X_{1i}' \pi_{1j} + X_{2i}' \pi_{2j} + \varepsilon_{ji}, \quad j = 1, \ldots, m. \quad (6.2)$$
Write the model (6.1) as (the dependent variable is denoted by $y_i$ rather than $y_{1i}$),

$$y_i = X'_i \beta + u_i,$$  \hspace{1cm} (6.3)

where the regressor vector $X'_i = [X'_i, Y'_{2i}]$ combines the exogenous and endogenous variables. Let $X$ be the stacked $X'_i$.

Now, let $Z'_i = [X'_{1i} \ X'_{2i}]$, called collectively the vector of IVs, where $X_{1i}$ serves as the (ideal) instrument for itself, and $X_{2i}$ the instrument for $Y_{2i}$. The instruments satisfy the conditional moment restriction,

$$E(u_i | Z_i) = 0.$$  \hspace{1cm} (6.4)

This implies the following population moment condition:

$$E\{Z_i (y_i - X'_i \beta)\} = 0.$$  \hspace{1cm} (6.5)

Let $Z$ be the stacked $Z'_i$. "We regress $y$ on $X$ using instruments $Z$!"

The IV estimators are the solutions to the sample analogue of (6.5).
**IV Estimators: IV, 2SLS, and GMM**

**Case I:** \( \text{dim}(Z) = \text{dim}(X) \): the number of instruments exactly equals to the number of regressors, called the **just-identified** case.

The sample analogue of (6.5) is

\[
\frac{1}{n} \sum_{i=1}^{n} Z_i (y_i - X_i' \beta) = 0. \tag{6.6}
\]

which can be written in matrix form,

\[
\frac{1}{n} Z'(y - X\beta) = 0.
\]

Solving for \( \beta \) leads to the IV estimator, if \( Z'X \) is invertible:

\[
\hat{\beta}_{\text{IV}} = (Z'X)^{-1}Z'y.
\]

**Case II:** \( \text{dim}(Z) < \text{dim}(X) \), called the **not-identified** case, where there are fewer instruments than regressors.

In this case, no consistent IV estimator exists.
Case III: $\dim(Z) > \dim(X)$, called the over-identified case, where there are more instruments than regressors.

Then, $Z'(y - X\beta) = 0$ has no solution for $\beta$ because it is a system of $\dim(Z)$ equations for $\dim(X)$ unknowns.

One possibility is to arbitrarily drop instruments to get to the just-identified case. But there are more efficient estimators.

One is the two-stage least-squares (2SLS) estimator:

$$\hat{\beta}_{2SLS} = (X'Z(Z'Z)^{-1}Z'X)^{-1}X'Z(Z'Z)^{-1}Z'y.$$ 

which is obtained by running two OLS regressions:

- an OLS regression of $Y_2$ on $Z$ in (6.2) to get the predicted values $\hat{Y}_2 = P_ZY_2$, where $P_Z = Z(Z'Z)^{-1}Z'$, a projection matrix;
- an OLS of (6.1) with $Y_2$ replaced by $\hat{Y}_2$, using $P_ZX_1 = X_1!$ (??)

$\hat{\beta}_{2SLS}$ is the most efficient if $u_i$ are independent and homoscedastic.
Case III: GMM estimator. Minimizing the objective function:
\[
\left[ \frac{1}{n} Z' (y - X\beta) \right]' W \left[ \frac{1}{n} Z' (y - X\beta) \right],
\]
we obtain the generalized method of moments (GMM) estimator:
\[
\hat{\beta}_{\text{GMM}} = \left( X' Z W Z' X \right)^{-1} X' Z W Z' y,
\]
where \( W \) is any full rank symmetric weighting matrix.

- For just identified case, all choices of \( W \) lead to \( \hat{\beta}_{\text{IV}} \).
- Choosing \( W = (Z' Z)^{-1} \) gives \( \hat{\beta}_{\text{2SLS}} \).
- Choosing \( W = \hat{\Omega}^{-1} \), where \( \hat{\Omega} \) is an estimate of \( \text{Var}(n^{-1/2} Z' u) \) leads to the optimal GMM (OGMM) estimator:
\[
\hat{\beta}_{\text{OGMM}} = \left( X' Z \hat{\Omega}^{-1} Z' X \right)^{-1} X' Z \hat{\Omega}^{-1} Z' y.
\]
- If errors are independent, then \( \hat{\Omega} = n^{-1} \sum_{i=1}^{n} \hat{u}_i^2 Z_i Z_i' \), where \( \hat{u}_i = y_i - X_i' \hat{\beta} \), where \( \hat{\beta} \) is a consistent estimator, usually \( \hat{\beta}_{\text{2SLS}} \).
Instrument Validity and Relevance

