

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Introduction to Panel-Data Analysis using Stata

Gustavo Sánchez

StataCorp LLC

May 6, 2026
StataCorp LLC



Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

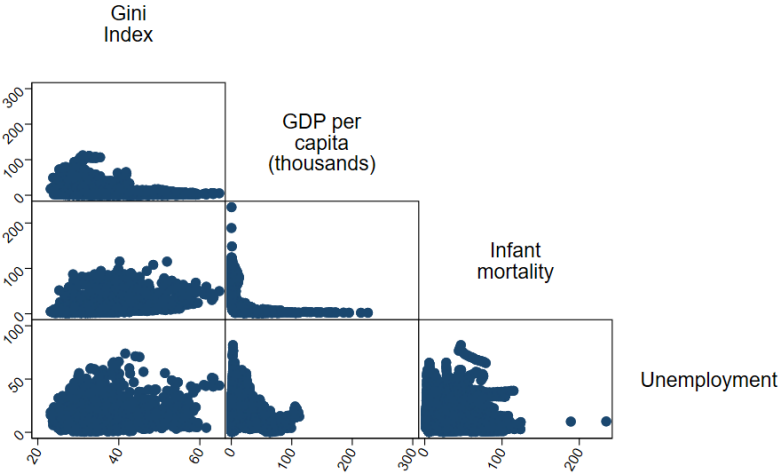
Final remarks

References

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
 - IV regression
- Dynamic panel-data linear models
 - Arellano-Bond
 - Arellano-Bover/Blundell-Bond
 - Panel VAR models
- Nonlinear panel data models

Pooled scatter plots



Source: World Bank via -import_fred-

Pooled vs. Panel

Scatter plots by country

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

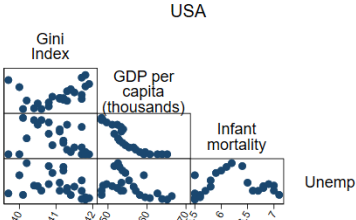
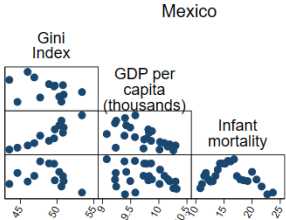
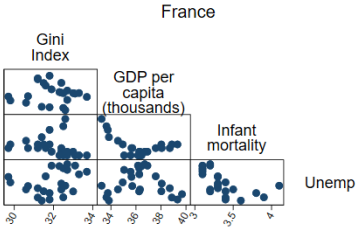
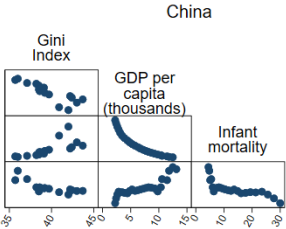
Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References



Source:World Bank via -import_fred-

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

```
. list country year gini_index gdp_capita inf_mrate ///
>           if (country==126 | country==214)           ///
>           & year>=2015 & year<=2023, noobs           ///
>           sepby(country) abbreviate(12)
```

country	year	gini_index	gdp_capita	inf_mrate
Mexico	2015	.	10.02124	13.9
Mexico	2016	46.9	10.1005	13.5
Mexico	2017	.	10.19377	13
Mexico	2018	46	10.29687	12.6
Mexico	2019	.	10.15945	12.2
Mexico	2020	44.6	9.234644	11.8
Mexico	2021	.	9.728057	11.5
Mexico	2022	43.5	10.01325	11.1
Mexico	2023	.	10.25918	10.8
US	2015	41.5	56.57292	5.8
US	2016	41.3	57.15147	5.7
US	2017	41.4	58.1517	5.6
US	2018	41.8	59.52666	5.6
US	2019	41.9	60.75099	5.5
US	2020	40	59.19467	5.5
US	2021	39.7	62.68025	5.5
US	2022	41.7	63.88613	5.5
US	2023	41.8	65.1866	5.5

How it Looks

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

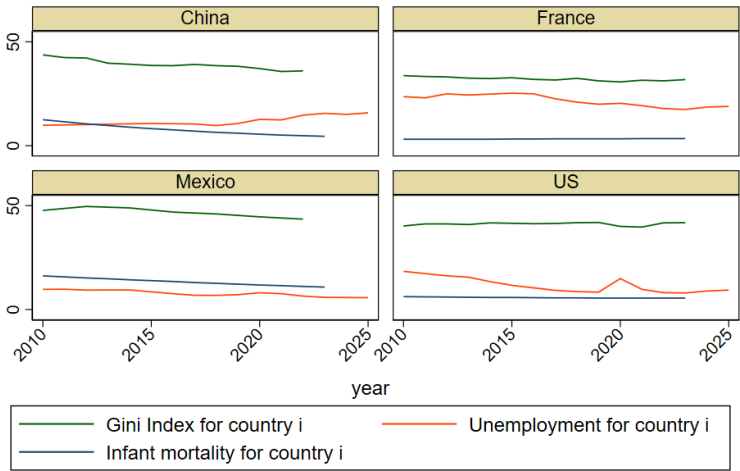
Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Gini index, unemployment, and mortality rate



Source: World Bank via -import_fred-

Stata Tools

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

Data management

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

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

LINEAR PANEL-DATA MODELS

Data Generating Process

- The data generating process is given by:

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \eta_{it}$$

$$\eta_{it} \equiv \alpha_i + \varepsilon_{it}$$

$$i = 1, \dots, n$$

$$t = 1, \dots, T$$

- We have redefined the nature of the random disturbance to include an unobservable component
 - Unobservable component is specific to each panel and independent of time.
 - Assumptions made on η_{it} , with particular emphasis on α_i , define the model we work with.

Example 1: Model for GINI (Income inequality) index

$$gini_index_{it} = \beta_0 + gdp_capita_{it} * \beta_1 + others_{it} * \beta_2 + \mu_i + \nu_{it}$$

Data

- World Bank public online data on:

gini_index: Gini inequality index for country i

gdp_capita: GDP per capita for country i

inf_mrate: Infant mortality rate for county i

unemployment: Unemployment for county i

- Example : 2000-2025 for 227 countries
- Source: <http://databank.worldbank.org/data/Home.aspx>

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Specifying the panel structure in Stata

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

```
. use $data_dir_gini\gini_data,clear
```

```
. xtset country year
```

Panel variable: country (unbalanced)

Time variable: year, 2000 to 2025

Delta: 1 unit

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

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:

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

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

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

$$E(\alpha_j | x_{it1}, \dots, x_{itk}) = E(\alpha_j)$$

- strict exogeneity, no lagged dependent variables:

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

- The previous two assumptions allow us to think about fitting a linear regression model. But:

$$E(\varepsilon_j \varepsilon_j' | x_j, \alpha_j) = \sigma_\varepsilon^2 I_T$$

$$E(\varepsilon_{it}^2) = \sigma_\varepsilon^2$$

$$E(\varepsilon_{it} \varepsilon_{is}) = 0$$

$$V(\alpha_j) = E(\alpha_j^2 | x_j) = \sigma_\alpha^2$$

Random Effects Variance-Matrix

- For each individual we have that:

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

- 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:

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

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure

Random Effects

Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Random Effects Variance-Matrix

- For each individual we have that:

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

- 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:

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

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure

Random Effects

Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Random Effects Estimation with Stata

```
. describe using $data_dir_gini\gini_data
```

Contains data
Observations: 5,876 23 Mar 2026 09:42
Variables: 7

Variable name	Storage type	Display format	Value label	Variable label
country	long	%13.0g	country	
year	float	%9.0g		
gdp_capita	float	%9.0g		GDP per capita for country i
gini_index	float	%9.0g		Gini Index for country i
le_	float	%9.0g		Life expectancy for country i
inf_mrate	float	%9.0g		Infant mortality for country i
unemployment	float	%9.0g		Unemployment for country i

Sorted by: country year

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Random Effects Estimation with Stata

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtreg gini_index gdp_capita inf_mrate unemployment, re
```

Random-effects GLS regression

Group variable: country

R-squared:

Within	=	0.1378
Between	=	0.1480
Overall	=	0.1645

corr(u_i, X) = 0 (assumed)

Number of obs	=	1,687
Number of groups	=	144
Obs per group:		
min	=	1
avg	=	11.7
max	=	24
Wald chi2(3)	=	266.00
Prob > chi2	=	0.0000

gini_index	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
gdp_capita	-.0807663	.0195266	-4.14	0.000	-.1190377	-.0424948
inf_mrate	.138271	.010222	13.53	0.000	.1182364	.1583057
unemployment	.04234	.012137	3.49	0.000	.0185519	.0661281
_cons	34.55571	.7570524	45.65	0.000	33.07191	36.03951
sigma_u	6.9820755					
sigma_e	2.5213379					
rho	.88463898	(fraction of variance due to u_i)				

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs REMarginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

- 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.
- $\text{sigma_u} = \sigma_\alpha$,
 $\text{sigma_e} = \sigma_\varepsilon$,
 $\text{rho} = \sigma_\alpha^2 (\sigma_\varepsilon^2 + \sigma_\alpha^2)^{-1}$

Random effects vs. Pooled OLS

```
. xttest0
```

Breusch and Pagan Lagrangian multiplier test for random effects

```
gini_index[country,t] = Xb + u[country] + e[country,t]
```

Estimated results:

	Var	SD = sqrt(Var)
gini_in_x	66.87245	8.177558
e	6.357145	2.521338
u	48.74938	6.982075

Test: Var(u) = 0

chibar2(01) = 11193.71
 Prob > chibar2 = 0.0000

Outline

Basic
concepts

Pooled vs. Panel
 Stata tools

Linear PD

DGP
 Data and model
 Panel structure

Random Effects

Fixed Effects
 FE vs RE
 Marginal Analysis
 Autocorrelation /
 Heteroskedasticity
 IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
Models

Probit Models
 Logit Models
 Poisson Models

Final remarks

References

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects

Fixed Effects

FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Fixed Effects Models

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

$$\text{Cov}(\alpha_i, x_i) \neq 0$$

- strict exogeneity, no lagged dependent variables:

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

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Interpretation

- In the model we have been discussing:

$$gini_index_{it} = \beta_0 + gdp_capita_{it} * \beta_1 + others_{it} * \beta_2 + \alpha_i + \nu_{it} \quad (1)$$

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

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \alpha_i + \varepsilon_{it} \quad (2)$$

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

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

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:

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

- And we can then estimate the parameters of interest from equation (2):

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

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects

Fixed Effects

FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

$$y_{it} = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \alpha_i + \varepsilon_{it} \quad (2)$$

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

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

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

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:

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

- And we can then estimate the parameters of interest from equation (2):

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

Example 2: Within Estimation

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtreg gini_index gdp_capita inf_mrate unemployment, fe
```

Fixed-effects (within) regression

Group variable: country

R-squared:

