Introduction to Panel-Data Analysis using Stata

Gustavo Sánchez

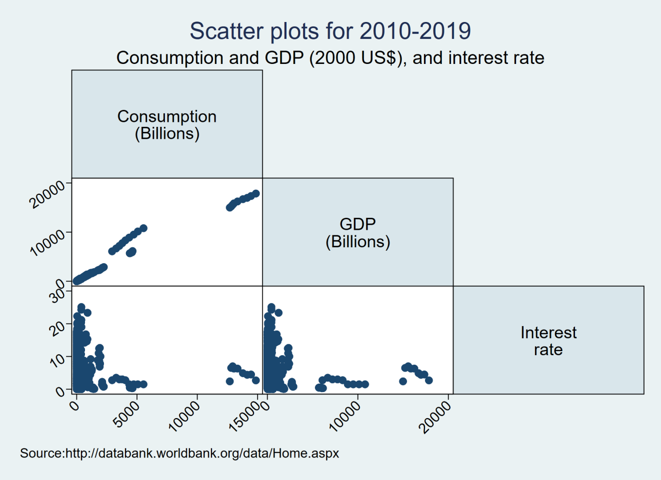
StataCorp LLC

July 28, 2022
StataCorp LLC

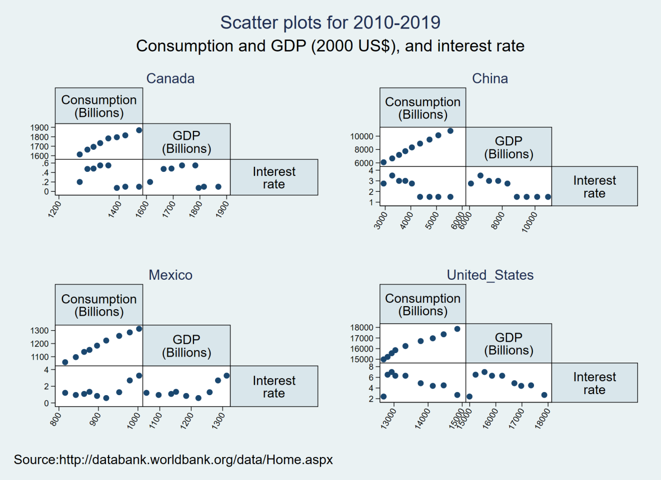
Outline

- Basic Concepts
  - Pooled vs. Panel
  - Stata tools
  - Data management
- Linear models for panel data
  - Data generating process
  - Random effects
  - Fixed effects
  - Fixed or random effects
  - Marginal analysis
- Dynamic panel-data linear models
  - Arellano-Bond
  - Arellano-Bover/Blundell-Bond
- Extended regression models for panel data
- Panel unit roots and cointegration

Pooled vs. Panel



Pooled vs. **Panel**

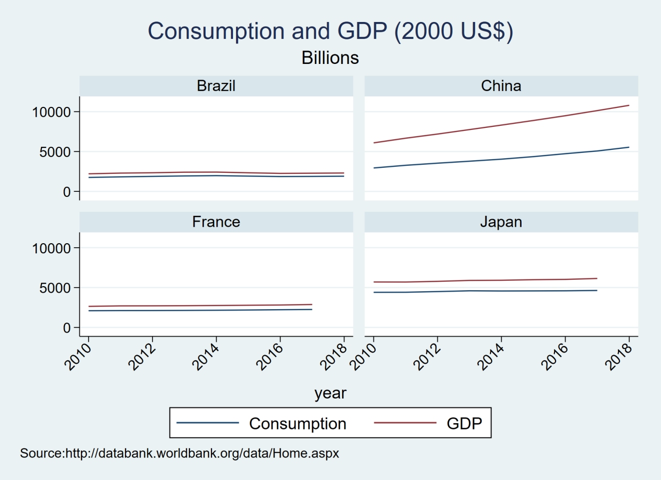


How it Looks

```
. list country year consumption gdp irate ///
>         if CountryName=="Mexico" |      ///
>         CountryName=="United States",   ///
>         sepby(country) abbreviate(12) noobs
```

country	year	consumption	gdp	irate
Mexico	2010	815.78416	1057.8013	1.2125
Mexico	2011	842.78459	1096.5486	.95583333
Mexico	2012	863.83937	1136.4885	1.0816667
Mexico	2013	877.37001	1151.8776	1.3316667
Mexico	2014	896.49265	1184.1801	.84
Mexico	2015	919.70082	1223.1159	.58916667
Mexico	2016	952.52868	1258.7152	1.2875
Mexico	2017	979.32119	1285.3759	2.6958333
Mexico	2018	1002.938	1312.831	3.2708333
United States	2010	12695.979	14992.053	2.4000001
United States	2011	12812.144	15224.555	6.5
United States	2012	12932.334	15567.038	7
United States	2013	13039.38	15853.796	6.3000002
United States	2014	13336.045	16242.526	6.3000002
United States	2015	13783.285	16710.459	4.9000001
United States	2016	14137.888	16972.348	4.4000001
United States	2017	14456.84	17348.627	4.5
United States	2018	14857.51	17856.477	2.7

How it Looks


Consumption and GDP (2000 US$)
Billions
Brazil
China
France
Japan
10000
5000
0
2010
2012
2014
2016
2018
year
Consumption
GDP
Source:http://databank.worldbank.org/data/Home.aspx

Stata Tools

- **Data management**
- **Linear regression estimators**
- **Dynamic panel-data estimators**
- Binary-outcome estimators
- Ordinal-outcome estimators
- Count-data estimators
- Survival-time estimators
- **Extended regression models**
- **Unit-root and cointegration tests**

Data management

- `reshape`: convert data (wide <–> long).
- `xtsum`: summarize `xt` (panel) data.
- Tabulate one-way generalization for `xt` (panel) data.
  - `xttab`: Counts decomposition between-within components.
  - `xttrans`: Transition probabilities report.

LINEAR PANEL-DATA MODELS

Data Generating Process

- The data generating process is given by:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \eta_{it} \\ \eta_{it} &\equiv \alpha_i + \varepsilon_{it} \\ i &= 1, \ldots n \\ t &= 1, \ldots T \end{aligned}$$

- We have redefined the nature of the random disturbance to include an unobservable component
  - The unobservable component is particular to each panel and is independent of time (e.g. for individuals: ability, intelligence, work ethic)
  - As in the regression case the assumptions made on η_{it}, with particular emphasis on α_i, define the models we work with.