- For all the IV-type estimators, the instruments must satisfy the condition (6.4), i.e., \( E(u_i|Z_i) = 0 \).
- This condition is impossible to test in the just-identified case.
- And even in the over-identified case, where a test is possible, the validity of instruments relies more on pervasive argument, economic theory, and norms established in prior related empirical studies.
- Instruments must be relevant: they (\( X_2 \)) must account for significant variation in the endogenous variables (\( Y_2 \)), after controlling the exogenous regressors (\( X_1 \)).
- Intuitively, the stronger the association between \( Z \) and \( X \), the stronger will be the identification of the model.
- Conversely, instruments that are only marginally relevant are referred to as **weak instruments**. The consequences of weak instruments: estimation much less precise; inference less reliable.
6.2. Panel IV estimation

IV methods have been extended from the cross-section data to panel data. Two main features for panels remain: estimation still needs to eliminate $\mu_i$ if the FE model is appropriate, and inference needs to control for the clustering inherent in panel data.

Consider the one-way individual effects model:

$$y_{it} = \alpha + X_{it}' \beta + \mu_i + \nu_{it}. \quad (6.7)$$

The FE and first-difference (FD) estimators provide consistent estimates of $\beta$ under a limited form of endogeneity: $X_{it}$ may be correlated with $\mu_i$ but not with $\nu_{it}$.

Now, we consider a richer type of endogeneity: (some) of $X_{it}$ are (also) correlated with $\nu_{it}$.

Assume the existence of instruments $Z_{it}$ that are correlated with $X_{it}$ but uncorrelated with $\nu_{it}$.

The FE panel IV procedure is to suitably transform the model to control for $\mu_i$ and then apply IV to the transformed model.
Panel IV estimation

The `xtivreg` command (extending the `ivregress` for cross-section data) implements 2SLS regression on the transformed model, with options as those for `xtreg`: `fe`, `fd`, `re`, and `be`.

- Note in case of short panels, one can also imbed the time FE in $X_{it}$.
- However, the `xtivreg` command has no `vce(robust)` option, but has the `vce(bootstrap)` option for cluster-robust standard errors.

We use again the Cornwell and Rupert Returns to Schooling Data, with 595 individuals and over 7 years.

The variables are listed on the right.

- `ed` might be correlated with $\mu_i$, and FE model might be appropriate.
- `wks` might be correlated with $v_{it}$, and the panel IV estimation might be more valid.
- `ms` (marital status) might be a suitable IV for `wks`.

\[
\begin{align*}
EXP &= \text{work experience} \\
WKS &= \text{weeks worked} \\
OCC &= \text{occupation, 1 if blue collar,} \\
IND &= \text{1 if manufacturing industry} \\
SOUTH &= \text{1 if resides in south} \\
SMSA &= \text{1 if resides in a city (SMSA)} \\
MS &= \text{1 if married} \\
FEM &= \text{1 if female} \\
UNION &= \text{1 if wage set by union contract} \\
ED &= \text{years of education} \\
BLK &= \text{1 if individual is black} \\
LWAGE &= \text{log of wage}
\end{align*}
\]
### Panel IV FE Estimation

```plaintext
. xtivreg lwage exp expsq (wks = ms), fe
Fixed-effects (within) IV regression
Group variable: id
R-sq:
   within = .
   between = 0.0172
   overall = 0.0284
Wald chi2(3) = 700142.43
corr(u_i, Xb) = -0.8499

|     | Coef.  | Std. Err. |   z  | P>|z| | [95% Conf. Interval] |
|-----|--------|-----------|------|-----|---------------------|
| wks | -.1149742 | .2316926 | -0.50 | 0.620 | -0.5690832 - 0.3391349 |
| exp | .1408101 | .0547014 | 2.57  | 0.010 | 0.0335974 - 0.2480228 |
| expsq | -.0011207 | .0014052 | -0.80 | 0.425 | -0.0038748 - 0.0016334 |
| _cons | 9.83932 | 10.48955 | 0.94  | 0.348 | -10.71983 - 30.39847 |

sigma_u | 1.0980369
sigma_e | .51515503
rho | .81959748 (fraction of variance due to u_i)

F test that all u_i=0: F(594, 3567) = 4.62  Prob > F = 0.0000
```

Number of obs = 4,165
Number of groups = 595
Obs per group:
   min = 7
   avg = 7.0
   max = 7

Panel IV FE Estimation
### Panel FE Estimation

```stata
.xtreg lwage exp expsq wks, fe vce(cluster id)
```