Within	= 0.1382	Number of obs	= 1,687
Between	= 0.1446	Number of groups	= 144
Overall	= 0.1646	Obs per group:	
		min	= 1
		avg	= 11.7
		max	= 24
		F(3, 1540)	= 82.33
		Prob > F	= 0.0000

corr(u_i, Xb) = -0.0870

gini_index	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
gdp_capita	-.0899808	.0228099	-3.94	0.000	-.1347225	-.0452391
inf_mrate	.1485641	.0108791	13.66	0.000	.1272247	.1699035
unemployment	.0343001	.012572	2.73	0.006	.0096401	.0589602
_cons	35.35043	.5546774	63.73	0.000	34.26243	36.43844
sigma_u	7.3031792					
sigma_e	2.5213379					
rho	.89350369	(fraction of variance due to u_i)				

F test that all u_i=0: F(143, 1540) = 92.28 **Prob > F = 0.0000**

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

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

The test favors fixed effects estimation

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtreg gini_index gdp_capita      ///  
>           inf_mrate unemployment, fe  
. estimates store eq_fe  
  
. xtreg gini_index gdp_capita      ///  
>           inf_mrate unemployment, re  
. estimates store eq_re
```

```
. hausman eq_fe eq_re
```

	Coefficients		(b-B) Difference	sqrt (diag (V_b-V_B)) Std. err.
	(b) eq_fe	(B) eq_re		
gdp_capita	-.0899808	-.0807663	-.0092145	.0117899
inf_mrate	.1485641	.138271	.0102931	.0037238
unemployment	.0343001	.04234	-.0080398	.0032786

```
           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)  
                   = 10.99  
Prob > chi2 = 0.0118
```

Mundlak Test

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

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

- This implies that:

$$E(y_{it} | x_{it},) = (\beta_0 + \theta_0) + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \theta_1 \bar{x}_{i1} + \dots + \theta_k \bar{x}_{ik}$$

$$E(y_{it} | x_{it},) = \gamma_0 + \gamma_1 x_{it} + \gamma_2 \bar{x}_i$$

- If we have a random effects model:

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

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

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

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects

FE vs RE

Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Mundlak Test

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

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

- This implies that:

$$E(y_{it} | x_{it},) = (\beta_0 + \theta_0) + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \theta_1 \bar{x}_{i1} + \dots + \theta_k \bar{x}_{ik}$$

$$E(y_{it} | x_{it},) = \gamma_0 + \gamma_1 x_{it} + \gamma_2 \bar{x}_i$$

- If we have a random effects model:

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

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

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

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects

FE vs RE

Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

The Mundlak test also favors FE, why?

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. quietly xtreg gini_index gdp_capita inf_mrate unemployment, re  
. estat mundlak
```

Mundlak specification test
H0: Covariates are uncorrelated with unobserved panel-level effects

```
chi2(3) = 10.57  
Prob > chi2 = 0.0143
```

**Notes: Fixed effects and correlated random effects are
consistent under H0 and Ha.
Random effects are efficient under H0.**

In StataNow 18.5 we added **xtreg, cre**

- With the **cre** option, you can fit the Mundlak specification
- The result for **estat Mundlak** will be exactly the same

```
. quietly xtreg gini_index gdp_capita inf_mrate unemployment, cre
.
. estat mundlak
```

Mundlak specification test

H0: Covariates are uncorrelated with unobserved panel-level effects

chi2(3) = 10.57

Prob > chi2 = 0.0143

Notes: Fixed effects and correlated random effects are consistent under H0 and Ha.

Random effects are efficient under H0.

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Also in 18.5, we added high dimensional fixed effects

- Instead of adding a set of dummies for a second dimension with **xtreg**, you can now use **absorb()**
- You can actually add more dimensions with **absorb()**
- This option can also be used with **areg** and **ivregress 2sls**

Let's absorb time as a second dimension for FE

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtreg gini_index gdp_capita inf_mrate unemployment, fe absorb(year) nolog
note: term year has 2023 levels in a dataset of size 1687
Alternating projection maximum absolute difference = 2.899e-09
Fixed-effects (within) regression               Number of obs   =       1,687
Group variable: country                        Number of groups =        144
R-squared:                                     Obs per group:
    Within   = 0.1974                               min =          1
    Between  = 0.0114                               avg  =       11.7
    Overall  = 0.0114                               max  =         24
                                                F(3, 1517)      =       13.10
                                                Prob > F        =       0.0000
```

corr(u_i, Xb) = -0.2796

Absorbed variable	Levels
year	24

gini_index	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
gdp_capita	.0660246	.0278079	2.37	0.018	.0114786	.1205705
inf_mrate	.059164	.0145829	4.06	0.000	.0305592	.0877689
unemployment	.0439484	.0127251	3.45	0.001	.0189878	.0689091
_cons	33.84444	.5620178	60.22	0.000	32.74202	34.94685
sigma_u	7.6345941					
sigma_e	2.4516415					
rho	.90651986	(fraction of variance due to u_i)				

F test that all u_i=0: F(143, 1517) = 98.38 Prob > F = 0.0000

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE

Marginal Analysis

Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

MARGINAL ANALYSIS

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE

Marginal Analysis

Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

- Notice that all the variables are in levels . Therefore:

$$E(y_{it}|x_{it}, \alpha_i) = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \alpha_i$$

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

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

- `margins` is a powerful command that allows you to interpret your results after estimation.
- Use `margins` to get the marginal effects (`dydx()`):

```
. xtreg gini_index                                ///
>      gdp_capita inf_mrate unemployment, fe
.
. margins, dydx(*)
```

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE

Marginal Analysis

Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

- Notice that all the variables are in levels . Therefore:

$$E(y_{it}|x_{it}, \alpha_i) = \beta_0 + \beta_1 x_{it1} + \dots + \beta_k x_{itk} + \alpha_i$$

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

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

- margins** is a powerful command that allows you to interpret your results after estimation.
- Use **margins** to get the marginal effects (dydx()):

```
. xtreg gini_index                                ///
>      gdp_capita inf_mrate unemployment, fe
.
. margins, dydx(*)
```


Example 3: GDP per capita interacted with life expectancy

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtreg gini_index c.gdp_capita##c.le_ inf_mrate unemployment, fe
Fixed-effects (within) regression                               Number of obs   =       1,687
Group variable: country                                       Number of groups  =        144
R-squared:                                                    Obs per group:
    Within   = 0.1475                                           min   =           1
    Between  = 0.1482                                           avg   =          11.7
    Overall  = 0.1564                                           max   =           24
                                                                    F(5, 1538)      =       53.21
corr(u_i, Xb) = -0.2274                                       Prob > F         =       0.0000
```

gini_index	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
gdp_capita	-.8819946	.2017265	-4.37	0.000	-1.277683	-.4863066
le_	-.1724262	.0657653	-2.62	0.009	-.3014254	-.0434271
c. gdp_capita# c.le_	.0093218	.0023314	4.00	0.000	.0047487	.0138949
inf_mrate	.1036206	.0194303	5.33	0.000	.0655078	.1417333
unemployment	.019137	.0131347	1.46	0.145	-.0066268	.0449007
_cons	50.02288	5.155652	9.70	0.000	39.91003	60.13573
sigma_u	7.5024458					
sigma_e	2.5094036					
rho	.89938121	(fraction of variance due to u_i)				

F test that all u_i=0: F(143, 1538) = 91.31 Prob > F = 0.0000

Marginal effects at different levels of life_expectancy

```
. margins, dydx(gdp_capita) at(le_=(25 40 60 85))
```

Average marginal effects

Number of obs = 1,687

Model VCE: Conventional

Expression: Linear prediction, predict()

dy/dx wrt: gdp_capita

- 1._at: le_ = 25
- 2._at: le_ = 40
- 3._at: le_ = 60
- 4._at: le_ = 85

		Delta-method				
		dy/dx	std. err.	z	P> z	[95% conf. interval]
gdp_capita						
	_at					
	1	-.6489502	.1440633	-4.50	0.000	-.931309 -.3665914
	2	-.5091235	.1097813	-4.64	0.000	-.724291 -.293956
	3	-.322688	.0652211	-4.95	0.000	-.450519 -.1948569
	4	-.0896435	.0249221	-3.60	0.000	-.1384899 -.0407972

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE

Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

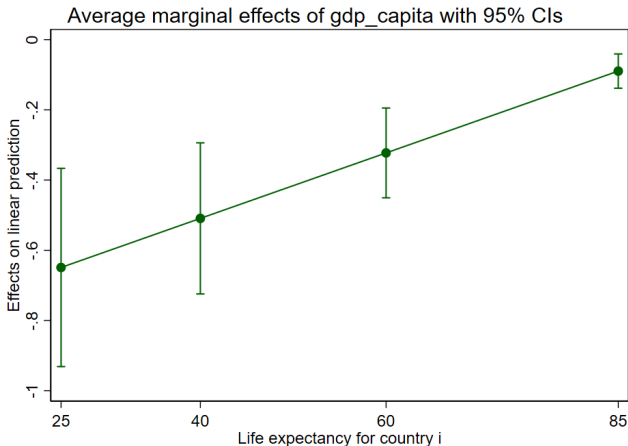
Final remarks

References

Smaller GDP-capita impact for higher life expectancy

```
. marginsplot
```

Variables that uniquely identify margins: le_



How about interacting GDP per capita with unemployment

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE

Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. quietly xtreg gini_index c.gdp_capita##c.unemployment inf_mrate, fe
. margins, dydx(gdp_capita) at(unemployment=(4 25 35 60))
```

Average marginal effects Number of obs = 1,687

Model VCE: Conventional

Expression: Linear prediction, predict()

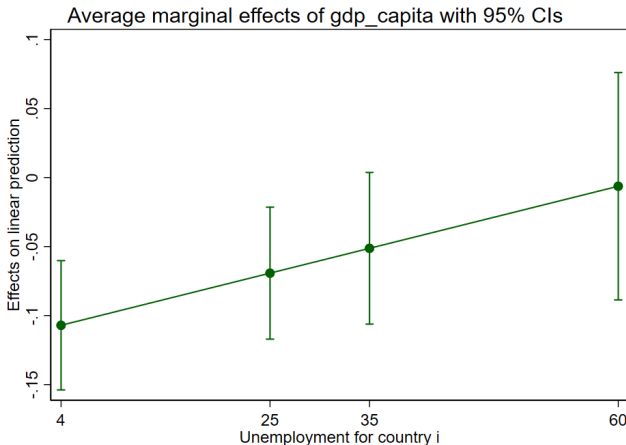
```
dy/dx wrt:  gdp_capita
1._at: unemployment = 4
2._at: unemployment = 25
3._at: unemployment = 35
4._at: unemployment = 60
```

		Delta-method					
		dy/dx	std. err.	z	P> z	[95% conf. interval]	
gdp_capita							
	_at						
	1	-.1070546	.0238881	-4.48	0.000	-.1538744	-.0602348
	2	-.069273	.0243944	-2.84	0.005	-.1170852	-.0214607
	3	-.0512817	.0280249	-1.83	0.067	-.1062096	.0036462
	4	-.0063036	.0420175	-0.15	0.881	-.0886564	.0760492

Smaller GDP-capita impact for higher unemployment

```
. marginsplot
```

Variables that uniquely identify margins: unemployment



Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
**Autocorrelation /
Heteroskedasticity**
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Testing and accounting for serial correlation and heteroskedasticity

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Test for autocorrelation

- Wooldridge (2002, pag. 2823) derives a simple test for autocorrelation in panel-data models.
 - Regress first difference pooled (OLS) model, and predict the residuals
 - Regress the residuals on its own first lag and test the coefficient associated to those lagged residuals
- Drukker (2003) implements the test with the user-written command **xtserial**