Model for aggregate consumption

$$consumption_{it} = \alpha + gdp_{it} * \beta_1 + irate_{it} * \beta_3 + \mu_i + \nu_{it}$$

Data

- World Bank public online data on:
  consumption: Final consumption expenditure (2010 US$)
  gdp: Gross domestic product (2010 US$)
  irate: deposit interest rate
- Example : 2010-2018 for 131 countries
- Source:http://databank.worldbank.org/data/Home.aspx

Specifying the panel structure in Stata

Assuming that the second dimension corresponds to time series, we use the -xtset- command to specify the panel structure with:

- Panel identifier variable (e.g. country)
- Time identifier variable (e.g. year)

```
. xtset country year

       panel variable:  country (unbalanced)
        time variable:  year, 2010 to 2018, but with a gap
                delta:  1 unit
```

Specifying the panel structure in Stata

Assuming that the second dimension corresponds to time series, we use the -xtset- command to specify the panel structure with:

- Panel identifier variable (e.g. country)
- Time identifier variable (e.g. year)

```
. xtset country year

       panel variable:  country (unbalanced)
        time variable:  year, 2010 to 2018, but with a gap
                delta:  1 unit
```

Theoretical Framework

- As in the classical linear regression all models are defined by two components:
  - The data generating process (DGP)
  - The relationship between the random disturbance or idiosyncratic shock and the explanatory variables
- From the first consideration, we can distinguish the DGP for the panel data case:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \eta_{it} \\ \eta_{it} &= \alpha_i + \varepsilon_{it} \end{aligned}$$

Random Effects Model

- The regressors are unrelated to the unobserved time invariant component α_i

$$E(\alpha_i | x_{it1}, \ldots x_{itk}) = E(\alpha_i)$$

- strict exogeneity, no lagged dependent variables:

$$E(\varepsilon_{it} | x_{it1}, \ldots, x_{itk}, \alpha_i) = 0$$

- The previous two assumptions allow us to think about using a regression. But:

$$\begin{aligned} E(\varepsilon_i \varepsilon_i' | x_i, \alpha_i) &= \sigma_\varepsilon^2 I_T \\ E(\varepsilon_{it}^2) &= \sigma_\varepsilon^2 \\ E(\varepsilon_{it}\varepsilon_{is}) &= 0 \\ V(\alpha_i) &= E(\alpha_i^2 | x_i) = \sigma_\alpha^2 \end{aligned}$$

Random Effects Model

- The regressors are unrelated to the unobserved time invariant component α_i

$$E\left(\alpha_i | x_{it1}, \ldots x_{itk}\right) = E\left(\alpha_i\right)$$

- strict exogeneity, no lagged dependent variables:

$$E\left(\varepsilon_{it} | x_{it1}, \ldots, x_{itk}, \alpha_i\right) = 0$$

- The previous two assumptions allow us to think about using a regression. But:

$$\begin{aligned} E\left(\varepsilon_i \varepsilon_i' | x_i, \alpha_i\right) &= \sigma_\varepsilon^2 I_T \\ E\left(\varepsilon_{it}^2\right) &= \sigma_\varepsilon^2 \\ E\left(\varepsilon_{it}\varepsilon_{is}\right) &= 0 \\ V\left(\alpha_i\right) &= E\left(\alpha_i^2 | x_i\right) = \sigma_\alpha^2 \end{aligned}$$

Random Effects Model

- The regressors are unrelated to the unobserved time invariant component α_i

$$E\left(\alpha_i | x_{it1}, \ldots x_{itk}\right) = E\left(\alpha_i\right)$$

- strict exogeneity, no lagged dependent variables:

$$E\left(\varepsilon_{it} | x_{it1}, \ldots, x_{itk}, \alpha_i\right) = 0$$

- The previous two assumptions allow us to think about using a regression. But:

$$\begin{aligned}
E\left(\varepsilon_i \varepsilon_i' | x_i, \alpha_i\right) &= \sigma_\varepsilon^2 I_T \\
E\left(\varepsilon_{it}^2\right) &= \sigma_\varepsilon^2 \\
E\left(\varepsilon_{it}\varepsilon_{is}\right) &= 0 \\
V\left(\alpha_i\right) &= E\left(\alpha_i^2 | x_i\right) = \sigma_\alpha^2
\end{aligned}$$

Random Effects Variance-Matrix

- For each individual we have that:

$$\Omega = E\left(\eta_i \eta_i'\right) = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\varepsilon^2 + \sigma_\alpha^2 & \cdots & \vdots \\ \vdots & & \ddots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\varepsilon^2 + \sigma_\alpha^2 \end{pmatrix}$$

- This gives rise to an efficient estimator:

$$\Omega^{-1/2} y_i = \Omega^{-1/2} x_i \beta + \Omega^{-1/2} \eta_i$$
$$\Omega^{-1/2} z_i = z_i^*$$

- This implies that we have the following model:

$$y_i^* = x_i^* \beta + \eta_i^*$$
$$E\left(\eta_i^* \eta_i^{*\prime}\right) = I_T$$

Random Effects Variance-Matrix

- For each individual we have that:

$$\Omega = E\left(\eta_i \eta_i'\right) = \begin{pmatrix} \sigma_\varepsilon^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\varepsilon^2 + \sigma_\alpha^2 & \dots & \vdots \\ \vdots & & \ddots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \dots & \sigma_\varepsilon^2 + \sigma_\alpha^2 \end{pmatrix}$$

- This gives rise to an efficient estimator:

$$\begin{aligned} \Omega^{-1/2} y_i &= \Omega^{-1/2} x_i \beta + \Omega^{-1/2} \eta_i \\ \Omega^{-1/2} z_i &\equiv z_i^* \end{aligned}$$

- This implies that we have the following model:

$$\begin{aligned} y_i^* &= x_i^* \beta + \eta_i^* \\ E\left(\eta_i^* \eta_i^{*\prime}\right) &= I_T \end{aligned}$$

Random Effects Estimation with Stata

```
. use panel_slides
. xtset country year
       panel variable:  country (unbalanced)
        time variable:  year, 2010 to 2018, but with a gap
                delta:  1 unit

. describe
Contains data from panel_slides.dta
  obs:         1,016
 vars:             9                          6 Mar 2020 13:40
------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------
country         long    %30.0g     country    Country Name
year            float   %10.0g
consumption     double  %10.0g                Consumption (Billions 2000 US$)
gdp             double  %10.0g                GDP (Billions 2000 US$)
irate           double  %10.0g                Deposit interest rate
region          long    %12.0g     region     Regional groups
ln_cons         float   %9.0g                 Log of consumption
ln_gdp          float   %9.0g                 Log of gdp
ln_irate        float   %9.0g                 Log of irate
------------------------------------------------------------------------------
Sorted by: country  year
     Note: Dataset has changed since last saved.
```