**Fixed-effects (within) regression**

- **Number of obs** = 4,165
- **Number of groups** = 595

**Group variable:** id

**R-sq:**
- within = 0.6566
- between = 0.0276
- overall = 0.0476

|                | Coef. | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|----------------|-------|-----------|-------|-----|----------------------|
| lwage          |       |           |       |     |                      |
| exp            | 0.1138 | 0.0040    | 28.24 | 0.000 | 0.1059 to 0.1217     |
| expsq          | -0.0004 | 0.0001     | -5.16 | 0.000 | -0.0006 to -0.0002   |
| wks            | 0.0008 | 0.0009     | 0.96  | 0.337 | -0.0009 to 0.0025    |
| _cons          | 4.5964 | 0.0601     | 76.49 | 0.000 | 4.4784 to 4.7144     |
| sigma_u        | 1.0362 |           |       |     |                      |
| sigma_e        | 0.1522 |           |       |     |                      |
| rho            | 0.9789 |           |       |     | (fraction of variance due to u_i) |

**Obs per group:**
- **min** = 7
- **avg** = 7.0
- **max** = 7

**F(3, 594)** = 1059.72
**Prob > F** = 0.0000

(Std. Err. adjusted for 595 clusters in id)

Compare the above panel IV FE estimation with panel FE estimation:

**corr(u_i, Xb) =** -0.9107
Panel IV estimation

The panel IV estimates imply that, surprisingly, log-wages decrease by 11.5% for each additional week worked, though the coefficient is statistically insignificant.

Log-wages increases with experience until a peak at 64 years.

Comparing the panel IV results with those using `xtreg, fe`:

• the coefficient of the endogenous regressor `wks` has changed sign, and is many times larger in absolute value;
• The coefficient of the exogenous regressors `exp` and `expsq` are less affected, but the corresponding standard errors are
• For these data, the IV standard errors are more than ten times larger.
• Because the instrument `ms` is not very correlated with `wks`, the panel IV regression leads to a huge loss in estimator efficiency.

In case of weak instruments, alternative methods for better handling of endogeneity in regressors are needed (a topic beyond our scope).
Hausman-Taylor Estimator

In case of limited form of endogeneity (FE), it has been well discussed that the FE and FD methods cannot estimate the coefficients of time-invariant regressors because they are ‘transformed away’.

Hausman-Taylor estimator is an IV estimator for the FE model that enables the estimation of the coefficients of time-invariant regressors.

**Assumption**: some specified regressors are uncorrelated with the FE.

Let $X_{it} = [X_{1it} \ X_{2it}]$ and $W_i = [W_{1i} \ W_{2i}]$ be, respectively, the time-varying and the time-invariant regressors, where the regressors with subscript 1 are uncorrelated with $\mu_i$, and the regressors with subscript 2 are correlated with $\mu_i$. The one-way FE model now has the form:

$$y_{it} = X'_{1it} \beta_1 + X'_{2it} \beta_2 + W'_{1it} \gamma_1 + W'_{2it} \gamma_2 + \mu_i + \nu_{it}, \quad (6.8)$$

which is transformed by a random effects transformation:

$$\tilde{y}_{it} = \tilde{X}'_{1it} \beta_1 + \tilde{X}'_{2it} \beta_2 + \tilde{W}'_{1it} \gamma_1 + \tilde{W}'_{2it} \gamma_2 + \tilde{\mu}_i + \tilde{\nu}_{it}, \quad (6.9)$$

where, e.g., $\tilde{y}_{it} = y_{it} - \hat{\theta}_i \bar{y}_i$. See [XT] xhtaylor for formula for $\hat{\theta}_i$. 
Hausman-Taylor Estimator

The instruments for the regressors in (6.9):

- $\tilde{X}_{1it} = x_{1it} - \overline{x}_{1i}$
- $\tilde{X}_{2it} = x_{2it} - \overline{x}_{2i}$
- $\tilde{W}_{1i} = w_{1i}$
- $\tilde{W}_{2i} = \overline{x}_{1i}$.

The method requires that $\text{dim}(\overline{x}_{1i}) \geq \text{dim}(\tilde{W}_{2i})$.

Consider the Returns to Schooling Data, analyzed by Cornwell and Rupert (1988) using Hausman-Taylor estimator.

- The goal is to obtain a consistent estimate of the coefficient of $ed$ because there is great interest in the impact of education on wage.
- Education is clearly endogenous. It is assumed that it is correlated only with the individual-specific components of the errors.