```
. xtserial gini_index gdp_capita inf_mrate unemployment  
Wooldridge test for autocorrelation in panel data  
H0: no first-order autocorrelation  
F( 1, 69) = 67.780  
Prob > F = 0.0000
```

Test for heteroskedasticity

- Poi and Wiggins (2001) suggest an LR test for panel-level heteroskedasticity:
 - iterated GLS with heteroskedastic panels produces MLE.
 - Thus, we can use a LR test with **xtgls, igls panels(heteroskdastic)** versus **xtgls, igls**

```
. xtgls gini_index gdp_capita inf_mrate unemployment, ///  
>           igls panel(heterosk) iterate(200)  
. estimates store hetero  
. xtgls gini_index gdp_capita inf_mrate unemployment, ///  
>           igls  
. estimates store homosk  
. local df = e(N_g) - 1  
  
. lrtest hetero homosk,df(`df')  
Likelihood-ratio test  
Assumption: homosk nested within hetero  
LR chi2(143) = 1576.98  
Prob > chi2 = 0.0000
```

Test for heteroskedasticity

- Poi and Wiggins (2001) suggest an LR test for panel-level heteroskedasticity:
 - iterated GLS with heteroskedastic panels produces MLE.
 - Thus, we can use a LR test with **xtgls, igls panels(heteroskdastic)** versus **xtgls, igls**

```
. xtgls gini_index gdp_capita inf_mrate unemployment, ///
>          igls panel(heterosk) iterate(200)
. estimates store hetero
. xtgls gini_index gdp_capita inf_mrate unemployment, ///
>          igls
. estimates store homosk
. local df = e(N_g) - 1

. lrtest hetero homosk,df(`df')

Likelihood-ratio test
Assumption: homosk nested within hetero
LR chi2(143) = 1576.98
Prob > chi2 = 0.0000
```

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Linear model with first order autoregressive error term -xtregar-:

$$Y_{it} = \beta_0 + X_{it}^1 * \beta_1 + \dots + X_{it}^K * \beta_K + \mu_i + \nu_{it}$$

$$\nu_{it} = \rho * \nu_{it-1} + \eta_{it}; \quad \eta \text{ is } iid(0, \sigma_\eta^2)$$

Some Relevant Assumptions

- **Fixed effects:** Non-observable individual effects (μ_i) may be correlated with the covariates in X.
- **Random effects:** μ_i assumed to be independent of ν_i and independent of the covariates in X. $\mu_i \tilde{iid} (0, \sigma_\mu^2)$.

Fit model accounting for autocorrelation

```
. xtregar gini_index gdp_capita inf_mrate unemployment, fe
FE (within) regression with AR(1) disturbances      Number of obs      =      1,543
Group variable: country                            Number of groups    =      140
R-squared:                                          Obs per group:
    Within = 0.1343                                min = 1
    Between = 0.0609                              avg = 11.0
    Overall = 0.0679                              max = 23
                                                    F(3,1400)          =      72.37
                                                    Prob > F            =      0.0000

corr(u_i, Xb) = -0.5738
```

gini_index	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
gdp_capita	.1270717	.0437086	2.91	0.004	.0413302	.2128132
inf_mrate	.588277	.0440945	13.34	0.000	.5017786	.6747754
unemployment	.0854057	.0154952	5.51	0.000	.0550094	.115802
_cons	28.82591	.1634147	176.40	0.000	28.50535	29.14648
rho_ar	.87319678					
sigma_u	11.67754					
sigma_e	1.5626029					
rho_fov	.98240914	(fraction of variance because of u_i)				

F test that all u_i=0: F(139,1400) = 4.25 Prob > F = 0.0000

Feasible Generalized Least Squares (**xtgls**):

$$Y_{it} = \beta_0 + X_{it}^1 * \beta_1 + \dots + X_{it}^K * \beta_K + \epsilon_{it}$$

Where the variance matrix of the disturbances would be:

$$E[\epsilon\epsilon'] = \Omega = \begin{pmatrix} \sigma_{11}\Omega_{11} & \sigma_{12}\Omega_{12} & \cdots & \sigma_{1m}\Omega_{1m} \\ \sigma_{21}\Omega_{21} & \sigma_{22}\Omega_{22} & \cdots & \sigma_{2m}\Omega_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{m1}\Omega_{m1} & \sigma_{m2}\Omega_{m2} & \cdots & \sigma_{mm}\Omega_{mm} \end{pmatrix}$$

xtgls supports the variance-covariance structures

- Heteroskedasticity across panels
- Correlation across panels
- Autocorrelation within panels

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Feasible Generalized Least Squares `xtgls`:

$$Y_{it} = \beta_0 + X_{it}^1 * \beta_1 + \dots + X_{it}^K * \beta_K + \epsilon_{it}$$

- Heteroskedasticity across panels

`xtgls, panels(heteroskedastic)`:

$$E[\epsilon\epsilon] = \Omega = \begin{pmatrix} \sigma_1 I & 0 & \dots & 0 \\ 0 & \sigma_2 I & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_m I \end{pmatrix}$$

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Feasible Generalized Least Squares `xtgls`:

$$Y_{it} = \beta_0 + X_{it}^1 * \beta_1 + \dots + X_{it}^K * \beta_K + \epsilon_{it}$$

$$\epsilon_{it} = \rho * \epsilon_{it-1} + \eta_{it} \quad ; \quad \eta \text{ is } iid(0, \sigma_\eta^2)$$

- Heteroskedasticity across panels and autocorrelation within panels
`xtgls, panels(heteroskedastic) corr(psar1)`:

$$E[\epsilon\epsilon] = \Omega = \begin{pmatrix} \sigma_1 \Omega_{11} & 0 & \dots & 0 \\ 0 & \sigma_2 \Omega_{22} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_m \Omega_{mm} \end{pmatrix}$$

Fit model accounting for heteroskedasticity

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis

Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtgls gini_index gdp_capita inf_mrate unemployment, ///
> panels(heterosk) nolog
Cross-sectional time-series FGLS regression
Coefficients: generalized least squares
Panels: heteroskedastic
Correlation: no autocorrelation

Estimated covariances      =      144      Number of obs      =      1,687
Estimated autocorrelations =      0      Number of groups   =      144
Estimated coefficients     =      4      Obs per group:
                                         min =      1
                                         avg  = 11.71528
                                         max  =      24
                                         Wald chi2(3) = 1711.30
                                         Prob > chi2   = 0.0000
```

gini_index	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
gdp_capita	-.0847204	.0042164	-20.09	0.000	-.0929845	-.0764563
inf_mrate	.1095668	.0062672	17.48	0.000	.0972834	.1218503
unemployment	.0425709	.0056975	7.47	0.000	.031404	.0537378
_cons	35.26954	.2347229	150.26	0.000	34.80949	35.72959

Fit model accounting for autocorrelation and heteroskedasticity

```
. xtgls gini_index gdp_capita inf_mrate unemployment, ///
>      panels(heterosk) corr(psarl) nolog force
(note: 4 observations dropped because only 1 obs in group)
```

Cross-sectional time-series FGLS regression

Coefficients: generalized least squares

Panels: heteroskedastic

Correlation: panel-specific AR(1)

Estimated covariances	=	140	Number of obs	=	1,683
Estimated autocorrelations	=	140	Number of groups	=	140
Estimated coefficients	=	4	Obs per group:		
			min	=	2
			avg	=	12.02143
			max	=	24
			Wald chi2(3)	=	6183.07
			Prob > chi2	=	0.0000

gini_index	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
gdp_capita	-.1014412	.005435	-18.66	0.000	-.1120937	-.0907887
inf_mrate	.0984335	.0023165	42.49	0.000	.0938933	.1029737
unemployment	.0350265	.0017579	19.93	0.000	.0315811	.0384719
_cons	36.33825	.1288917	281.93	0.000	36.08562	36.59087

Outline

Basic
concepts

Pooled vs. Panel

Stata tools

Linear PD

DGP

Data and model

Panel structure

Random Effects

Fixed Effects

FE vs RE

Marginal Analysis

Autocorrelation /
Heteroskedasticity

IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models

Logit Models

Poisson Models

Final remarks

References

Instrumental Variables

Instrumental Variables Definition

- Let's look at the linear cross-sectional model:

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \varepsilon_i$$

$$E(x_{i1}\varepsilon_i) = E(x_{i2}\varepsilon_i) = 0$$

- The definition of endogeneity implies:

$$E(x_{i1}\varepsilon_i) = 0$$

$$E(x_{i2}\varepsilon_i) \neq 0$$

- Letting $z_i \equiv (x_{i1}, z_{i1}, \dots, z_{ip})$ a set of models arises from:

$$E(z_i\varepsilon_i) = 0 \tag{3}$$

$$E(x_{i2}|z_i) = z_i\pi \tag{4}$$

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

- From condition (3) the following estimator arises:

$$\hat{\beta}_{gmm} = (X'ZWZ'X)^{-1} X'ZWZ'y$$

Where: W is a weighting matrix

- From condition (4) a two stage estimation arises (2SLS):

$$P_Z = Z(Z'Z)^{-1}Z'$$

$$\hat{\beta}_{2sls} = (X'P_ZX)^{-1}X'P_Zy$$

- A reliable instrumental variables estimator should satisfy:

- 1 $E(z_i \varepsilon_i) = 0$
- 2 The relation between the instruments and the regressors should be strong (weak instruments)

Panel data instrumental variables estimation

- Going back to our linear panel data specification:

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

- When we studied FE versus RE we analyzed the assumption:

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

- We now look at the violation of the following assumption:

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

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

FE and RE estimation with instrumental variables

- **xtivreg, fe** implements 2SLS using the variables after performing the within transformation."
- **xtivreg, re** implements a couple of 2SLS estimators on two alternative FGLS transformations.