Random Effects Estimation with Stata

```
. xtreg ln_cons ln_gdp ln_irate,re

Random-effects GLS regression                   Number of obs     =      1,016
Group variable: country                         Number of groups  =        131
R-sq:                                           Obs per group:
     within  = 0.8033                                         min =          1
     between = 0.9859                                         avg =        7.8
     overall = 0.9847                                         max =          9
                                                Wald chi2(2)      =   13277.81
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
     ln_cons |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ln_gdp |    .958856   .0084128   113.98   0.000     .9423672    .9753449
    ln_irate |  -.0039294   .0041147    -0.95   0.340     -.011994    .0041352
       _cons |    .760708   .2065915     3.68   0.000     .3557961     1.16562
-------------+----------------------------------------------------------------
     sigma_u |   .2339765
     sigma_e |  .05205235
         rho |  .95284182   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Interpreting Results

- The `Wald chi2(df)` statistic is the equivalent of the `F` and regards the overall relevance of the model
- The three different `R-sq` statistics represent the variability of y explained by its predicted values. But there are three possible measures of y:
  - y_{it} OVERALL
  - $\bar{y}_i$ BETWEEN
  - $y_{it} - \bar{y}_i$ WITHIN
- `corr(u_i, X)` refers to the correlation between the time invariant component α_i, in this case called `u_i`, and the regressors. For the random effects we assume it is zero.
- `sigma_u`$= \sigma_\alpha$,
  `sigma_e`$= \sigma_\varepsilon$,
  `rho`$= \sigma^2_\alpha \left(\sigma^2_\varepsilon + \sigma^2_\alpha\right)^{-1}$

Random effects vs. Pooled OLS

```
. xttest0

Breusch and Pagan Lagrangian multiplier test for random effects

    ln_cons[country,t] = Xb + u[country] + e[country,t]

        Estimated results:
                         |       Var     sd = sqrt(Var)
                ---------+-----------------------------
                 ln_cons |   4.027035       2.006747
                       e |   .0027094       .0520524
                       u |    .054745       .2339765

        Test:   Var(u) = 0
                             chibar2(01) =   3108.85
                          Prob > chibar2 =    0.0000
```

Fixed Effects Models

- The regressors can be correlated with the unobserved time invariant component α_i

$$Cov\left(\alpha_i, x_i\right) \neq 0$$

- strict exogeneity, no lagged dependent variables:

$$E\left(\varepsilon_{it} | x_{it1}, \ldots, x_{itk}, \alpha_i\right) = 0$$

Fixed Effects Models

- The regressors can be correlated with the unobserved time invariant component α_i

$$Cov(\alpha_i, x_i) \neq 0$$

- strict exogeneity, no lagged dependent variables:

$$E(\varepsilon_{it} | x_{it1}, \ldots, x_{itk}, \alpha_i) = 0$$

Interpretation

- In the model we have been discussing:

$$ln(\text{consumption}_{it}) = \beta_0 + \beta_1 ln(\text{gdp}_{it}) + \beta_2 ln(\text{irate}_{it}) + \alpha_i + \varepsilon_{it}$$

- It is difficult to maintain, for a particular model, that the unobserved individual component is independent of all regressors

Within Transform

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \alpha_i + \varepsilon_{it} \qquad (1)$$

- If we take the average over the T observations of each panel, we obtain:

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_{i1} + \ldots + \beta_k \bar{x}_{ik} + \alpha_i + \bar{\varepsilon}_i$$

Where:

$$\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it},$$

$$\bar{x}_{ij} = T^{-1} \sum_{t=1}^{T} x_{itj}$$

- We now can construct the following object:

$$y_{it} - \bar{y}_i = (\beta_0 - \beta_0) + \beta_1 (x_{it1} - \bar{x}_{i1}) + \ldots + \beta_k (x_{itk} - \bar{x}_{ik}) + (\alpha_i - \alpha_i) + (\varepsilon_{it} - \varepsilon_i)$$

- And we can then estimate the parameters of interest from equation (1):

$$\tilde{y}_i = \beta_1 \tilde{x}_{i1} + \ldots + \beta_k \tilde{x}_{ik} + \tilde{\varepsilon}_i$$

Within Transform

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \alpha_i + \varepsilon_{it} \tag{1}$$

- If we take the average over the T observations of each panel, we obtain:

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_{i1} + \ldots + \beta_k \bar{x}_{ik} + \alpha_i + \bar{\varepsilon}_i$$

Where:

$$\bar{y}_i = T^{-1} \textstyle\sum_{t=1}^{T} y_{it},$$

$$\bar{x}_{ij} = T^{-1} \textstyle\sum_{t=1}^{T} x_{itj}$$

- We now can construct the following object:

$$y_{it} - \bar{y}_i = (\beta_0 - \beta_0) + \beta_1 (x_{it1} - \bar{x}_{i1}) + \ldots + \beta_k (x_{itk} - \bar{x}_{ik}) + (\alpha_i - \alpha_i) + (\varepsilon_{it} - \varepsilon_i)$$

- And we can then estimate the parameters of interest from equation (1):

$$\tilde{y}_i = \beta_1 \tilde{x}_{i1} + \ldots + \beta_k \tilde{x}_{ik} + \tilde{\varepsilon}_i$$

Within Transform

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \alpha_i + \varepsilon_{it} \quad (1)$$

- If we take the average over the T observations of each panel, we obtain:

$$\bar{y}_i = \beta_0 + \beta_1 \bar{x}_{i1} + \ldots + \beta_k \bar{x}_{ik} + \alpha_i + \bar{\varepsilon}_i$$

Where:

$$\bar{y}_i = T^{-1} \sum_{t=1}^{T} y_{it},$$

$$\bar{x}_{ij} = T^{-1} \sum_{t=1}^{T} x_{itj}$$

- We now can construct the following object:

$$y_{it} - \bar{y}_i = (\beta_0 - \beta_0) + \beta_1 (x_{it1} - \bar{x}_{i1}) + \ldots + \beta_k (x_{itk} - \bar{x}_{ik}) + (\alpha_i - \alpha_i) + (\varepsilon_{it} - \varepsilon_i)$$

- And we can then estimate the parameters of interest from equation (1):

$$\tilde{y}_i = \beta_1 \tilde{x}_{i1} + \ldots + \beta_k \tilde{x}_{ik} + \tilde{\varepsilon}_i$$

Within Estimation

```
. xtreg ln_cons ln_gdp ln_irate,fe

Fixed-effects (within) regression               Number of obs     =      1,016
Group variable: country                         Number of groups  =        131

R-sq:                                           Obs per group:
     within  = 0.8034                                         min =          1
     between = 0.9858                                         avg =        7.8
     overall = 0.9845                                         max =          9

                                                F(2,883)          =    1804.60
corr(u_i, Xb)  = -0.0175                        Prob > F          =     0.0000

