

# Panel data

Take full advantage of the extra information that panel data provide while simultaneously handling the peculiarities of panel data.

Study the time-invariant features within each panel, the relationships across panels, and how outcomes of interest change over time.

Fit linear models or nonlinear models for binary, count, ordinal, censored, or survival outcomes with fixed-effects, random-effects, or population-averaged estimators. Even perform Bayesian estimation.

Use difference in differences (DID) to estimate treatment effects.

Fit dynamic models or models with endogeneity.

And much more.

$$x_{it} \beta \epsilon_{it} \alpha_i y_{it} \sigma^2 y_{i,t-1}$$

## Linear panel-data models

- Fixed effects, random effects, correlated random effects **New**, and population averaged
- High-dimensional fixed effects (HDFE) **New**
- Fixed and random effects with AR(1) disturbances
- Random coefficients
- Multiple levels of random effects

## Nonlinear panel-data models

- Probit
- Logit
- Ordered logit and probit
- Multinomial logit
- Mixed logit
- Interval regression
- Tobit
- Poisson
- Negative binomial
- Random coefficients
- Multiple levels of random effects

## Dynamic panel-data estimators

- Arellano–Bond
- Arellano–Bover
- Blundell–Bond
- Panel-data vector autoregression **New**
- Build your own dynamic model

## Parametric survival models

- Weibull, exponential, lognormal, loglogistic, or gamma

## Instrumental-variables models

- Fixed effects
- Random effects
- First-differenced
- Between effects

## Extended regression models

- Combine endogenous covariates, sample selection, and treatment effects
- Linear, probit, ordered probit, and interval regression
- Random effects

## Test and diagnostics

- Panel-data unit-root tests
- Panel-data cointegration tests
- Hausman test
- Mundlak test **New**
- Overidentification and autocorrelation tests
- Breusch and Pagan Lagrange multiplier test for random effects

## Graphic and tabular analysis

- Summary statistics and tabulations within and between panels
- Patterns of panel participation
- Graphs of marginal effects, elasticities, treatment effects, and marginal means

## Bayesian estimation

Before fitting any panel-data model in Stata, we specify the panel and time identifiers, in this case the variables **id** and **year**.

```
. xtset id year
```

Now we are ready to fit a model. Let's start by fitting a random-effects linear regression model for **y** on **x1**, **x2**, and **x3** and storing the results.

Viewer - view xt1.smcl

view xt1.smcl X

Dialog ▾ Also see ▾ Jump to ▾

+  
. xtreg y x1 x2 x3, re

Random-effects GLS regression  
Group variable: id

Number of obs = 18,760  
Number of groups = 2,345

R-squared:  
    Within = 0.6354  
    Between = 0.5341  
    Overall = 0.5604

Obs per group:  
    min = 8  
    avg = 8.0  
    max = 8

corr(u\_i, X) = 0 (assumed)

Wald chi2(3) = 30754.84  
Prob > chi2 = 0.0000

	y	Coefficient	Std. err.	z	P> z	[95% conf. interval]
x1		1.819258	.023037	78.97	0.000	1.774106 1.86441
x2		-.9579329	.005927	-161.62	0.000	-.9695497 -.9463162
x3		.9732638	.0227792	42.73	0.000	.9286175 1.01791
_cons		1.046447	.0292991	35.72	0.000	.9890223 1.103872

sigma_u	1.2557466
sigma_e	1.0487969
rho	.58908237 (fraction of variance due to u_i)

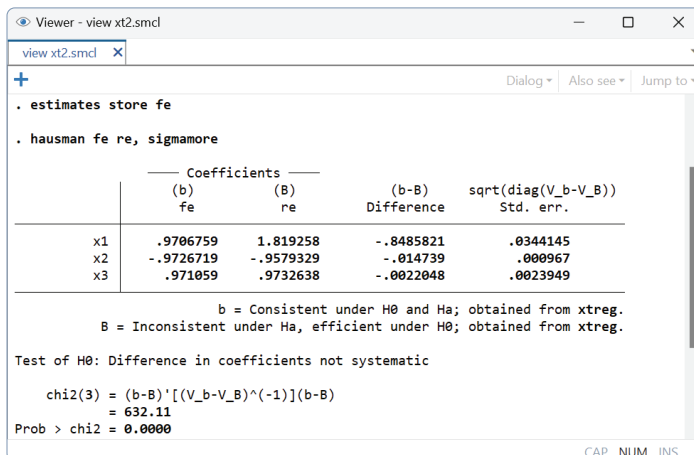
. estimates store re

CAP NUM INS

Fitting a fixed-effects model is just as easy. We can type

```
. xtreg y x1 x2 x3, fe
```

We can now store the results of this model and compare the models using a Hausman test.