Example 4: IV estimation for the FE (gini_index) model

- Endogenous: gdp_capita. Instruments: unemployment, le

```
. xtivreg gini_index inf_mrate (gdp_capita = unemployment le), fe
```

Fixed-effects (within) IV regression Number of obs = 1,687
Group variable: country Number of groups = 144
R-squared: Obs per group: min = 1
 Within = 0.1257 avg = 11.7
 Between = 0.1426 max = 24
 Overall = 0.1668 Wald chi2(2) = 352700.81
corr(u_i, Xb) = -0.3048 Prob > chi2 = 0.0000

gini_index	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
gdp_capita	-.1897486	.0492316	-3.85	0.000	-.2862408	-.0932565
inf_mrate	.1401105	.0116751	12.00	0.000	.1172276	.1629933
_cons	37.83177	.9560125	39.57	0.000	35.95802	39.70552
sigma_u	7.6751692					
sigma_e	2.5387423					
rho	.90137913	(fraction of variance due to u_i)				
F test that all u_i=0: F(143,1541) =			91.23	Prob > F	= 0.0000	

Endogenous: gdp_capita
Exogenous: inf_mrate unemployment le_

```
. estimates store xtivfe
```

How about the IV random effects estimation

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects

FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtivreg gini_index inf_mrate (gdp_capita = unemployment le),re
G2SLS random-effects IV regression      Number of obs      =      1,687
Group variable: country                  Number of groups    =      144
R-squared:                               Obs per group:
      Within = 0.1256                      min = 1
      Between = 0.1431                     avg = 11.7
      Overall = 0.1663                     max = 24
                                           Wald chi2(2)        =      247.13
                                           Prob > chi2         =      0.0000
corr(u_i, X) = 0 (assumed)
```

gini_index	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
gdp_capita	-.1920802	.0444252	-4.32	0.000	-.279152	-.1050084
inf_mrate	.1294964	.0114243	11.34	0.000	.1071052	.1518877
_cons	36.76172	1.037219	35.44	0.000	34.72881	38.79463
sigma_u	9.2895161					
sigma_e	2.5387423					
rho	.93050259	(fraction of variance due to u_i)				

```
Endogenous: gdp_capita
Exogenous: inf_mrate unemployment le_
```

```
. estimates store xtivre
```

Let's compare the results from FE and RE

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. estimates table xtivfe xtivre, star
```

Variable	xtivfe	xtivre
gdp_capita	-.18974862***	-.1920802***
inf_mrate	.14011045***	.12949643***
_cons	37.831771***	36.761721***

Legend: * p<0.05; ** p<0.01; *** p<0.001

The Hausman test favors the FE estimation

```
. hausman xtivfe xtivre
```

	Coefficients		(b-B) Difference	sqrt (diag (V_b-V_B)) Std. err.
	(b) xtivfe	(B) xtivre		
gdp_capita	-.1897486	-.1920802	.0023316	.0212168
inf_mrate	.1401105	.1294964	.010614	.0024071

```
                b = Consistent under H0 and Ha; obtained from xtivreg
                B = Inconsistent under Ha, efficient under H0; obtained from xtivreg

Test of H0: Difference in coefficients not systematic

      chi2(2) = (b-B) ' [ (V_b-V_B) ^ (-1) ] (b-B)
              = 21.30
Prob > chi2 = 0.0000
```

- Outline
- Basic concepts
 - Pooled vs. Panel
 - Stata tools
- Linear PD
 - DGP
 - Data and model
 - Panel structure
 - Random Effects
 - Fixed Effects
 - FE vs RE
 - Marginal Analysis
 - Autocorrelation / Heteroskedasticity
 - IV Regression
- Dynamic PD Models
 - Arellano/Bond
- Panel VAR models
- Nonlinear PD Models
 - Probit Models
 - Logit Models
 - Poisson Models
- Final remarks
- References

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

DYNAMIC PANEL-DATA MODELS

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

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:

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

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

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:

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

Instrumental Variable (GMM) Estimation

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

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

- This gives rise to the following models:
 - Anderson-Hsiao (1981) use the instruments $y_{i(t-2)}$ and $\Delta y_{i(t-2)}$ (**xtivreg, fd**).
 - Arellano-Bond (1991) suggest using all available lag levels (not only the second lag) for the first difference equation (**xtabond**).

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Example 5: Model for aggregate consumption

$$consumption_{it} = \beta_0 + gdp_capita_{it} * \beta_1 + irate_{it} * \beta_3 + \mu_i + \nu_{it}$$

Data

- World Bank public online data on:
consumption: Consumption expenditure (2015 US\$)
gdp: Gross domestic product per capita (2015 US\$)
irate: deposit interest rate
- Example : 2010-2024 for 123 countries
- Source: <http://databank.worldbank.org/data/Home.aspx>

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. describe
Contains data from C:\Users\gas\Documents\conferences\paa\data\consumption_gdpc
> apita_paa.dta
Observations:      1,514
Variables:          10                               25 Mar 2026 15:40
```

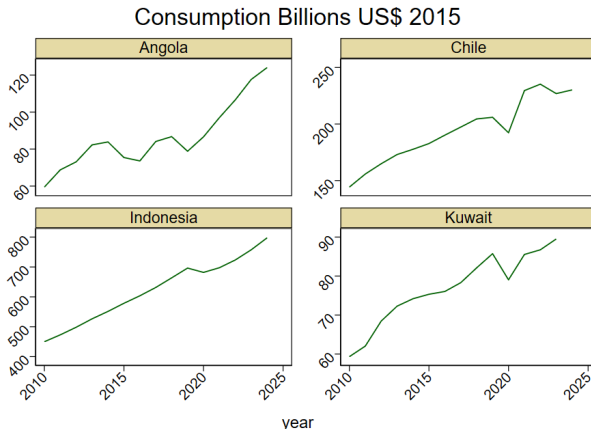
Variable name	Storage type	Display format	Value label	Variable label
country	long	%73.0g	country	Country Name
year	float	%8.0g		
consumption	double	%10.0g		Consumption (Billions 2015 US\$)
gdp	double	%10.0g		GDP (Billions 2015 US\$)
irate	double	%10.0g		Deposit interest rate
gdp_capita	double	%10.0g		GDP p/capita (Thousands 2015 US\$)
ln_cons	float	%9.0g		Log of consumption
ln_gdp	float	%9.0g		Log of gdp
ln_irate	float	%9.0g		Log of irate
ln_gdp_capita	float	%9.0g		Log of gdp_capita

```
Sorted by: country year
Note: Dataset has changed since last saved.
```

```
.
.
. xtset country year
Panel variable: country (unbalanced)
Time variable: year, 2010 to 2024, but with gaps
Delta: 1 unit
```

Upward trended series for consumption

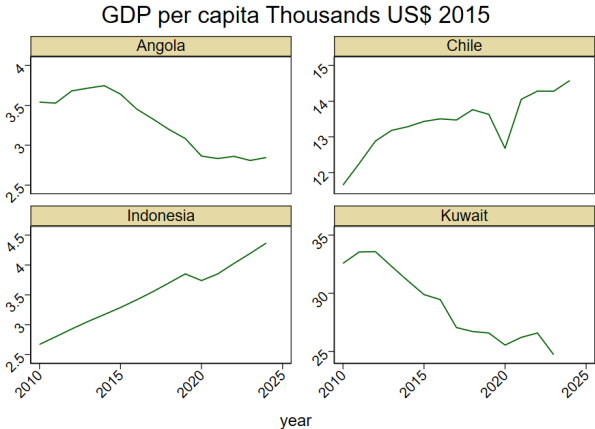
- I(1) series -> Specify first difference for this variable.



Source: <https://data.worldbank.org/indicator/NE.CON.TOTL.KD>

Trended series also for GDP per capita

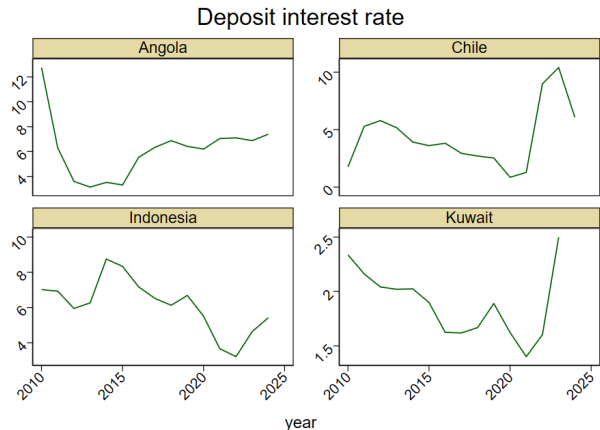
- $I(1)$ series -> Also first difference for this variable.



Source: <https://data.worldbank.org/indicator/NY.GDP.PCAP.KD>

We don't observe general trends in the interest rates

- I(0) series, we will include the level of this variable.



Source: <https://data.worldbank.org/indicator/FR.INR.DPST>

Let's start with a fixed effects model

- First difference of the logs for `ln_cons` and `gdp_capita`.

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtreg D.ln_cons LD.ln_cons D.ln_gdp_capita irate, fe
Fixed-effects (within) regression      Number of obs   =      1,256
Group variable: country                Number of groups =      123
R-squared:                             Obs per group:
    Within   = 0.2922                      min =      2
    Between  = 0.4110                      avg  =     10.2
    Overall  = 0.3001                      max  =     13
                                         F(3, 1130)      =    155.52
corr(u_i, Xb) = 0.0492                  Prob > F        =    0.0000
```

D.ln_cons	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
ln_cons						
LD.	.0380762	.0250794	1.52	0.129	-.0111312	.0872836
ln_gdp_capita						
D1.	.4223618	.0195618	21.59	0.000	.3839804	.4607432
irate	.0001112	.0003141	0.35	0.723	-.000505	.0007274
_cons	.0235804	.0020796	11.34	0.000	.0195002	.0276606
sigma_u	.01874881					
sigma_e	.03941836					
rho	.18449241	(fraction of variance due to u_i)				

F test that all u_i=0: F(122, 1130) = 1.37 Prob > F = 0.0072

Now fit the model with the Arellano/Bond estimator

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE

Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtabond D.ln_cons D.ln_gdp_capita irate,twostep maxldep(3)
Arellano-Bond dynamic panel-data estimation      Number of obs   =       1,132
Group variable: country                          Number of groups  =        123
Time variable: year

Obs per group:
      min =          1
      avg =     9.203252
      max =          12

Number of instruments =        36                Wald chi2(3)      =       280.62
                                                Prob > chi2       =       0.0000

Two-step results
```

D.ln_cons	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
ln_cons						
LD.	.0467469	.016675	2.80	0.005	.0140645	.0794293
ln_gdp_cap_a						
D1.	.6406092	.042884	14.94	0.000	.5565581	.7246604
irate	-.0004106	.000229	-1.79	0.073	-.0008594	.0000382
_cons	.0210111	.0017985	11.68	0.000	.0174862	.024536

Warning: gmm two-step standard errors are biased; robust standard errors are recommended.

Instruments for differenced equation

GMM-type: L(2/4).D.ln_cons

Standard: D2.ln_gdp_capita D.irate

Instruments for level equation

Standard: _cons

Use Sargan test to check the model specification

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. estat sargan
```

Sargan test of overidentifying restrictions
H0: Overidentifying restrictions are valid

```
chi2(32)          = 45.58921  
Prob > chi2       = 0.0565
```

- The result supports the null hypothesis.