------------------------------------------------------------------------------
     ln_cons |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      ln_gdp |    .958713    .016523    58.02   0.000      .926284     .991142
    ln_irate |  -.0074047   .0042761    -1.73   0.084     -.0157972    .0009878
       _cons |   .7750608   .4063998     1.91   0.057     -.0225615    1.572683
-------------+----------------------------------------------------------------
     sigma_u |  .24585324
     sigma_e |  .05205235
         rho |  .95709727   (fraction of variance due to u_i)
------------------------------------------------------------------------------
F test that all u_i=0: F(130, 883) = 152.63                  Prob > F = 0.0000
```

Fixed effects vs. Random effects

- Theory should be one of the main factors guiding your modeling decision
- However, you should present statistical test to back up your claims
    - Hausman test for fixed effects vs random effects
    - Mundlak test for fixed effects vs random effects

Hausman Test

- The following object has a Chi-Squared distribution with degrees of freedom equal to the number of regressors:

$$H = \left(\hat{\beta}_{fe} - \hat{\beta}_{re}\right)' \left[\widehat{VCE}_{fe} - \widehat{VCE}_{re}\right]^{-1} \left(\hat{\beta}_{fe} - \hat{\beta}_{re}\right)$$

- The test implicitly assumes that the random effects model is efficient, which in turn makes $\left[\widehat{VCE}_{fe} - \widehat{VCE}_{re}\right]$ positive definite.
- The test rules out heteroskedasticity and serial correlation

Hausman Test

```
. quietly xtreg ln_cons ln_gdp ln_irate,fe
. estimates store eq_fe
. quietly xtreg ln_cons ln_gdp ln_irate,re
. estimates store eq_re

. hausman eq_fe eq_re
```

	---- Coefficients ----			
	(b) eq_fe	(B) eq_re	(b-B) Difference	sqrt(diag(V_b-V_B)) S.E.
ln_gdp	.958713	.958856	-.000143	.0142209
ln_irate	-.0074047	-.0039294	-.0034753	.0011638

```
                 b = consistent under Ho and Ha; obtained from xtreg
  B = inconsistent under Ha, efficient under Ho; obtained from xtreg

   Test:  Ho:  difference in coefficients not systematic

                 chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                         =        17.25
               Prob>chi2 =       0.0002
```

Mundlak Test

- The main idea is to model the correlation between the unobserved component and the regressors.

$$E(\alpha_i | x_{it}) = \theta_0 + \theta_1 \bar{x}_{i1} + \ldots + \theta_k \bar{x}_{ik} + \nu_i$$

- This implies that:

$$\begin{aligned} E(y_{it} | x_{it},) &= (\beta_0 + \theta_0) + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \theta_1 \bar{x}_{i1} + \ldots + \theta_k \bar{x}_{ik} + \nu_i + \varepsilon_{it} \\ E(y_{it} | x_{it},) &= \gamma_0 + \gamma_1 x_{it} + \gamma_2 \bar{x}_i + \epsilon_{it} \end{aligned}$$

- If we have a random effects model:

$$\begin{aligned} \theta_1 &= \ldots = \theta_k = 0 \\ \gamma_2 &= 0 \end{aligned}$$

- If not the coefficients will have some meaning
- Therefore:

$$\begin{aligned} H_o &: \theta_1 = \ldots = \theta_k = 0 \\ H_o &: \gamma_2 = 0 \end{aligned}$$

Mundlak Test

- The main idea is to model the correlation between the unobserved component and the regressors.

$$E(\alpha_i|x_{it}) = \theta_0 + \theta_1\bar{x}_{i1} + \ldots + \theta_k\bar{x}_{ik} + \nu_i$$

- This implies that:

$$\begin{aligned} E(y_{it}|x_{it},) &= (\beta_0 + \theta_0) + \beta_1 x_{it1} + \ldots + \beta_k x_{itk} + \theta_1\bar{x}_{i1} + \ldots + \theta_k\bar{x}_{ik} + \nu_i + \varepsilon_{it} \\ E(y_{it}|x_{it},) &= \gamma_0 + \gamma_1 x_{it} + \gamma_2\bar{x}_i + \epsilon_{it} \end{aligned}$$

- If we have a random effects model:

$$\begin{aligned} \theta_1 &= \ldots = \theta_k = 0 \\ \gamma_2 &= 0 \end{aligned}$$

- If not the coefficients will have some meaning
- Therefore:

$$\begin{aligned} H_o &: \theta_1 = \ldots = \theta_k = 0 \\ H_o &: \gamma_2 = 0 \end{aligned}$$

Constructing the Mundlak Test

- First lets construct a list

```
. local explain ln_gdp ln_irate
. local explainm ln_gdpm ln_iratem
```

- Now lets generate a sample mean for each object of the list, and then run the auxiliary regression for the test:

```
. foreach var of local explain {
  2.    by id: egen double `var´m = mean(`var´)
  3. }
. xtreg ln_cons `explain´ `explainm´, re
```

Constructing the Mundlak Test

- First lets construct a list

```
. local explain ln_gdp ln_irate
. local explainm ln_gdpm ln_iratem
```

- Now lets generate a sample mean for each object of the list, and then run the auxiliary regression for the test:

```
. foreach var of local explain {
  2.    by id: egen double `var´m = mean(`var´)
  3. }
. xtreg ln_cons `explain´ `explainm´, re
```

Mundlak Results

```
. xtreg ln_cons `explain´ `explainm´, re
Random-effects GLS regression                   Number of obs     =      1,016
Group variable: country                         Number of groups  =        131
R-sq:                                           Obs per group:
     within  = 0.8034                                         min =          1
     between = 0.9872                                         avg =        7.8
     overall = 0.9864                                         max =          9
                                                Wald chi2(4)      =   13443.43
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000
```

ln_cons	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_gdp	.958713	.0165226	58.02	0.000	.9263292	.9910968
ln_irate	-.0074047	.004276	-1.73	0.083	-.0157855	.0009761
ln_gdpm	.0031773	.0191767	0.17	0.868	-.0344083	.0407629
ln_iratem	.0680654	.0184904	3.68	0.000	.0318248	.104306
_cons	.6099668	.2413124	2.53	0.011	.1370031	1.08293
sigma_u	.2339765					
sigma_e	.05205235					
rho	.95284182	(fraction of variance due to u_i)				

```
. test `explainm´
 ( 1)  ln_gdpm = 0
 ( 2)  ln_iratem = 0

           chi2(  2) =    13.60
         Prob > chi2 =    0.0011
```

MARGINAL ANALYISIS

Interpreting the Coefficients

- Notice that all the variables are in natural logs. Therefore:

$$E\left(ln\left(y_{it}\right)|lnx_{it},\alpha_i\right)=\beta_0+\beta_1 lnx_{it1}+\ldots+\beta_k lnx_{itk}+\alpha_i$$

- If you want the impact of a continuous regressor on y_{it}:

$$\frac{\partial E\left(y_{it}|x_{it},\alpha_i\right)}{\partial x_{itj}}\frac{x_{itj}}{E\left(y_{it}|x_{it},\alpha_i\right)}=\beta_j$$

- Use `margins` to get the elasticities (dydx() in this particular case):

```
. quietly xtreg ln_cons ln_gdp ln_irate,fe
. margins, dydx(*)

Average marginal effects                        Number of obs     =      1,016
Model VCE    : Conventional