Cornwell and Rupert (1988) assumed that

- $X_{1it} = [occ, south, smsa, ind]$;
- $X_{2it} = [exp, expsq, wks, ms, union]$.
- $W_{1i} = [fem, blk]$; $W_{2i} = [ed]$. 
Hausman-Taylor Estimator

```
.xtaylor lwage occ south smsa ind exp expsq wks ms union fem blk ed,
> endog(exp expsq wks ms union ed)
```

Hausman-Taylor estimation

| Coef.   | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---------|-----------|------|------|----------------------|
| lwage   |           |      |      |                      |
| TVexogenous |
| occ     | -.0207047 | .0137809 | -1.50 | 0.133 | -.0477149 | .0063055 |
| south   | .0074398  | .031955  | 0.23  | 0.816 | -.0551908 | .0700705  |
| smsa    | -.0418334 | .0189581 | -2.21 | 0.027 | -.0789906 | -.0046761 |
| ind     | .0136039  | .0152374 | 0.89  | 0.372 | -.0162608 | .0434686  |

Number of obs = 4,165
Number of groups = 595
Obs per group:
min = 7
avg = 7
max = 7
Random effects u_i ~ i.i.d.
Wald chi2(12) = 6891.87
Prob > chi2 = 0.0000

Chapter 6
Hausman-Taylor Estimator

Cont’d from last page

| TVendogenous | exp | .1131328 | .002471 | 45.79 | 0.000 | .1082898 | .1179758 |
| | expsq | -.0004189 | .0000546 | -7.67 | 0.000 | -.0005259 | -.0003119 |
| | wks | .0008374 | .0005997 | 1.40 | 0.163 | -.0003381 | .0020129 |
| | ms | -.0298508 | .01898 | -1.57 | 0.116 | -.0670508 | .0073493 |
| | union | .0327714 | .0149084 | 2.20 | 0.028 | .0035514 | .0619914 |
| TIexogenous | fem | -.1309236 | .126659 | -1.03 | 0.301 | -.3791707 | .1173234 |
| | blk | -.2857479 | .1557019 | -1.84 | 0.066 | -.5909179 | .0194221 |
| TIendogenous | ed | .137944 | .0212485 | 6.49 | 0.000 | .0962977 | .1795902 |
| | _cons | 2.912726 | .2836522 | 10.27 | 0.000 | 2.356778 | 3.468674 |
| sigma_u | .94180304 |
| sigma_e | .15180273 |
| rho | .97467788 | (fraction of variance due to u_i) |

Note: TV refers to time varying; TI refers to time invariant.
### Hausman-Taylor Estimator

Compared with the RE estimation given next:

- the coefficient of \( \text{ed} \) and standard errors have both increased,
- \( \text{ed} \) remains highly significant.

```plaintext
. xtreg lwage occ south smsa ind exp expsq wks ms union fem blk ed, re vce(robust)
```

Random-effects GLS regression

<table>
<thead>
<tr>
<th>Number of obs = 4,165</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group variable: id</td>
</tr>
<tr>
<td>Number of groups = 595</td>
</tr>
</tbody>
</table>

R-sq:

<table>
<thead>
<tr>
<th>within = 0.6124</th>
</tr>
</thead>
<tbody>
<tr>
<td>between = 0.2539</td>
</tr>
<tr>
<td>overall = 0.2512</td>
</tr>
</tbody>
</table>

Obs per group:

<table>
<thead>
<tr>
<th>min = 7</th>
</tr>
</thead>
<tbody>
<tr>
<td>avg = 7.0</td>
</tr>
<tr>
<td>max = 7</td>
</tr>
</tbody>
</table>

Wald chi2(12) = 1528.82

Prob > chi2 = 0.0000

Continued on next page …
### Hausman-Taylor Estimator

(Std. Err. adjusted for 595 clusters in id)