Use `estat abond` to test for serial correlation

```
. estat abond
```

Arellano-Bond test for zero autocorrelation in first-differenced errors
H0: No autocorrelation

Order	z	Prob > z
1	-3.3035	0.0010
2	.23011	0.8180

- The Arellano-Bond test is testing that $H_0: E[\Delta\varepsilon_{it}\Delta\varepsilon_{i(t-1)}] \neq 0$:

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

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

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

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

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:

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

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

- These are defined by Arellano-Bover (1995) / Blundell-Bond (1998).
- 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:

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

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

- These are defined by Arellano-Bover (1995) / Blundell-Bond (1998).
- Notice that you have moments for the equation in levels and for the equation in first difference
- Fit this model with **xtdpdsys**

We now use the Blundell-Bond system GMM estimator

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects

Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtdpdsys D.ln_cons D.ln_gdp_capita irate,twostep maxldep(3)

System dynamic panel-data estimation      Number of obs      =      1,256
Group variable: country                  Number of groups    =      123
Time variable: year

Obs per group:
      min =      2
      avg =    10.21138
      max =     13

Number of instruments =      48           Wald chi2(3)        =     447.47
                                           Prob > chi2         =     0.0000
```

Two-step results

D.ln_cons	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
ln_cons						
LD.	.040489	.0117938	3.43	0.001	.0173735	.0636044
ln_gdp_capita						
D1.	.635923	.0369796	17.20	0.000	.5634443	.7084017
irate	-.0006207	.0002539	-2.44	0.014	-.0011184	-.0001231
_cons	.0219494	.0017821	12.32	0.000	.0184567	.0254422

Warning: gmm two-step standard errors are biased; robust standard errors are recommended.

Instruments for differenced equation

GMM-type: L(2/4).D.ln_cons

Standard: D2.ln_gdp_capita D.irate

Instruments for level equation

GMM-type: LD2.ln_cons

Standard: _cons

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Let's check autocorrelation and overidentification

```
. estat abond
```

Arellano-Bond test for zero autocorrelation in first-differenced errors
H0: No autocorrelation

Order	z	Prob > z
1	-3.2666	0.0011
2	.20423	0.8382

```
.
```

```
.
```

```
. estat sargan
```

Sargan test of overidentifying restrictions
H0: Overidentifying restrictions are valid

```
chi2(44)      =    51.0574  
Prob > chi2   =    0.2160
```

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

- Generalizes time-series VAR models to panel data
- Generalizes dynamic panel data to multiple outcomes
- A panel VAR model can be written as:

$$\mathbf{y}_{it} = \mathbf{A}_1 \mathbf{y}_{it-1} + \dots + \mathbf{A}_p \mathbf{y}_{it-p} + \mathbf{B} \mathbf{x}_{it} + \mathbf{C} \mathbf{w}_{it} + \mathbf{D} \mathbf{v}_{it} + \mathbf{u}_i + \varepsilon_{it}$$

Where:

 \mathbf{y}_{it} is a $K \times 1$ vector of dependent variables; \mathbf{A}_1 through \mathbf{A}_p are $K \times K$ matrices of parameters; \mathbf{x}_{it} is an $M_1 \times 1$ vector of strictly exogenous variables; \mathbf{B} is a $K \times M_1$ matrix of parameters; \mathbf{w}_{it} is an $M_2 \times 1$ vector of endogenous variables; \mathbf{C} is a $K \times M_2$ matrix of parameters; \mathbf{v}_{it} is a $M_3 \times 1$ vector of predetermined variables; \mathbf{D} is a $K \times M_3$ matrix of parameters; \mathbf{u}_i is a $K \times 1$ fixed effects vector; ε_{it} is a $K \times 1$ vector of serially uncorrelated errors; $i=1, \dots, N$ denotes the i th panel; $t=1, \dots, T$ denotes the t th time period

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

GMM estimator, assumptions similar to Arellano-Bond

- The asymptotics rely on a large number of panels.
- The more time periods,
 - The larger the number of moment conditions.
 - The number of instruments increases dramatically.
 - A number of instruments could be weakly relevant
- In practice, the number of time periods is generally not large when using this estimator.
- You may need to use options `maxldep()` and `collapse` to reduce the number of instruments.

Example 6: Panel VAR for consumption and gdp_capita

```
. xtvar D.ln_cons D.ln_gdp, maxldep(2) pre(irate) ///
> lags(2) collapse fd
note: 2 panels dropped due to no within-panel variation in irate.
note: the estimation sample has gaps; see xtdescribe.
```

Panel-data vector autoregression
Group variable: country
Time variable: year

Number of obs = 996
Number of groups = 118
Obs per group:
min = 1
avg = 8.4
max = 11

Number of moment conditions = 14 (collapsed)

Fixed-effects transform: FD
Two-step results

(Std. err. adjusted for 118 clusters in country)						
	Coefficient	WC robust std. err.	z	P> z	[95% conf. interval]	
D.ln_cons						
ln_cons						
LD.	.1157647	.0700688	1.65	0.099	-.0215676	.2530971
L2D.	-.0142077	.0608584	-0.23	0.815	-.133488	.1050726
ln_gdp						
LD.	.0236653	.0421092	0.56	0.574	-.0588673	.1061978
L2D.	.0877439	.0450175	1.95	0.051	-.0004887	.1759765
irate	-.0086637	.0028626	-3.03	0.002	-.0142744	-.0030531
D.ln_gdp						
ln_cons						
LD.	-.0640195	.1385575	-0.46	0.644	-.3355873	.2075483
L2D.	-.1419558	.1596415	-0.89	0.374	-.4548474	.1709358
ln_gdp						
LD.	.1505274	.1688794	0.89	0.373	-.1804701	.4815248
L2D.	.2936375	.2784721	1.05	0.292	-.2521577	.8394328
irate	-.0091913	.0031592	-2.91	0.004	-.0153832	-.0029994

Hansen's test of overid. restrictions: chi2(4) = 3.63 Prob > chi2 = 0.458
Added predetermined: irate
GMM-type instruments: L(2/3). (D.ln_cons D.ln_gdp)
L(1/3).irate

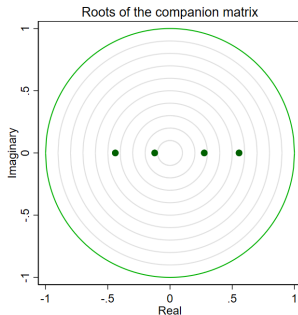
Check the model stability

```
. varstable, graph
```

Eigenvalue stability condition

Eigenvalue	Modulus
.5546634	.554663
-.4397675	.439767
.2749245	.274924
-.1235283	.123528

All the eigenvalues lie inside the unit circle.
VAR satisfies stability condition.



Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
Models

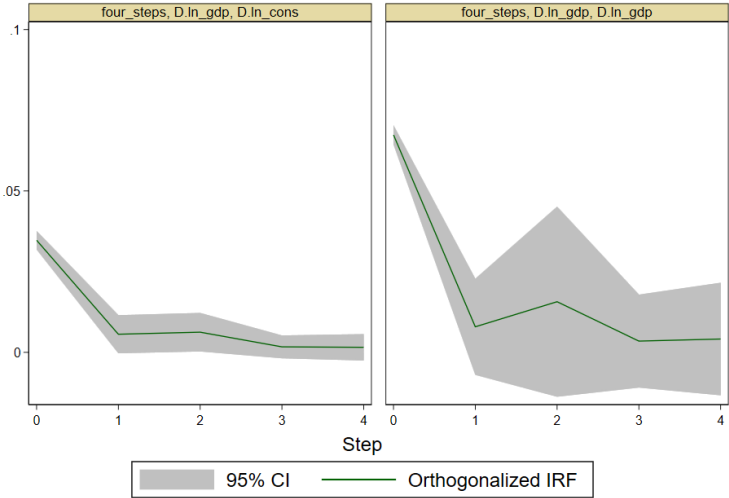
Probit Models
Logit Models
Poisson Models

Final remarks

References

Orthogonalized impulse-response functions

```
. irf graph oirf, impulse(D.ln_gdp)
```



Graphs by irfname, impulse variable, and response variable

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

NONLINEAR PANEL-DATA MODELS

Probit and Logit Models for the Cross-Sectional Case

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

- The data generating process is given by:

$$y_i = \begin{cases} 1 & \text{if } y_i^* = x_i' \beta + \varepsilon_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

- By construction $P(y_i = 1|x_i) = F(x_i' \beta)$.
 - This is the distribution of the random disturbance ε
- This implies that:

$$\begin{aligned} E(y_i|x_i) &= 1P(y_i = 1|x_i) + 0P(y_i = 0|x_i) \\ &= F(x_i' \beta) \end{aligned}$$

- This gives rise to two models:
 - If $F(\cdot)$ is the standard normal distribution we have a **Probit**
 - If $F(\cdot)$ is the logistic distribution we have a **Logit** model

Probit and Logit Models for the Cross-Sectional Case

- The data generating process is given by:

$$y_i = \begin{cases} 1 & \text{if } y_i^* = x_i' \beta + \varepsilon_i > 0 \\ 0 & \text{otherwise} \end{cases}$$

- By construction $P(y_i = 1|x_i) = F(x_i' \beta)$.
 - This is the distribution of the random disturbance ε
- This implies that:

$$\begin{aligned} E(y_i|x_i) &= 1P(y_i = 1|x_i) + 0P(y_i = 0|x_i) \\ &= F(x_i' \beta) \end{aligned}$$

- This gives rise to two models:
 - 1 If $F(\cdot)$ is the standard normal distribution we have a **Probit**
 - 2 If $F(\cdot)$ is the logistic distribution we have a **Logit** model

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

Example 7: Probability to visit the doctor

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. describe
Contains data from C:\Users\gas\Documents\conferences\paa\data\German_healthcar
> e_probit.dta
Observations:      27,326
Variables:          6
15 Apr 2026 21:17
```

Variable name	Storage type	Display format	Value label	Variable label
id	int	%8.0g		person - identification number
year	int	%8.0g		calendar year of the observation
doctor	byte	%8.0g		1 if visited doctor ; 0 otherwise
hhkids	byte	%8.0g		1 if kids<16 in hhold;0 otherwise
age	byte	%8.0g		age in years (25 - 64)
income	float	%9.0g		Monthly Income (German marks/1000)

Sorted by: id year

Source:<https://pages.stern.nyu.edu/~wgreene/Econometrics/PanelDataSets.htm>

Estimating a pooled probit model

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. probit doctor age income i.hhkids
```

Iteration 0: Log likelihood = -18019.552

Iteration 1: Log likelihood = -17708.885

Iteration 2: Log likelihood = -17708.685

Iteration 3: Log likelihood = -17708.685

Probit regression

Number of obs = 27,326
LR chi2(3) = 621.73
Prob > chi2 = 0.0000
Pseudo R2 = 0.0173

Log likelihood = -17708.685

doctor	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
age	.0145427	.0007398	19.66	0.000	.0130926	.0159928
income	-.1944154	.0437552	-4.44	0.000	-.280174	-.1086569
1.hhkids	-.117389	.0167616	-7.00	0.000	-.1502412	-.0845368
_cons	-.1807133	.038602	-4.68	0.000	-.2563718	-.1050549

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Margins

- In non-linear models the coefficients of the model are usually not the object of interest.
- Before we move forward we will analyze how **margins** works in this case.

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

Interpreting the Results: Marginal Effects

- The change in the conditional expectation due to a change in a covariate is given by

$$\begin{aligned}\frac{\partial E(y_i|x_i)}{\partial x_{ik}} &= \frac{\partial F(x_i'\beta)}{\partial x_{ik}} \beta_k \\ &= f(x_i'\beta) \beta_k\end{aligned}$$

- This implies that:
 - The value of the object of interest depends on x , conventionally \bar{x}
 - The β coefficients only tell us the sign of the effect given that $f(x_i'\beta) > 0$ almost surely
- For a categorical variable (factor variables)

$$F(x_i'\beta + \beta_{treatment}) - F(x_i'\beta)$$

Marginal effects at the mean

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation / Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. margins, dydx(age income) atmeans
```