Expression   : Linear prediction, predict()
dy/dx w.r.t. : ln_gdp ln_irate
```

	Delta-method dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_gdp	.958713	.016523	58.02	0.000	.9263284	.9910976
ln_irate	-.0074047	.0042761	-1.73	0.083	-.0157857	.0009763

Interpreting the Coefficients

- Notice that all the variables are in natural logs. Therefore:

$$E\left(ln\left(y_{it}\right)|lnx_{it},\alpha_i\right) = \beta_0 + \beta_1 lnx_{it1} + \ldots + \beta_k lnx_{itk} + \alpha_i$$

- If you want the impact of a continuous regressor on y_{it}:

$$\frac{\partial E\left(y_{it}|x_{it},\alpha_i\right)}{\partial x_{itj}} \frac{x_{itj}}{E\left(y_{it}|x_{it},\alpha_i\right)} = \beta_j$$

- Use `margins` to get the elasticities (dydx() in this particular case):

```
. quietly xtreg ln_cons ln_gdp ln_irate,fe
. margins, dydx(*)
```

```
Average marginal effects                        Number of obs     =      1,016
Model VCE    : Conventional

Expression   : Linear prediction, predict()
dy/dx w.r.t. : ln_gdp ln_irate
```

	Delta-method dy/dx	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_gdp	.958713	.016523	58.02	0.000	.9263284	.9910976
ln_irate	-.0074047	.0042761	-1.73	0.083	-.0157857	.0009763

Marginal Effects with interactions

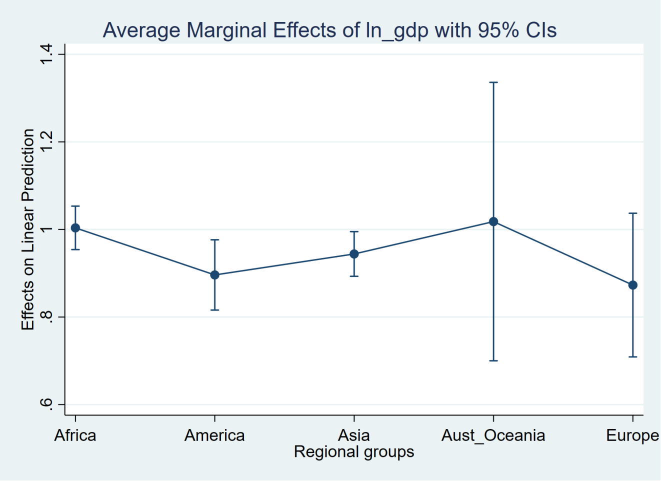
- Regional interactions with ln_gdp:

```
. quietly xtreg ln_cons i.region#c.ln_gdp ln_irate,fe

. margins, dydx(ln_gdp) over(region)

Average marginal effects                    Number of obs     =        986
Model VCE    : Conventional

Expression   : Linear prediction, predict()
dy/dx w.r.t. : ln_gdp
over         : region
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_gdp						
region						
Africa	1.003669	.0253091	39.66	0.000	.9540644	1.053274
America	.8961536	.0409304	21.89	0.000	.8159314	.9763758
Asia	.9440403	.0260334	36.26	0.000	.8930157	.9950649
Aust_Oceania	1.017993	.1622033	6.28	0.000	.70008	1.335905
Europe	.8729883	.0837015	10.43	0.000	.7089363	1.03704

Marginal effects by region

```
. marginsplot
 Variables that uniquely identify margins: region
```



Marginal Effects with interactions

- Regional interactions with ln_irate:

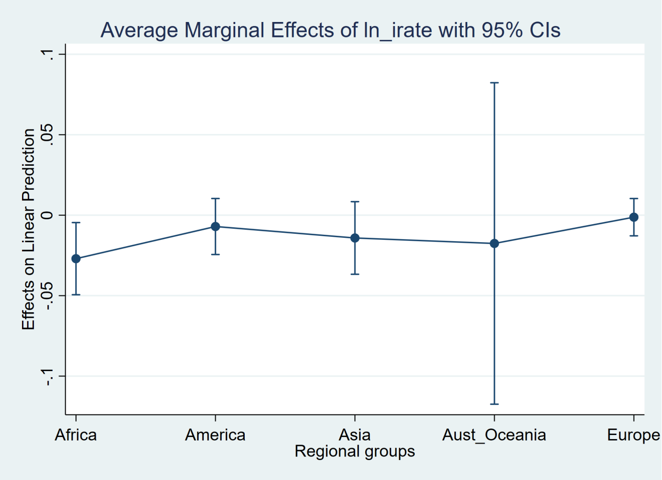
```
. quietly xtreg ln_cons ln_gdp ln_irate ///
                      i.region#c.ln_irate,fe

. margins,dydx(ln_irate) over(region)
```

```
Average marginal effects                        Number of obs     =        986
Model VCE      : Conventional

Expression     : Linear prediction, predict()
dy/dx w.r.t.   : ln_irate
over           : region
```

	dy/dx	Delta-method Std. Err.	z	P>|z|	[95% Conf.	Interval]
ln_irate						
region						
Africa	-.0270155	.0114371	-2.36	0.018	-.0494319	-.0045991
America	-.0070408	.0088792	-0.79	0.428	-.0244436	.0103621
Asia	-.0141604	.0115014	-1.23	0.218	-.0367028	.0083819
Aust_Oceania	-.0175721	.0509582	-0.34	0.730	-.1174484	.0823041
Europe	-.0012435	.0059118	-0.21	0.833	-.0128304	.0103435

Marginal effects by region

```
. marginsplot
 Variables that uniquely identify margins: region
```



DYNAMIC PANEL-DATA MODELS

Dynamic Models

$$y_{it} = \beta_0 + \beta_1 y_{i(t-1)} + x_{it}'\beta_2 + \alpha_i + \varepsilon_{it}$$

- In the model above x_{it} could also include lagged variables.
- Taking first differences:

$$\Delta y_{it} = \beta_1 \Delta y_{i(t-1)} + \Delta x_{it}'\beta_2 + \Delta\varepsilon_{it}$$

- We have eliminated the fixed effect but notice that:

$$E\left(\Delta y_{i(t-1)} \Delta\varepsilon_{it}\right) \neq 0$$

Dynamic Models

$$y_{it} = \beta_0 + \beta_1 y_{i(t-1)} + x_{it}'\beta_2 + \alpha_i + \varepsilon_{it}$$

- In the model above x_{it} could also include lagged variables.
- Taking first differences:

$$\Delta y_{it} = \beta_1 \Delta y_{i(t-1)} + \Delta x_{it}'\beta_2 + \Delta\varepsilon_{it}$$

- We have eliminated the fixed effect but notice that:

$$E\left(\Delta y_{i(t-1)} \Delta\varepsilon_{it}\right) \neq 0$$

Dynamic Models

$$y_{it} = \beta_0 + \beta_1 y_{i(t-1)} + x_{it}'\beta_2 + \alpha_i + \varepsilon_{it}$$

- In the model above x_{it} could also include lagged variables.
- Taking first differences:

$$\Delta y_{it} = \beta_1 \Delta y_{i(t-1)} + \Delta x_{it}'\beta_2 + \Delta\varepsilon_{it}$$

- We have eliminated the fixed effect but notice that:

$$E\left(\Delta y_{i(t-1)} \Delta\varepsilon_{it}\right) \neq 0$$

Instrumental Variable (GMM) Estimation

- The key to estimation is to find a set of instruments that satisfy:

$$E\left(z_{it}\Delta\varepsilon_{it}\right) = 0$$

- This gives rise to the following models:
  - Anderson-Hsiao $y_{i(t-2)}$ and $\Delta y_{i(t-2)}$ (`xtivreg, fd`).
  - Arellano and Bond suggest using all available lag levels (not only the second lag) for the first difference equation (`xtabond`).