|       | Coef.  | Std. Err. | z     | P>|z|  | [95% Conf. Interval] |
|-------|--------|-----------|-------|------|----------------------|
| lwage |        |           |       |      |                      |
| occ   | -0.0500664 | 0.0207608 | -2.41 | 0.016 | -0.0907567 to -0.009376 |
| south | -0.0166176 | 0.0460333 | -0.36 | 0.718 | -0.1068411 to 0.073606  |
| smsa  | -0.0138231 | 0.0297818 | -0.46 | 0.643 | -0.0721944 to 0.0445482 |
| ind   | 0.0037441  | 0.0232112 | 0.16  | 0.872 | -0.0417489 to 0.0492372 |
| exp   | 0.0820544  | 0.0040168 | 20.43 | 0.000 | 0.0741816 to 0.0899272  |
| expsq | -0.0008084 | 0.0000895 | -9.03 | 0.000 | -0.000984 to -0.0006329 |
| wks   | 0.0010347  | 0.000941  | 1.10  | 0.272 | -0.0008097 to 0.0028791 |
| ms    | -0.0746283 | 0.0274262 | -2.72 | 0.007 | -0.1283826 to -0.020874 |
| union | 0.0632232  | 0.0249276 | 2.54  | 0.011 | 0.0143661 to 0.1120803  |
| fem   | -0.3392101 | 0.0630206 | -5.38 | 0.000 | -0.4627283 to -0.2156919 |
| blk   | -0.2102803 | 0.0826686 | -2.54 | 0.011 | -0.3723078 to -0.0482528 |
| ed    | 0.0996585  | 0.0080237 | 12.42 | 0.000 | 0.0839324 to 0.1153847  |
| _cons | 4.26367    | 1.359373  | 31.36 | 0.000 | 3.997238 to 4.530103    |

|       |         |           |       |      |                      |
| sigma_u | 0.26265814 |
| sigma_e | 0.15199444 |
| rho     | 0.74913774  | (fraction of variance due to u_i) |
6.3. Dynamic Panel Data Models

Extending Model (6.7) by including time-lags of the dependent variable as regressors, we have a dynamic panel data model with individual-specific effects, referred to as DPD model:

\[
y_{it} = \gamma_1 y_{i,t-1} + \ldots + \gamma_p y_{i,t-p} + X'_{it} \beta + \mu_i + v_{it}, \quad t = p + 1, \ldots, T,
\]

where \(\gamma_1, \ldots, \gamma_p\) are called the dynamic parameter, and the goal is to consistently estimate \(\gamma_1, \ldots, \gamma_p\) and \(\beta\) when \(\{\mu_i\}\) are fixed effects.

In this model, the response is \(y\) affected

- directly by its own proceeding (lagged) values – dynamic impacts or state dependence;
- directly through the observables \(X\) – observed heterogeneity;
- indirectly through time-invariant individual effect – unobserved heterogeneity.

Being able to control dynamic effects and unobserved heterogeneity may be the biggest advantages of using the panel data.
Arellano-Bond Estimator -- *xtabond*

Consider the case where $\mu_i$ are fixed effects. Assume $v_{it}$ are independent over $i$ and serially uncorrelated over $t$.

**Arellano-Bond estimator** is the IV estimation of the first-differenced (FD) model, using appropriate lags of regressors as instruments, which is implemented in STATA by the command *xtabond* with the syntax:

```
xtabond  depvar  [ indepvar ]  [ if ]  [ in ]  [ ,  options ]
```

- **Strictly exogenous regressors** are uncorrelated with $v_{it}$, and are entered as *indepvar*.
- **Predetermined or weakly exogenous regressors** are correlated with the past errors but are uncorrelated with the current and future errors: $E(X_{it}v_{is}) \neq 0$ for $s < t$, and $E(X_{it}v_{is}) = 0$ for $s \geq t$. These regressors are entered by using *pre* (*varlist*) option.
- **Contemporaneously endogenous regressors**: $E(X_{it}v_{is}) \neq 0$, for $s \leq t$, and $E(X_{it}v_{is}) = 0$ for $s > t$. These regressors are entered by using the *endogenous* (*varlist*) option.
Arellano-Bond Estimator -- \textit{xtabond}

The option \texttt{vce (robust)} provides a heteroskedasticity-consistent estimate of the VC matrix, but not cluster-robust;

Consider a pure dynamic model with FE for \texttt{lwage} in Return to Schooling Example, after FD: $\Delta y_{it} = \alpha + \gamma_1 \Delta y_{i,t-1} + \gamma_2 \Delta y_{i,t-2} + \Delta v_{it}, \quad t = 4, 5, 6, 7$.

. * 2SLS or one-step GMM for a pure time-series AR(2) panel data model

. \texttt{xtabond lwage, lags(2) vce(robust)}

Arellano-Bond dynamic panel-data estimation

\begin{verbatim}
Number of obs = 2,380
Group variable: id
Number of groups = 595
Time variable: year

There all together 14 instruments for the two lagged dependent variables: 2 at t=4, 3 at t=5, 4 at t=6, and 5 at t=7.