Viewer - view xt2.smcl

```
view xt2.smcl X
```

Dialog ▾ Also see ▾ Jump to ▾

```
. estimates store fe
```

```
. hausman fe re, sigmamore
```

	Coefficients		(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
	(b) fe	(B) re		
x1	.9706759	1.819258	-.8485821	.0344145
x2	-.9726719	-.9579329	-.014739	.000967
x3	.971059	.9732638	-.0022048	.0023949

b = Consistent under H0 and Ha; obtained from xtreg.  
B = Inconsistent under Ha, efficient under H0; obtained from xtreg.

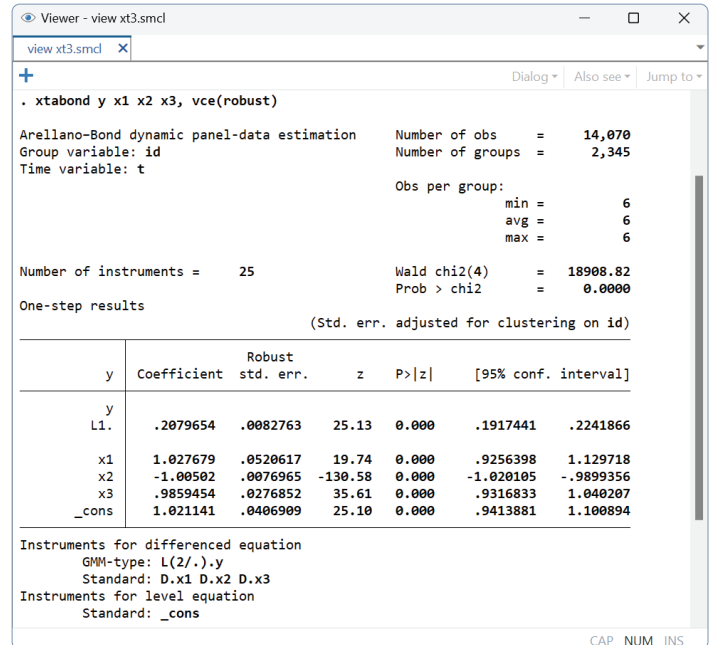
Test of H0: Difference in coefficients not systematic

chi2(3) = (b-B)'[(V\_b-V\_B)^(-1)](b-B)  
= 632.11  
Prob > chi2 = 0.0000

CAP NUM INS

If the model is correctly specified, the Hausman test indicates that time-invariant unobservables are not modeled correctly using random effects.

Commands for more complex models are just as straightforward. For instance, let's fit a dynamic panel-data model using the Arellano–Bond estimator.



Viewer - view xt3.smcl

```
view xt3.smcl X
```

Dialog ▾ Also see ▾ Jump to ▾

```
. xtabond y x1 x2 x3, vce(robust)
```

Arellano-Bond dynamic panel-data estimation      Number of obs = 14,070  
Group variable: id                                  Number of groups = 2,345  
Time variable: t

Obs per group:                                      min = 6  
    avg = 6  
    max = 6

Number of instruments = 25                          Wald chi2(4) = 18908.82  
    Prob > chi2 = 0.0000

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

	y	Coefficient	Robust std. err.	z	P> z	[95% conf. interval]
y	L1.	.2079654	.0082763	25.13	0.000	.1917441 .2241866
x1		1.027679	.0520617	19.74	0.000	.9256398 1.129718
x2		-1.00502	.0076965	-130.58	0.000	-1.020105 -.9899356
x3		.9859454	.0276852	35.61	0.000	.9316833 1.040207
_cons		1.021141	.0406909	25.10	0.000	.9413881 1.100894

Instruments for differenced equation  
GMM-type: L(2/.)y  
Standard: D.x1 D.x2 D.x3  
Instruments for level equation  
Standard: \_cons

CAP NUM INS

This is just the beginning.

*Do you have a binary outcome?*

You can fit a random-effects probit model.

```
. xtprobit y x1 x2 x3
```

*Do you have a count outcome?*

You can fit a conditional fixed-effects Poisson model.

```
. xtpoisson y x1 x2 x3, fe
```

*Do you have prior information or want to make probability statements about the results?*

You can perform Bayesian estimation.

```
. bayes: xtreg y x1 x2 x3
```

*Do you have random coefficients for x1?*

You can fit a mixed-effects model.

```
. mixed y x1 x2 x3 || id: x1
```

*Do you have a panel of individuals nested within companies?*

You can fit a three-level random-effects model.

```
. mixed y x1 x2 x3 || company: || id:
```

*Do you want to estimate the average treatment effect on the treated?* You can use DID estimation.

```
. xtddidregress (y x) (treatment),  
                                    group(company) time(t)
```

And so much more.