Conditional marginal effects

Number of obs = 27,326

Model VCE: OIM

Expression: Pr(doctor), predict()

dy/dx wrt: age income

At: age = 43.52569 (mean)
income = .3521352 (mean)
0.hhkids = .59727 (mean)
1.hhkids = .40273 (mean)

	Delta-method					
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
age	.0054823	.0002786	19.68	0.000	.0049362	.0060284
income	-.0732906	.0164935	-4.44	0.000	-.1056173	-.0409638

```
. estat summarize
```

Estimation sample probit

Number of obs = 27,326

Variable	Mean	Std. dev.	Min	Max
doctor	.6291078	.4830525	0	1
age	43.52569	11.33025	25	64
income	.3521352	.176857	.0015	3.0671
1.hhkids	.40273	.4904563	0	1

Marginal effects at the mean

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. margins, dydx(age income) atmeans
```

Conditional marginal effects

Number of obs = 27,326

Model VCE: OIM

Expression: Pr(doctor), predict()

dy/dx wrt: age income

At: age = 43.52569 (mean)
income = .3521352 (mean)
0.hhkids = .59727 (mean)
1.hhkids = .40273 (mean)

	Delta-method					
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
age	.0054823	.0002786	19.68	0.000	.0049362	.0060284
income	-.0732906	.0164935	-4.44	0.000	-.1056173	-.0409638

```
. estat summarize
```

Estimation sample probit

Number of obs = 27,326

Variable	Mean	Std. dev.	Min	Max
doctor	.6291078	.4830525	0	1
age	43.52569	11.33025	25	64
income	.3521352	.176857	.0015	3.0671
1.hhkids	.40273	.4904563	0	1

Get the marginal effects at the mean manually

- Matrix from `estat summarize`

```
. matrix list r(stats)
r(stats) [4,4]
```

	mean	sd	min	max
doctor	.62910781	.48305249	0	1
age	43.52569	11.330248	25	64
income	.35213516	.17685695	.0015	3.0671
1.hhkids	.40273	.49045627	0	1

- Get the predicted linear combination

```
. scalar xb = _b[_cons] + _b[age]*r(stats)[2,1] ///
>               + _b[income]*r(stats)[3,1] ///
>               + _b[1.hhkids]*r(stats)[4,1]
```

- Plug into the marginal effects formula

```
. display _n(2) _col(8) "d(y)/dxi"          normalden(xb)*_b[xi]" ///
>           _n _col(11) "age" _col(28) normalden(xb)*_b[age] ///
>           _n _col(9) "income" _col(27) normalden(xb)*_b[income]
```

	d(y)/dxi	normalden(xb)*_b[xi]
age		.00548229
income		-.07329059

Get the marginal effects at the mean manually

- Matrix from `estat summarize`

```
. matrix list r(stats)
r(stats) [4,4]
```

	mean	sd	min	max
doctor	.62910781	.48305249	0	1
age	43.52569	11.330248	25	64
income	.35213516	.17685695	.0015	3.0671
1.hhkids	.40273	.49045627	0	1

- Get the predicted linear combination

```
. scalar xb = _b[_cons] + _b[age]*r(stats)[2,1] ///
> + _b[income]*r(stats)[3,1] ///
> + _b[1.hhkids]*r(stats)[4,1]
```

- Plug into the marginal effects formula

```
. display _n(2) _col(8) "d(y)/dxi" normalden(xb)*_b[xi]" ///
> _n _col(11) "age" _col(28) normalden(xb)*_b[age] ///
> _n _col(9) "income" _col(27) normalden(xb)*_b[income]
```

	d(y)/dxi	normalden(xb)*_b[xi]
age		.00548229
income		-.07329059

Get the marginal effects at the mean manually

- Matrix from `estat summarize`

```
. matrix list r(stats)
r(stats) [4,4]
```

	mean	sd	min	max
doctor	.62910781	.48305249	0	1
age	43.52569	11.330248	25	64
income	.35213516	.17685695	.0015	3.0671
1.hhkids	.40273	.49045627	0	1

- Get the predicted linear combination

```
. scalar xb = _b[_cons] + _b[age]*r(stats)[2,1] ///
>               + _b[income]*r(stats)[3,1] ///
>               + _b[1.hhkids]*r(stats)[4,1]
```

- Plug into the marginal effects formula

```
. display _n(2) _col(8) "d(y)/dxi"          normalden(xb)*_b[xi]" ///
>           _n _col(11) "age" _col(28) normalden(xb)*_b[age] ///
>           _n _col(9) "income" _col(27) normalden(xb)*_b[income]
```

	d(y)/dxi	normalden(xb)*_b[xi]
age		.00548229
income		-.07329059

Marginal effects for the categorical variable: hhkids

```
. quietly probit doctor age income i.hhkids  
. margins, dydx(1.hhkids) atmeans noatlegend
```

Conditional marginal effects
Model VCE: OIM

Number of obs = 27,326

Expression: Pr(doctor), predict()
dy/dx wrt: 1.hhkids

	Delta-method					[95% conf. interval]
	dy/dx	std. err.	z	P> z		
1.hhkids	-.0443979	.0063558	-6.99	0.000	-.056855	-.0319408

Note: dy/dx for factor levels is the discrete change from the base level.

Manual computation: dydx for hhkids

```
. quietly estat summarize  
. scalar xbl = _b[_cons] + _b[age] *r(stats)[2,1] ///  
> + _b[income] *r(stats)[3,1] ///  
> + _b[1.hhkids]*1  
. scalar xb0 = _b[_cons] + _b[age] *r(stats)[2,1] ///  
> + _b[income] *r(stats)[3,1] ///  
> + _b[1.hhkids]*0  
. display _n(2) "d(y)/dx for hhkids: "normal(xbl) - normal(xb0)  
d(y)/dx for hhkids: -.04439789
```

Marginal effects for the categorical variable: hhkids

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

```
. quietly probit doctor age income i.hhkids
. margins, dydx(1.hhkids) atmeans noatlegend
```

Conditional marginal effects

Number of obs = 27,326

Model VCE: OIM

```
Expression: Pr(doctor), predict()
dy/dx wrt: 1.hhkids
```

	Delta-method				
	dy/dx	std. err.	z	P> z	[95% conf. interval]
1.hhkids	-.0443979	.0063558	-6.99	0.000	-.056855 -.0319408

Note: dy/dx for factor levels is the discrete change from the base level.

Manual computation: dydx for hhkids

```
. quietly estat summarize
. scalar xb1 = _b[_cons] + _b[age] *r(stats)[2,1] ///
> + _b[income] *r(stats)[3,1] ///
> + _b[1.hhkids]*1
.
. scalar xb0 = _b[_cons] + _b[age] *r(stats)[2,1] ///
> + _b[income] *r(stats)[3,1] ///
> + _b[1.hhkids]*0
. display _n(2) "d(y)/dx for hhkids: "normal(xb1) - normal(xb0)
d(y)/dx for hhkids: -.04439789
```

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Interpreting the Results: Marginal Effects II

- If you do not specify the option **atmeans** you are getting the average marginal effect.
- For a change in x_{ik} this is equal to:

$$\frac{1}{N} \sum_{i=1}^N f(x_i' \beta) \beta_k$$

- Before, we were getting the effect for the average person
- If we do not specify **atmeans** we are getting the average effect over the sample

```
. quietly probit doctor age income i.hhkids
. margins, dydx(age income)
```

Average marginal effects Number of obs = 27,326
Model VCE: OIM

Expression: Pr(doctor), predict()
dy/dx wrt: age income

	Delta-method					
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
age	.0053918	.0002684	20.09	0.000	.0048657	.0059179
income	-.0720812	.0162049	-4.45	0.000	-.1038423	-.0403201

Manual computation for dydx with predict, xb

```
. predict xb, xb
. generate double me_age = normalden(xb)*_b[age]
. generate double me_income = normalden(xb)*_b[income]

. summarize me_age me_income
```

Variable	Obs	Mean	Std. dev.	Min	Max
me_age	27,326	.0053918	.0003821	.0044175	.0058017
me_income	27,326	-.0720812	.0051085	-.0775605	-.0590554

Average Marginal Effects

```
. quietly probit doctor age income i.hhkids
. margins, dydx(age income)
```

Average marginal effects Number of obs = 27,326
Model VCE: OIM

Expression: Pr(doctor), predict()
dy/dx wrt: age income

	Delta-method					
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
age	.0053918	.0002684	20.09	0.000	.0048657	.0059179
income	-.0720812	.0162049	-4.45	0.000	-.1038423	-.0403201

Manual computation for dydx with predict, xb

```
. predict xb, xb
. generate double me_age = normalden(xb)*_b[age]
. generate double me_income = normalden(xb)*_b[income]
```

```
. summarize me_age me_income
```

Variable	Obs	Mean	Std. dev.	Min	Max
me_age	27,326	.0053918	.0003821	.0044175	.0058017
me_income	27,326	-.0720812	.0051085	-.0775605	-.0590554

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Panel-Data for Binary Latent Dependent Variables

$$\begin{aligned}y_{it}^* &= x_{it}'\beta_{it} + z_i'\gamma + \alpha_i + \varepsilon_{it} \\ &= w_{it}'\delta + \alpha_i + \varepsilon_{it}\end{aligned}$$

$$y_{it} = \begin{cases} 1 & \text{if } y_{it}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

- As in the linear panel models
 - if α_i is related to x_{it} or z_i , it is known as a fixed effect
 - if α_i is unrelated to x_{it} and z_i , it is known as a random effect

General Comments

- For fixed effects models:
 - We require a way of dealing with α_i . This is feasible for logit models.
 - The Mundlak approach is an alternative in this scenario.
 - Partial Effects cannot be estimated given our model specification, unless we assume fixed effects equal to zero
- For random effects models
 - Integrate the unobserved component
 - We can estimate partial effects
- We can once more test random vs. fixed effects

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

xtprobit

xtprobit *depvar varlist*, [{**re**|**pa**} **quad**(#)]

- The **re** and **pa** options stand for random effects and population averaged models
- The **quad**() option specifies the number of quadrature points to be used in the numerical integration
- **quadcheck** Re-estimates the model for two different values for **quad**

Example 8: Panel probit regression for doctor visits

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. describe
Contains data from C:\Users\gas\Documents\conferences\paa\data\German_healthcar
> e_xtprobit.dta
Observations:      27,326
Variables:          11
3 Apr 2026 10:13
```

Variable name	Storage type	Display format	Value label	Variable label
id	int	%8.0g		person - identification number
year	int	%8.0g		calendar year of the observation
doctor	byte	%8.0g		1 if visited doctor ; 0 otherwise
female	byte	%8.0g		1 if female ; 0 if male
working	byte	%8.0g		1 if employed ; 0 otherwise
univ	byte	%8.0g		1 if university degree; 0 if not
hhkids	byte	%8.0g		1 if kids<16 in hhold;0 otherwise
healthy	byte	%8.0g		1 if hsat=7-10 ; 0 if hsat=0-6
hsat	byte	%8.0g		Health satisfaction (0 - 10)
age	byte	%8.0g		age in years (25 - 64)
educ	float	%9.0g		years of schooling (7 - 18)