Number of instruments = 15

Wald chi2(2) = 1253.03
Prob > chi2 = 0.0000

Continued on next page . . .
\end{verbatim}
### Arellano-Bond Estimator -- xtabond

Number of instruments = 15

|                  | Coef.   | Robust Std. Err. | z    | P>|z|    | [95% Conf. Interval] |
|------------------|---------|------------------|------|--------|---------------------|
| lwage            |         |                  |      |        |                     |
| L1.              | .5707517| .0333941         | 17.09| 0.000  | .5053005            | .6362029           |
| L2.              | .2675649| .0242641         | 11.03| 0.000  | .2200082            | .3151216           |
| _cons            | 1.203588| .164496          | 7.32 | 0.000  | .8811814            | 1.525994           |

Instruments for differenced equation

GMM-type: L(2/).lwage

Instruments for level equation

Standard: _cons

There are $4 \times 595 = 2380$ observations because the **first three years of data are lost** in order to construct $\Delta y_{i,t-2}$.

Results are reported in levels although the FD model is fitted.
Arellano-Bond Estimator -- `xtabond`

* Optimal or two-step GMM for a pure time-series AR(2) panel data model

```
. xtabond lwage, lags(2) twostep vce(robust)
```

Arellano-Bond dynamic panel-data estimation

<table>
<thead>
<tr>
<th></th>
<th>WC-Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>l wage</strong></td>
<td></td>
</tr>
<tr>
<td><strong>Coef.</strong></td>
<td><strong>Std. Err.</strong></td>
</tr>
<tr>
<td>L1.</td>
<td>.6095931</td>
</tr>
<tr>
<td>L2.</td>
<td>.2708335</td>
</tr>
<tr>
<td>_cons</td>
<td>.9182262</td>
</tr>
</tbody>
</table>
Arellano-Bond Estimator -- xtabond

. * Reduce no. of instruments for a pure time-series AR(2) panel data model
. xtabond lwage, lags(2) vce(robust) maxldep(1) <- Use only the 1st lag as IVs.

Arellano-Bond dynamic panel-data estimation
Number of obs = 2,380
Group variable: id
Number of groups = 595
Time variable: year
Obs per group:
  min = 4
  avg = 4
  max = 4

Number of instruments = 5
Wald chi2(2) = 1372.33
Prob > chi2 = 0.0000

One-step results
(Std. Err. adjusted for clustering on id)

|        | Coef.     | Std. Err. |      z   | P>|z|   | [95% Conf. Interval] |
|--------|-----------|-----------|----------|-------|----------------------|
| lwage  |           |           |          |       |                      |
| L1.    | .4863642  | .1919353  | 2.53     | 0.011 | .110178 .8625505    |
| L2.    | .3647456  | .1661008  | 2.20     | 0.028 | .039194 .6902973    |
| _cons  | 1.127609  | .2429357  | 4.64     | 0.000 | .6514633 1.603754   |
Arellano-Bond Estimator -- *xtabond*

**Optimal or two-step GMM for a dynamic panel data model**

*xtabond lwage occ south smsa ind, lags(2) maxldep(3) pre(wks, lag(1,2))
endogenous(ms, lag(0,2)) endogenous(union, lag(0,2)) twostep vce(robust)*

Arellano-Bond dynamic panel-data estimation

<table>
<thead>
<tr>
<th></th>
<th>WC-Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td>lwage</td>
<td>Coef.</td>
</tr>
<tr>
<td>lwage</td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.611753</td>
</tr>
<tr>
<td>L2.</td>
<td>.2409058</td>
</tr>
</tbody>
</table>
Arellano-Bond Estimator -- *xtabond*

<table>
<thead>
<tr>
<th>wks</th>
<th>-.0159751</th>
<th>.0082523</th>
<th>-1.94</th>
<th>0.053</th>
<th>-.0321493</th>
<th>.000199</th>
</tr>
</thead>
<tbody>
<tr>
<td>L1</td>
<td>.0039944</td>
<td>.0027425</td>
<td>1.46</td>
<td>0.145</td>
<td>-.0013807</td>
<td>.0093695</td>
</tr>
<tr>
<td>ms</td>
<td>.1859324</td>
<td>.144458</td>
<td>1.29</td>
<td>0.198</td>
<td>-.0013807</td>
<td>.4690649</td>
</tr>
<tr>
<td>union</td>
<td>-.1531329</td>
<td>.1677842</td>
<td>-0.91</td>
<td>0.361</td>
<td>-.4819839</td>
<td>.1757181</td>
</tr>
<tr>
<td>occ</td>
<td>-.0357509</td>
<td>.0347705</td>
<td>-1.03</td>
<td>0.304</td>
<td>-.1038999</td>
<td>.032398</td>
</tr>
<tr>
<td>south</td>
<td>-.0250368</td>
<td>.2150806</td>
<td>-0.12</td>
<td>0.907</td>
<td>-.446587</td>
<td>.3965134</td>
</tr>
<tr>
<td>smsa</td>
<td>-.0848223</td>
<td>.0525243</td>
<td>-1.61</td>
<td>0.106</td>
<td>-.187768</td>
<td>.0181235</td>
</tr>
<tr>
<td>ind</td>
<td>.0227008</td>
<td>.0424207</td>
<td>0.54</td>
<td>0.593</td>
<td>-.0604422</td>
<td>.1058437</td>
</tr>
<tr>
<td>_cons</td>
<td>1.639999</td>
<td>.4981019</td>
<td>3.29</td>
<td>0.001</td>
<td>.6637377</td>
<td>2.616261</td>
</tr>
</tbody>
</table>