Instrumental Variable (GMM) Estimation

- The key to estimation is to find a set of instruments that satisfy:

$$E\left(z_{it}\Delta\varepsilon_{it}\right) = 0$$

- This gives rise to the following models:
  - Anderson-Hsiao $y_{i(t-2)}$ and $\Delta y_{i(t-2)}$ (`xtivreg, fd`).
  - Arellano and Bond suggest using all available lag levels (not only the second lag) for the first difference equation (`xtabond`).

Arellano-Bond

```
. xtabond ln_cons ln_gdp ln_irate,twostep
Arellano-Bond dynamic panel-data estimation     Number of obs      =       761
Group variable: country                         Number of groups   =       121
Time variable: year
                                                Obs per group:
                                                              min =          1
                                                              avg =   6.289256
                                                              max =          7

Number of instruments =     31                  Wald chi2(3)       =   9345.33
                                                Prob > chi2        =    0.0000
Two-step results
------------------------------------------------------------------------------
     ln_cons |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ln_cons |
         L1. |   .3616734   .0072234    50.07   0.000     .3475158     .375831
             |
      ln_gdp |    .602238   .0073699    81.72   0.000     .5877932    .6166828
    ln_irate |  -.0085773   .0024087    -3.56   0.000    -.0132982   -.0038564
       _cons |   .7702696   .2566304     3.00   0.003     .2672833    1.273256
------------------------------------------------------------------------------
Warning: gmm two-step standard errors are biased; robust standard
         errors are recommended.
Instruments for differenced equation
        GMM-type: L(2/.).ln_cons
        Standard: D.ln_gdp D.ln_irate
Instruments for level equation
        Standard: _cons
```

Model Specification

```
. estat sargan

Sargan test of overidentifying restrictions
      H0: overidentifying restrictions are valid

      chi2(27)       =  32.56842
      Prob > chi2    =    0.2117
```

- The overidentification restriction is a test of the validity of the instruments under correct specification.

Model Specification

```
. estat abond
Arellano-Bond test for zero autocorrelation
in first-differenced errors
```

Order	z	Prob > z
1	-2.5248	0.0116
2	1.5938	0.1110

H0: no autocorrelation

- The arellano-bond test is testing that Ho: $E\left[\Delta\varepsilon_{it}\Delta\varepsilon_{i(t-1)}\right] \neq 0$:

$$E\left[\Delta\varepsilon_{it}\Delta\varepsilon_{i(t-1)}\right] = E\left[\left(\varepsilon_{it}-\varepsilon_{i(t-1)}\right)\left(\varepsilon_{i(t-1)}-\varepsilon_{i(t-2)}\right)\right] = E\left[\varepsilon^2_{i(t-1)}\right] + 0$$

- According to our asssumptions we should reject this hypothesis. Also, according to our hypothesis:

$$\begin{aligned} E\left[\Delta\varepsilon_{it}\Delta\varepsilon_{i(t-2)}\right] &= E\left[\left(\varepsilon_{it}-\varepsilon_{i(t-1)}\right)\left(\varepsilon_{i(t-2)}-\varepsilon_{i(t-3)}\right)\right] \\ &= E\left(\varepsilon_{it}\varepsilon_{i(t-2)}\right) - E\left(\varepsilon_{it}\varepsilon_{i(t-3)}\right) + E\left(\varepsilon_{i(t-1)}\varepsilon_{i(t-2)}\right) \\ &\quad - E\left(\varepsilon_{i(t-1)}\varepsilon_{i(t-3)}\right) \\ &= 0 \end{aligned}$$

A New Set of Moment Conditions

- The lagged-level instruments in `xtabond` become weak as the AR process becomes too persistent or σ_u^2/σ_e^2 becomes too large, so a new set of moments conditions are proposed:

$$E\left(z_{it}\Delta\varepsilon_{it}\right) = 0$$
$$E\left(\Delta z_{it}\varepsilon_{it}\right) = 0$$

- These are defined by Arellano-Bover/Blundell-Bond.
- Notice that you have moments for the equation in levels and for the equation in first difference
- Fit this model with `xtdpdsys`

A New Set of Moment Conditions

- The lagged-level instruments in `xtabond` become weak as the AR process becomes too persistent or σ_u^2/σ_e^2 becomes too large, so a new set of moments conditions are proposed:

$$
\begin{aligned}
E\left(z_{it}\Delta\varepsilon_{it}\right) &= 0 \\
E\left(\Delta z_{it}\varepsilon_{it}\right) &= 0
\end{aligned}
$$

- These are defined by Arellano-Bover/Blundell-Bond.
- Notice that you have moments for the equation in levels and for the equation in first difference
- Fit this model with `xtdpdsys`

Arellano-Bover/Blundell-Bond

```
. xtdpdsys ln_cons ln_gdp ln_irate,twostep
System dynamic panel-data estimation          Number of obs      =       884
Group variable: country                       Number of groups   =       122
Time variable: year
                                              Obs per group:
                                                            min =         1
                                                            avg =  7.245902
                                                            max =         8

Number of instruments =     38                Wald chi2(3)      =  36908.02
                                              Prob > chi2       =    0.0000
Two-step results
------------------------------------------------------------------------------
     ln_cons |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     ln_cons |
         L1. |    .464653   .0063034    73.71   0.000     .4522985    .4770074
             |
      ln_gdp |   .4918536   .0051095    96.26   0.000     .4818391     .501868
    ln_irate |  -.0092232   .0029176    -3.16   0.002    -.0149415   -.0035049
       _cons |   .9754017   .1538629     6.34   0.000     .6738359    1.276967
------------------------------------------------------------------------------
Warning: gmm two-step standard errors are biased; robust standard
         errors are recommended.
Instruments for differenced equation
        GMM-type: L(2/.).ln_cons
        Standard: D.ln_gdp D.ln_irate
Instruments for level equation
        GMM-type: LD.ln_cons
        Standard: _cons
```