Sorted by: id year

Random Effects Estimates

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtprobit doctor c.age##healthy educ female working univ hhkids, nolog
Random-effects probit regression
Group variable: id
Random effects u_i ~ Gaussian
Integration method: mvaghermite
Log likelihood = -15735.979
Number of obs      = 27,326
Number of groups   = 7,293
Obs per group:
    min = 1
    avg  = 3.7
    max  = 7
Integration pts.   = 12
Wald chi2(8)       = 1463.56
Prob > chi2        = 0.0000
```

doctor	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
age	.0163626	.0018509	8.84	0.000	.012735	.0199902
1.healthy	-.4430871	.0918739	-4.82	0.000	-.6231566	-.2630176
healthy# c.age 1	-.0045654	.0020279	-2.25	0.024	-.0085401	-.0005907
educ	.0211495	.0093003	2.27	0.023	.0029213	.0393777
female	.4272152	.0293703	14.55	0.000	.3696506	.4847799
working	-.0682233	.0272355	-2.50	0.012	-.1216038	-.0148428
univ	-.2881139	.0799262	-3.60	0.000	-.4447664	-.1314615
hhkids	-.1444234	.0250242	-5.77	0.000	-.1934699	-.0953768
_cons	-.1715246	.1450356	-1.18	0.237	-.4557892	.1127399
/lnsig2u	-.4509572	.0444464			-.5380707	-.3638438
sigma_u	.7981341	.0177371			.7641163	.8336664
rho	.3891332	.0105653			.3686365	.4100294

LR test of rho=0: chibar2(01) = 1996.30

Prob >= chibar2 = 0.000

Check sensitivity with quadcheck (Ideally <.1%)

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. quietly xtprobit doctor age i.healthy educ  
. quadchk, nooutput
```

```
Refitting model intpoints() = 8
```

```
Refitting model intpoints() = 16
```

Quadrature check				
	Fitted quadrature 12 points	Comparison quadrature 8 points	Comparison quadrature 16 points	
Log likelihood	-15889.701	-15889.713	-15889.701	Difference
		-.01255563	-.00012256	Relative difference
		7.902e-07	7.713e-09	
doctor: age	.0162238	.0162238	.0162238	Difference
		1.875e-12	-2.099e-09	Relative difference
		1.155e-10	-1.294e-07	
doctor: 1.healthy	-.64813857	-.64813857	-.64813853	Difference
		-2.377e-11	3.584e-08	Relative difference
		3.667e-11	-5.529e-08	
doctor: educ	-.01966104	-.01966104	-.01966104	Difference
		-1.703e-12	4.768e-09	Relative difference
		8.661e-11	-2.425e-07	
doctor: _cons	.39438356	.39438356	.39438353	Difference
		-1.351e-11	-2.874e-08	Relative difference
		-3.425e-11	-7.288e-08	
/: lnsig2u	-.36523047	-.36523047	-.36522977	Difference
		1.732e-10	6.960e-07	Relative difference
		-4.743e-10	-1.906e-06	

Marginal effects at () levels of continuous variable

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects

Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. quietly xtprobit doctor c.age##healthy educ female working univ hhkids  
. margins,dydx(healthy) at(educ = (8 12 14 16 18))
```

Average marginal effects
Model VCE: OIM

Number of obs = 27,326

```
Expression: Pr(doctor=1), predict(pr)  
dy/dx wrt: 1.healthy  
1._at: educ = 8  
2._at: educ = 12  
3._at: educ = 14  
4._at: educ = 16  
5._at: educ = 18
```

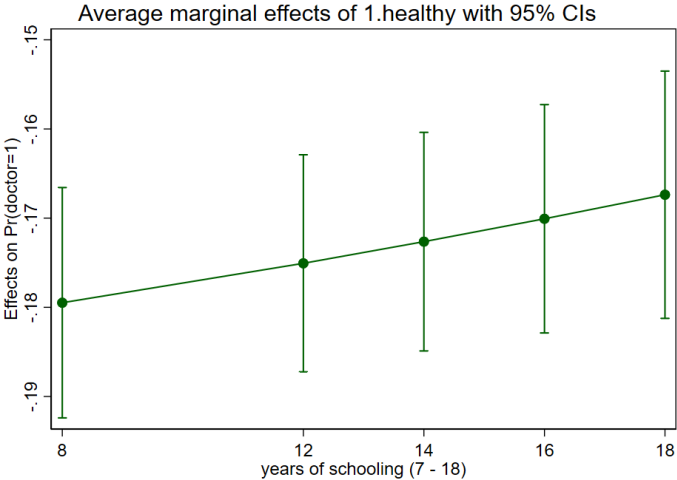
	Delta-method					
	dy/dx	std. err.	z	P> z	[95% conf. interval]	
0.healthy	(base outcome)					
1.healthy						
_at						
1	-.1794885	.0065906	-27.23	0.000	-.1924057	-.1665712
2	-.1750679	.006212	-28.18	0.000	-.1872432	-.1628925
3	-.1726386	.0062547	-27.60	0.000	-.1848976	-.1603796
4	-.1700724	.006533	-26.03	0.000	-.1828768	-.1572679
5	-.1673763	.0070749	-23.66	0.000	-.1812429	-.1535097

Note: dy/dx for factor levels is the discrete change from the base level.

Probability of visiting a doctor if feeling healthy, increases with more years of schooling

```
. marginsplot
```

Variables that uniquely identify margins: educ



Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

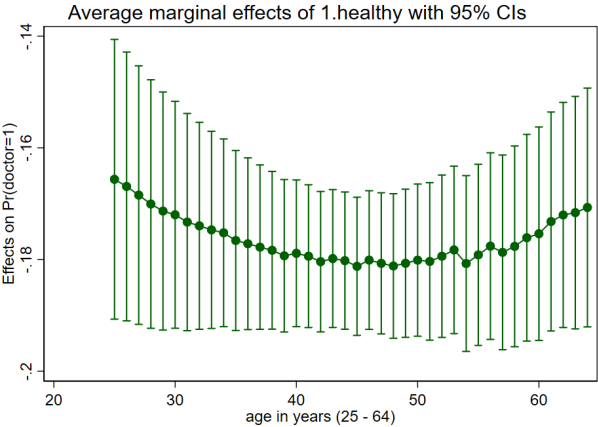
Final remarks

References

Probability of visiting a doctor increases at younger and older ages

```
. quietly margins, dydx(healthy) over(age)
. marginsplot
```

Variables that uniquely identify margins: age



Outline

Basic concepts

- Pooled vs. Panel
- Stata tools

Linear PD

- DGP
- Data and model
- Panel structure
- Random Effects
- Fixed Effects
- FE vs RE
- Marginal Analysis
- Autocorrelation / Heteroskedasticity
- IV Regression

Dynamic PD Models

- Arellano/Bond

Panel VAR models

Nonlinear PD Models

- Probit Models
- Logit Models
- Poisson Models

Final remarks

References

Let's now look at the random effects logit estimation

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtlogit doctor c.age##healthy educ female working univ hhkids, re nolog
Random-effects logistic regression
Group variable: id
Random effects u_i ~ Gaussian
Integration method: mvaghermite
Log likelihood = -15738.183
```

Number of obs = 27,326
Number of groups = 7,293
Obs per group:
min = 1
avg = 3.7
max = 7
Integration pts. = 12
Wald chi2(8) = 1403.21
Prob > chi2 = 0.0000

doctor	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
age	.0281942	.0031926	8.83	0.000	.0219369	.0344515
1.healthy	-.732481	.1574612	-4.65	0.000	-1.041099	-.4238627
healthy# c.age 1	-.0083444	.0034846	-2.39	0.017	-.0151741	-.0015147
educ	.0341623	.0158322	2.16	0.031	.0031318	.0651929
female	.7313507	.0502127	14.57	0.000	.6329355	.8297658
working	-.1107481	.0465192	-2.38	0.017	-.2019241	-.0195721
univ	-.4775681	.1358105	-3.52	0.000	-.7437517	-.2113845
hhkids	-.2460232	.0425557	-5.78	0.000	-.3294308	-.1626156
_cons	-.2914805	.247572	-1.18	0.239	-.7767127	.1937518
/lnsig2u	.6073147	.0461673			.5168284	.697801
sigma_u	1.354805	.0312739			1.294875	1.417508
rho	.3581201	.0106125			.3375975	.3791762

How about the fixed-effects logit estimator

Outline

Basic concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD Models

Arellano/Bond

Panel VAR models

Nonlinear PD Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

- There is no transform to remove ν_i from the panel-probit likelihood function
- However, there exists a transform to remove ν_i from the panel-logit likelihood function.
 - The trick is to look at the probability of transitioning from 0 (1) in the previous period to 1 (0) in the current period.
 - Obtaining those conditional probabilities with the logit functional form do not depend on ν_i
- Only transitions from 1 to 0 or from 0 to 1 provide information
- As in linear models, coefficients on time-invariant variables are not identified

Notice the sample reduction with the FE logit estimation

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtlogit doctor c.age##healthy educ female working univ hhkids, fe nolog
note: multiple positive outcomes within groups encountered.
note: 3,960 groups (11,014 obs) omitted because of all positive or
      all negative outcomes.
note: female omitted because of no within-group variance.
Conditional fixed-effects logistic regression
Group variable: id

Number of obs      = 16,312
Number of groups   =  3,333
Obs per group:
      min =      2
      avg =     4.9
      max =      7

LR chi2(7)         = 450.17
Prob > chi2        = 0.0000

Log likelihood = -6182.9605
```

doctor	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
age	.0754335	.0068851	10.96	0.000	.0619389	.0889281
1.healthy	-.6747047	.2020007	-3.34	0.001	-1.070619	-.2787906
healthy# c.age 1	-.000827	.0044929	-0.18	0.854	-.0096329	.007979
educ	-.0170087	.0997681	-0.17	0.865	-.2125507	.1785333
female	0 (omitted)					
working	-.0225414	.066679	-0.34	0.735	-.1532298	.108147
univ	-.4313305	.5107612	-0.84	0.398	-1.432404	.5697431
hhkids	-.0963817	.065153	-1.48	0.139	-.2240792	.0313158

Look for a model specification with the reduced sample

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis

Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtlogit doctor c.age##healthy educ working univ hhkids, fe  
. keep if e(sample)
```

```
. xtlogit doctor age univ healthy,fe nolog  
note: multiple positive outcomes within groups encountered.
```

Conditional fixed-effects logistic regression
Group variable: id

Number of obs = 16,312
Number of groups = 3,333
Obs per group:
min = 2
avg = 4.9
max = 7
LR chi2(3) = 447.76
Prob > chi2 = 0.0000

Log likelihood = -6184.1657

doctor	Coefficient	Std. err.	z	P> z	[95% conf. interval]	
age	.0764022	.0059246	12.90	0.000	.0647901	.0880142
univ	-.5476045	.2920267	-1.88	0.061	-1.119966	.0247573
healthy	-.7104429	.0473608	-15.00	0.000	-.8032684	-.6176174

Let's compare FE versus RE with the Hausman Test

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
.  
. xtlogit doctor c.age##healthy educ female working univ hhkids, fe  
. keep if e(sample)  
. xtlogit doctor age univ healthy, re nolog  
. estimates store xtlogit_re  
. xtlogit doctor age univ healthy, fe nolog  
. estimates store xtlogit_fe
```

```
. hausman xtlogit_fe xtlogit_re
```

	Coefficients			
	(b) xtlogit_fe	(B) xtlogit_re	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
age	.0764022	.0090162	.067386	.0057044
univ	-.5476045	-.1105003	-.4371042	.2852839
healthy	-.7104429	-.7330712	.0226283	.0311295

```
          b = Consistent under H0 and Ha; obtained from xtlogit.  
          B = Inconsistent under Ha, efficient under H0; obtained from xtlogit.  
  