Instruments for differenced equation
GMM-type: L(2/4).lwage L(1/2).L.wks L(2/3).ms L(2/3).union
Standard: D.occ D.south D.smsa D.ind

Instruments for level equation
Standard: _cons

- With the inclusion of additional regressors, dynamic effects change little, and the standard errors are 10-15% higher;
- The additional regressors are all statistically insignificant at 5%;
- By contrast, some are statistically significant using the within estimation of a static panel data model without including the lagged dependent variables.
A General Method for Dynamic Panel Data Model

• The proceeding method depends builds on the assumptions that the errors $v_{it}$ are serially uncorrelated.

• In case this is violated, an alternative is to use the `xtdpd` command, allowing $v_{it}$ to follow a moving average (MA) process of low order.

```
. xtdpd L(0/2).lwage L(0/1).wks occ south smsa ind ms union, div(occ south smsa > ind) dgmmiv(lwage, lagrange(2 4)) dgmmiv(ms union, lagrange(2 3))
> dgmmiv(L.wks, lagrange(1 2)) lgmmiv(lwage wks ms union)
> twostep vce(robust) artests(3)
```

Dynamic panel-data estimation

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of obs</td>
<td>2,975</td>
</tr>
<tr>
<td>Number of groups</td>
<td>595</td>
</tr>
<tr>
<td>Obs per group:</td>
<td></td>
</tr>
<tr>
<td>min</td>
<td>5</td>
</tr>
<tr>
<td>avg</td>
<td>5</td>
</tr>
<tr>
<td>max</td>
<td>5</td>
</tr>
<tr>
<td>Number of instruments</td>
<td>60</td>
</tr>
<tr>
<td>Wald chi2(10)</td>
<td>2270.88</td>
</tr>
<tr>
<td>Prob &gt; chi2</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Two-step results

(Std. Err. adjusted for clustering on id)
A General Method for Dynamic Panel Data Model

<table>
<thead>
<tr>
<th></th>
<th>WC-Robust</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coef.</td>
</tr>
<tr>
<td>lwage</td>
<td></td>
</tr>
<tr>
<td>L1.</td>
<td>.6017533</td>
</tr>
<tr>
<td>L2.</td>
<td>.2880537</td>
</tr>
<tr>
<td>wks</td>
<td></td>
</tr>
<tr>
<td>--.</td>
<td>-.0014979</td>
</tr>
<tr>
<td>L1.</td>
<td>.0006786</td>
</tr>
<tr>
<td>occ</td>
<td>-.0508803</td>
</tr>
<tr>
<td>south</td>
<td>-.1062817</td>
</tr>
<tr>
<td>smsa</td>
<td>-.0483567</td>
</tr>
<tr>
<td>ind</td>
<td>.0144749</td>
</tr>
<tr>
<td>ms</td>
<td>.0395337</td>
</tr>
<tr>
<td>union</td>
<td>-.0422409</td>
</tr>
<tr>
<td>_cons</td>
<td>.9584113</td>
</tr>
</tbody>
</table>

Instruments for differenced equation
GMM-type: L(2/4).lwage L(2/3).ms L(2/3).union L(1/2).L.wks
Standard: D.occ D.south D.smsa D.ind