Arellano-Bover/Blundell-Bond

Overidentification and Autocorrelation Tests

```
. estat sargan

 Sargan test of overidentifying restrictions
       H0: overidentifying restrictions are valid
       chi2(34)      =  46.01339
       Prob > chi2   =    0.0819

. estat abond

Arellano-Bond test for zero autocorrelation
in first-differenced errors
```

Order	z	Prob > z
1	-2.6633	0.0077
2	1.6218	0.1048

```
   H0: no autocorrelation
```

Your Own Dynamic Model

- This models relies heavily on the idea that the dynamics are correctly specified
- For instance you could have:

$$\begin{aligned} y_{it} &= \beta_0 + \beta_1 y_{i(t-1)} + x'_{it}\beta_2 + \alpha_i + \varepsilon_{it} + \gamma\varepsilon_{i(t-1)} \\ \Delta y_{it} &= \Delta\beta_1 y_{i(t-1)} + \Delta x'_{it}\beta_2 + \Delta\varepsilon_{it} + \gamma\Delta\varepsilon_{i(t-1)} \end{aligned}$$

- You now need to construct a new set of instruments that satisfy the moment conditions.
- Stata allows you to do this with `xtdpd`. You need to specify the instruments for the level and difference equations.

Your Own Dynamic Model

- This models relies heavily on the idea that the dynamics are correctly specified
- For instance you could have:

$$
\begin{aligned}
y_{it} &= \beta_0 + \beta_1 y_{i(t-1)} + x_{it}'\beta_2 + \alpha_i + \varepsilon_{it} + \gamma\varepsilon_{i(t-1)} \\
\Delta y_{it} &= \Delta\beta_1 y_{i(t-1)} + \Delta x_{it}'\beta_2 + \Delta\varepsilon_{it} + \gamma\Delta\varepsilon_{i(t-1)}
\end{aligned}
$$

- You now need to construct a new set of instruments that satisfy the moment conditions.
- Stata allows you to do this with `xtdpd`. You need to specify the instruments for the level and difference equations.

Extended Regression Models

Extended regression models for panel data

- Random effects linear regression with endogenous covariates
  - ```
    xteregress y x1 x2, ///
              endogenous(w = x1 z1 z2)
    ```
- Random effects linear regression with sample selection
  - ```
    xteregress y x1 x2, ///
              select(selected = x2 w2)
    ```
- Random effects linear regression with endogenous treatment
  - ```
    xteregress y x1 x2, ///
              entreat(treatment = w z2 z3)
    ```
    ```

Extended regression models for panel data

- Binary dependent variables
  - ```
    xteprobit y x1 x2,                ///
          endogenous(w = x1 z1 z2) ///
          select(selected = x2 w2) ///
          entreat(treatment = w z2 z3)
    ```
- Random effects ordered probit regression
  - `xteoprobit`
- Random effects Interval regression
  - `xteintreg`
- Random effects Heckman model
  - `xtheckman`
    ```

Random effects probit regression with sample selection

```
. webuse womenhlthre,clear
(Women´s health status panel)
. xtset personid year
       panel variable:  personid (strongly balanced)
        time variable:  year, 2010 to 2013
                delta:  1 unit
. generate goodhlth = health>3


. describe
Contains data from https://www.stata-press.com/data/r16/womenhlthre.dta
  obs:         7,200                          Women´s health status panel
 vars:            10                          6 Sep 2018 16:14
------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------
grade           byte    %8.0g                 Years of education
personid        int     %9.0g                 Person ID
year            int     %9.0g                 Year
workschool      byte    %8.0g      yesno      Employed or in school
insured         byte    %8.0g      yesno      Has health insurance
regcheck        byte    %8.0g      yesno      Has regular checkups
select          byte    %8.0g                 In sample
exercise        byte    %8.0g      yesno      Exercises regularly
health          byte    %9.0g      status     Health status
goodhlth        float   %9.0g                 Good-Excellent Health condition
------------------------------------------------------------------------
Sorted by: personid  year
     Note: Dataset has changed since last saved.
```

Random effects probit regression with sample selection

```
. xteprobit goodhlth exercise grade, select(select = grade i.regcheck)

(setting technique to bhhh)
Iteration 0:   log likelihood =  -6840.671
Iteration 1:   log likelihood = -6808.6475
Iteration 2:   log likelihood = -6808.1535
Iteration 3:   log likelihood = -6808.1515
Iteration 4:   log likelihood = -6808.1515
Extended probit regression                      Number of obs     =      7,200
                                                      Selected    =      5,421
                                                      Nonselected =      1,779

Group variable: personid                        Number of groups  =      1,800

                                                Obs. per group:
                                                              min =          4
                                                              avg =        4.0
                                                              max =          4

Integration method: mvaghermite                 Integration pts.  =          7

                                                Wald chi2(2)      =     348.34
Log likelihood = -6808.1515                     Prob > chi2       =     0.0000
```

Random effects probit regression with sample selection

```
. xteprobit goodhlth exercise grade, select(select = grade i.regcheck)
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
goodhlth						
exercise	.3554439	.0400762	8.87	0.000	.276896	.4339919
grade	.1743015	.0095533	18.25	0.000	.1555774	.1930256
_cons	-2.252753	.1154867	-19.51	0.000	-2.479102	-2.026403
select						
grade	.0832256	.007392	11.26	0.000	.0687376	.0977137
regcheck						
yes	.4800144	.036039	13.32	0.000	.4093793	.5506495
_cons	-.5420435	.0964841	-5.62	0.000	-.731149	-.3529381
corr(e.select,e.goodhlth)	.8060986	.0855705	9.42	0.000	.5627727	.9208657
var(goodhlth[personid])	.2640095	.0364768			.2013787	.346119
var(select[personid])	.1538155	.0271043			.1088948	.2172667
corr(select[personid], goodhlth[personid])	.6224091	.0808206	7.70	0.000	.4384837	.7562961

Random effects probit regression with endogenous treatment

```
. xteprobit goodhlth exercise grade,      ///
>     entreat(insured = workschool,nore) nolog

Extended probit regression                      Number of obs     =      7,200
Group variable: personid                        Number of groups  =      1,800
                                                Obs. per group:
                                                              min =          4
                                                              avg =        4.0
                                                              max =          4
Integration method: mvaghermite                 Integration pts.  =          7
                                                Wald chi2(6)      =     265.10
Log likelihood =  -7572.592                     Prob > chi2       =     0.0000
```

Random effects probit regression with endogenous treatment