Instruments for level equation
GMM-type: LD.lwage LD.wks LD.ms LD.union
Standard: _cons
## A General Method for Dynamic Panel Data Model

|               | WC–Robust Coef. | Std. Err. | z    | P>|z| | [95% Conf. Interval] |
|---------------|-----------------|-----------|------|--------|----------------------|
| lwage         |                 |           |      |        |                      |
| L1.           | 0.6017533       | 0.0291502 | 20.64| 0.000  | 0.5446199            | 0.6588866            |
| L2.           | 0.2880537       | 0.0285319 | 10.10| 0.000  | 0.2321322            | 0.3439752            |
| wks           |                 |           |      |        |                      |
| --.           | -0.0014979      | 0.0056143 | -0.27| 0.790  | -0.0125017           | 0.009506             |
| L1.           | 0.0006786       | 0.0015694 | 0.43 | 0.665  | -0.0023973           | 0.0037545            |
| occ           | -0.0508803      | 0.0331149 | -1.54| 0.124  | -0.1157843           | 0.0140237            |
| south         | -0.1062817      | 0.083753  | -1.27| 0.204  | -0.2704346           | 0.0578713            |
| smsa          | -0.0483567      | 0.0479016 | -1.01| 0.313  | -0.1422422           | 0.0455288            |
| ind           | 0.0144749       | 0.031448  | 0.46 | 0.645  | -0.0471621           | 0.0761118            |
| ms            | 0.0395337       | 0.0558543 | 0.71 | 0.479  | -0.0699386           | 0.1490061            |
| union         | -0.0422409      | 0.0719919 | -0.59| 0.557  | -0.1833423           | 0.0988606            |
| _cons         | 0.9584113       | 0.3632287 | 2.64 | 0.008  | 0.2464961            | 1.670327             |

Instruments for differenced equation

- GMM-type: L(2/4).lwage L(2/3).ms L(2/3).union L(1/2).L.wks
- Standard: D.occ D.south D.smsa D.ind

Instruments for level equation

- GMM-type: LD.lwage LD.wks LD.ms LD.union
- Standard: _cons
A General Method for Dynamic Panel Data Model

If the error $v_{it}$ is correlated as MA(1), we need to change the `dgmmiv()` and `lgmmiv()` options for `lwage`. The command becomes:

```
xtdpd L(0/2).lwage L(0/1).wks occ south smsa ind ms union,
> div(occ south smsa ind) dgmmiv(lwage, lagrange(3 4))
> dgmmiv(ms union, lagrange(2 3)) dgmmiv(L.wks, lagrange(1 2))
> lgmmiv(L.lwage wks ms union) twostep vce(robust) artests(3)
```

See [XT] xtdpd for details; or type:
```
help xtdpd
```

<table>
<thead>
<tr>
<th>lwage</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>L1.</td>
<td>0.8508834</td>
<td>0.0950866</td>
<td>8.95</td>
<td>0.000</td>
<td>0.6645171</td>
<td>1.03725</td>
</tr>
<tr>
<td>L2.</td>
<td>0.0497</td>
<td>0.0835395</td>
<td>0.59</td>
<td>0.552</td>
<td>-0.1140345</td>
<td>0.2134345</td>
</tr>
<tr>
<td>wks</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>--.</td>
<td>-0.0011449</td>
<td>0.0047492</td>
<td>-0.24</td>
<td>0.810</td>
<td>-0.010453</td>
<td>0.0081633</td>
</tr>
<tr>
<td>L1.</td>
<td>0.0001078</td>
<td>0.0014174</td>
<td>0.08</td>
<td>0.939</td>
<td>-0.0026703</td>
<td>0.0028859</td>
</tr>
<tr>
<td>occ</td>
<td>-0.0495623</td>
<td>0.0313022</td>
<td>-1.58</td>
<td>0.113</td>
<td>-0.1109136</td>
<td>0.0117889</td>
</tr>
<tr>
<td>south</td>
<td>-0.1493551</td>
<td>0.0747589</td>
<td>-2.00</td>
<td>0.046</td>
<td>-0.2958797</td>
<td>-0.0028304</td>
</tr>
<tr>
<td>smsa</td>
<td>-0.0646732</td>
<td>0.0526654</td>
<td>-1.23</td>
<td>0.219</td>
<td>-0.1678956</td>
<td>0.0385491</td>
</tr>
<tr>
<td>ind</td>
<td>0.0136841</td>
<td>0.0354739</td>
<td>0.39</td>
<td>0.700</td>
<td>-0.0558435</td>
<td>0.0832117</td>
</tr>
<tr>
<td>ms</td>
<td>0.0404826</td>
<td>0.0482916</td>
<td>0.84</td>
<td>0.402</td>
<td>-0.0541672</td>
<td>0.1351323</td>
</tr>
<tr>
<td>union</td>
<td>-0.0256618</td>
<td>0.070258</td>
<td>-0.37</td>
<td>0.715</td>
<td>-0.1633649</td>
<td>0.1120413</td>
</tr>
<tr>
<td>_cons</td>
<td>0.8903751</td>
<td>0.3448392</td>
<td>2.58</td>
<td>0.010</td>
<td>0.2145028</td>
<td>1.566247</td>
</tr>
</tbody>
</table>