```
. xteprobit goodhlth exercise grade,      ///
>     entreat(insured = workschool,nore) nolog
```

	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
goodhlth						
insured#c.exercise						
no	.5563098	.0916258	6.07	0.000	.3767266	.735893
yes	.486376	.0454754	10.70	0.000	.3972458	.5755062
insured#c.grade						
no	.0125397	.0207005	0.61	0.545	-.0280325	.053112
yes	.0788714	.0100576	7.84	0.000	.0591589	.098584
insured						
no	-1.398234	.3668983	-3.81	0.000	-2.117342	-.679127
yes	-.6820556	.1458962	-4.67	0.000	-.9680069	-.3961043
insured						
workschool	.6620277	.058127	11.39	0.000	.5481008	.7759545
_cons	-.0088057	.0557336	-0.16	0.874	-.1180415	.1004301
corr(e.insured,e.goodhlth)	.3433395	.1522733	2.25	0.024	.0195374	.6019547
var(goodhlth[personid])	.3394691	.0451158			.2616222	.4404797

Panel-Data Unit-Root and Cointegration Tests

Panel-data unit-root tests

- Ho: All panels contain a unit root.
  H1: At least some panels contain unit roots.
  - `xtunitroot llc`: Levin-Lin-Chu test
  - `xtunitroot ht`: Harris-Tzavalis test
  - `xtunitroot breitung`: Breitung test
  - `xtunitroot ips`: Im-Pesaran-Shin test
  - `xtunitroot fisher`: Fisher-type test
- Ho: All the panels are (trend) stationary.
  H1: At least some panels contain unit roots.
  - `xtunitroot hadri`: Hadri LM stationarity test

Panel-data cointegration tests

- Ho: No cointegration.
  H1: Variables cointegrated in all panels.
  - `xtcointtest kao`
  - `xtcointtest pedroni`
  - `xtcointtest westerlund, allpanels`
- Ho: No cointegration.
  H1: Variables cointegrated in some panels.
  - `xtcointtest westerlund, somepanels`

Panel unit root test (fictitious data)

```
. use productivity
(Fictitious cointegration data)
. xtset
Panel variable: id (strongly balanced)
 Time variable: time, 1973q3 to 2010q4
         Delta: 1 quarter
. xtsum product rddomest rdfor
```

Variable		Mean	Std. dev.	Min	Max	Observations
product	overall	9.030543	5.980432	-7.10538	26.32578	N = 15000
	between		1.738041	3.524668	13.841	n = 100
	within		5.724926	-4.839727	26.2163	T = 150
rddomest	overall	74.09569	43.99242	-3.273872	170.2226	N = 15000
	between		6.873214	53.24847	90.06024	n = 100
	within		43.45757	-14.45431	161.2792	T = 150
rdfor	overall	44.6047	27.0383	-4.653028	116.5821	N = 15000
	between		6.410035	24.97034	58.58555	n = 100
	within		26.27526	-16.42538	109.3243	T = 150

Panel unit root test for productivity

```
. xtunitroot breitung product
Breitung unit-root test for product
-------------------------------------
H0: Panels contain unit roots           Number of panels  =    100
Ha: Panels are stationary               Number of periods =    150

AR parameter: Common                    Asymptotics:  T,N -> Infinity
Panel means:  Included                                   sequentially
Time trend:   Not included              Prewhitening: Not performed
------------------------------------------------------------------------
                     Statistic      p-value
------------------------------------------------------------------------
 lambda                5.8117        1.0000
------------------------------------------------------------------------

. xtunitroot breitung D.product
Breitung unit-root test for D.product
---------------------------------------
H0: Panels contain unit roots           Number of panels  =    100
Ha: Panels are stationary               Number of periods =    149

AR parameter: Common                    Asymptotics:  T,N -> Infinity
Panel means:  Included                                   sequentially
Time trend:   Not included              Prewhitening: Not performed
------------------------------------------------------------------------
                     Statistic      p-value
------------------------------------------------------------------------
 lambda              -61.6147        0.0000
------------------------------------------------------------------------
```

Panel unit root test for RD_domestic

```
. xtunitroot breitung rddomest
Breitung unit-root test for rddomest
------------------------------------
H0: Panels contain unit roots              Number of panels  =    100
Ha: Panels are stationary                  Number of periods =    150
AR parameter: Common                       Asymptotics:  T,N -> Infinity
Panel means:  Included                                     sequentially
Time trend:   Not included                 Prewhitening: Not performed
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 lambda              74.2929        1.0000
------------------------------------------------------------------------------
. xtunitroot breitung D.rddomest
Breitung unit-root test for D.rddomest
--------------------------------------
H0: Panels contain unit roots              Number of panels  =    100
Ha: Panels are stationary                  Number of periods =    149
AR parameter: Common                       Asymptotics:  T,N -> Infinity
Panel means:  Included                                     sequentially
Time trend:   Not included                 Prewhitening: Not performed
------------------------------------------------------------------------------
                    Statistic      p-value
------------------------------------------------------------------------------
 lambda             -62.9818        0.0000
------------------------------------------------------------------------------
```

Panel unit root test for RD_foreign

```
. xtunitroot breitung rdfor
Breitung unit-root test for rdfor
-----------------------------------
H0: Panels contain unit roots             Number of panels  =    100
Ha: Panels are stationary                 Number of periods =    150
AR parameter: Common                      Asymptotics:   T,N -> Infinity
Panel means:  Included                                  sequentially
Time trend:   Not included                Prewhitening: Not performed
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 lambda             52.9037        1.0000
------------------------------------------------------------------------------

. xtunitroot breitung D.rdfor
Breitung unit-root test for D.rdfor
-------------------------------------
H0: Panels contain unit roots             Number of panels  =    100
Ha: Panels are stationary                 Number of periods =    149
AR parameter: Common                      Asymptotics:   T,N -> Infinity
Panel means:  Included                                  sequentially
Time trend:   Not included                Prewhitening: Not performed
------------------------------------------------------------------------------
                   Statistic      p-value
------------------------------------------------------------------------------
 lambda            -57.7736        0.0000
------------------------------------------------------------------------------
```

Panel cointegration test for productivity, RD_domestic and RD_foreign

```
. xtcointtest kao product rddomest rdfor
Kao test for cointegration
-------------------------------------------------------------------------------
H0: No cointegration                          Number of panels       =    100
Ha: All panels are cointegrated               Number of periods      =    148
Cointegrating vector: Same
Panel means:          Included                Kernel:            Bartlett
Time trend:           Not included            Lags:              3.60 (Newey-West)
AR parameter:         Same                    Augmented lags:    1
-------------------------------------------------------------------------------
                                              Statistic          p-value
-------------------------------------------------------------------------------
 Modified Dickey-Fuller t                     -23.6733           0.0000
 Dickey-Fuller t                              -15.1293           0.0000
 Augmented Dickey-Fuller t                     -3.6909           0.0001
 Unadjusted modified Dickey-Fuller t          -46.7561           0.0000
 Unadjusted Dickey-Fuller t                   -20.2521           0.0000
-------------------------------------------------------------------------------
```

Summing up

- Basic Concepts
- Linear panel-data models
  - Random effects
  - Fixed effects
  - Marginal analysis
- Dynamic panel-data models
  - Arellano-Bond
  - Arellano-Bover/Blundell-Bond
- Extended regression models for panel data
- Panel unit roots and cointegration