Test of H0: Difference in coefficients not systematic  
  
      chi2(3) = (b-B)' [(V_b-V_B)^(-1)] (b-B)  
              = 139.67  
Prob > chi2 = 0.0000
```

We can also fit the CRE model

- As in the linear we can model the correlation between the observed component and the regressors

$$E(\alpha_i | x_i, z_i) = \bar{x}_i \theta_1 + \bar{z}_i \theta_2$$

- Then, substitute that expression in the functional form for the model:

$$Pr(y_{it} | x_i, z_i) = F(x_i \beta + \bar{x}_i \theta_1 + \bar{z}_i \theta_2)$$

- Let's compute the panel means:

```
.
. generate hlth_age=healthy*age
. foreach var of varlist age healthy hlth_age educ ///
>                                female working univ hhkids {
.     by id: egen double `var' _m = mean(`var')
. }
```

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

```

. local explain C.age##healthy educ female working ///
>               univ hhkids

. local explain_m age_m healthy_m hlth_age_m educ_m ///
>               female_m working_m univ_m hhkids_m

. quietly xtlogit doctor `explain' `explain_m', nolog

. test `explain_m'

( 1)  [doctor]age_m = 0
( 2)  [doctor]healthy_m = 0
( 3)  [doctor]hlth_age_m = 0
( 4)  [doctor]educ_m = 0
( 5)  [doctor]o.female_m = 0
( 6)  [doctor]working_m = 0
( 7)  [doctor]univ_m = 0
( 8)  [doctor]hhkids_m = 0
      Constraint 5 dropped

             chi2( 7) =   301.95
      Prob > chi2 =    0.0000

```


Odds Ratios

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtlogit doctor c.age##healthy educ female working univ hhkids,nolog or
Random-effects logistic regression
Group variable: id
Random effects u_i ~ Gaussian
Integration method: mvaghermite
Log likelihood = -15738.183

Number of obs      = 27,326
Number of groups   = 7,293
Obs per group:
    min = 1
    avg = 3.7
    max = 7
Integration pts.   = 12
Wald chi2(8)       = 1403.21
Prob > chi2        = 0.0000
```

doctor	Odds ratio	Std. err.	z	P> z	[95% conf. interval]	
age	1.028595	.0032839	8.83	0.000	1.022179	1.035052
1.healthy	.4807148	.075694	-4.65	0.000	.3530663	.6545138
healthy# c.age						
1	.9916903	.0034557	-2.39	0.017	.9849404	.9984865
educ	1.034753	.0163824	2.16	0.031	1.003137	1.067365
female	2.077885	.1043363	14.57	0.000	1.88313	2.292782
working	.8951642	.0416423	-2.38	0.017	.817157	.9806182
univ	.6202901	.0842419	-3.52	0.000	.4753273	.8094628
hhkids	.7819041	.0332745	-5.78	0.000	.7193331	.8499178
_cons	.7471566	.1849751	-1.18	0.239	.4599154	1.213795
/lnsig2u	.6073147	.0461673			.5168284	.697801
sigma_u	1.354805	.0312739			1.294875	1.417508
rho	.3581201	.0106125			.3375975	.3791762

Note: Estimates are transformed only in the first equation to odds ratios.
Note: _cons estimates baseline odds (conditional on zero random effects).
LR test of rho=0: chibar2(01) = 1991.19 Prob >= chibar2 = 0.000

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Count data and the Poisson model

- Count data are dependent variables that take on nonnegative integer values
 - Hospitals visits, arrests, and felony convictions are examples of count data
- Because many observations have observed value 0, modeling the natural log is usually not an option

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Count data and the Poisson model

- The most common strategy is to specify a Poisson distribution for y_i conditional on x
- Although the Poisson distribution has some restrictive features, this model is surprising useful because the QML estimator is consistent for the true parameters under a weak set of conditions

The Poisson model

Outline

Basic
conceptsPooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV RegressionDynamic PD
Models

Arellano/Bond

Panel VAR
modelsNonlinear PD
ModelsProbit Models
Logit Models
Poisson Models

Final remarks

References

- If $\mu(x) = E[y|x]$ and y has Poisson distribution, conditional on x , then the probability mass function for y , conditional on x is

$$f(y|x) = \exp[-\mu(x)] \frac{[\mu(x)]^y}{y!}$$

- The Poisson model imposes the restrictive condition that the conditional mean equals the conditional variance
 - Because the QML estimator is consistent without this condition, we can just correct our estimator of the VCE for inference
- Assuming that $\mu(x) = \exp(x\beta)$ the log-likelihood is given by

$$\mathcal{L}(\beta) = \sum_{i=1}^N y_i x_i' \beta - \exp(x_i' \beta) - \ln(y_i)$$

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Poisson Random-Effects and Fixed Effects Models

- The key assumptions of the multiplicative panel-effects models we use for panel count data are
 - **Assumption MPE 1:** $E[y_{it}|x_{i1}, \dots, x_{iT}, \alpha_i] = \alpha_i \exp(x'_{it}\beta)$
 - **Assumption MPE 2:** $E[\alpha_i|x_{i1}, \dots, x_{iT}] = 1$
- If both MPE 1 and MPE 2 hold, we have a random-effects model
- If only MPE 1 holds, the α_i depend on the x_{i1}, \dots, x_{iT} and we have a fixed-effects model

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

Poisson Random-Effects and Fixed Effects Models

- We will consider two estimators here for the parameters of these models
 - 1 ML (QML) estimator of the parameters of a random-effects model
 - 2 A conditional ML estimator of the parameters of the fixed-effects model

Example 9: Number of doctor visits

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. describe using $data_directory\German_healthcare_xtpoisson
```

Contains data

Observations: 27,326 17 Apr 2026 10:37
Variables: 9

Variable name	Storage type	Display format	Value label	Variable label
id	int	%8.0g		person - identification number
year	int	%8.0g		calendar year of the observation
docvis	int	%8.0g		number of doctor visits in last three months
age	byte	%8.0g		age in years (25 - 64)
hhkids	byte	%8.0g		1 if kids<16 in hhold;0 otherwise
healthy	byte	%8.0g		1 if hsat=7-10 ; 0 if hsat=0-6
female	byte	%8.0g		1 if female ; 0 if male
working	byte	%8.0g		1 if employed ; 0 otherwise
univ	byte	%8.0g		1 if university degree; 0 if not

Sorted by: id year

Random Effects Estimates

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtpoisson docvis c.age##healthy female working univ hhkids,nolog irr
Random-effects Poisson regression              Number of obs   = 27,326
Group variable: id                            Number of groups  = 7,293
Random effects u_i ~ Gamma                    Obs per group:
                                                min = 1
                                                avg = 3.7
                                                max = 7
Wald chi2(7) = 4241.50
Prob > chi2 = 0.0000
Log likelihood = -70129.935
```

docvis	IRR	Std. err.	z	P> z	[95% conf. interval]	
age	1.012921	.0009559	13.60	0.000	1.011049	1.014796
1.healthy	.4449601	.0181164	-19.89	0.000	.4108323	.4819229
healthy#						
c.age						
1	1.005543	.0008723	6.37	0.000	1.003835	1.007254
female	1.330774	.0336429	11.30	0.000	1.266442	1.398374
working	.8974939	.0111388	-8.71	0.000	.8759255	.9195933
univ	.7740507	.034792	-5.70	0.000	.7087771	.8453356
hhkids	.9517416	.0131318	-3.58	0.000	.9263487	.9778306
_cons	2.254132	.1109875	16.51	0.000	2.046767	2.482505
/lnalpha	-.0110847	.0197083			-.0497123	.0275428
alpha	.9889765	.019491			.9515032	1.027926

```
Note: Estimates are transformed only in the first equation to incidence-rate ratios.
Note: _cons estimates baseline incidence rate (conditional on zero random effects).
LR test of alpha=0: chibar2(01) = 5.0e+04                Prob >= chibar2 = 0.000
. estimates store poi_re
```


Fixed Effects Estimates

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects

Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. xtpoisson docvis c.age##healthy female working univ hhkids,nolog fe irr
note: 1525 groups (1525 obs) dropped because of only one obs per group
note: 589 groups (2148 obs) dropped because of all zero outcomes
note: female dropped because it is constant within group
Conditional fixed-effects Poisson regression      Number of obs   =   23,653
Group variable: id                               Number of groups =    5,179
                                                Obs per group:
                                                min   =         2
                                                avg   =         4.6
                                                max   =         7
                                                Wald chi2(6)    =  2725.81
                                                Prob > chi2     =   0.0000

Log likelihood = -46557.349
```

docvis	IRR	Std. err.	z	P> z	[95% conf. interval]	
age	1.01974	.0014322	13.92	0.000	1.016937	1.022551
1.healthy	.4601451	.0200798	-17.79	0.000	.4224254	.5012329
healthy#						
c.age						
1	1.00616	.000927	6.66	0.000	1.004344	1.007978
working	.9103053	.0121524	-7.04	0.000	.8867959	.934438
univ	.8160101	.0781546	-2.12	0.034	.6763484	.9845111
hhkids	.9975472	.0154515	-0.16	0.874	.9677178	1.028296

```
. estimates store poi_fe
```

The Hausman test favors the fixed effects model

Outline

Basic
concepts

Pooled vs. Panel
Stata tools

Linear PD

DGP
Data and model
Panel structure
Random Effects
Fixed Effects
FE vs RE
Marginal Analysis
Autocorrelation /
Heteroskedasticity
IV Regression

Dynamic PD
Models

Arellano/Bond

Panel VAR
models

Nonlinear PD
Models

Probit Models
Logit Models
Poisson Models

Final remarks

References

```
. hausman poi_fe poi_re
```

	Coefficients			
	(b) poi_fe	(B) poi_re	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
age	.0195478	.0128383	.0067094	.0010402
1.healthy	-.7762134	-.8097707	.0335573	.0157031
healthy#				
c.age				
1	.0061407	.0055281	.0006127	.0003105
working	-.0939752	-.108149	.0141738	.0049177
univ	-.2033285	-.2561179	.0527894	.0845743
hhkids	-.0024558	-.0494617	.0470059	.0070392

```
b = Consistent under H0 and Ha; obtained from xtpoisson.  
B = Inconsistent under Ha, efficient under H0; obtained from xtpoisson.  
Test of H0: Difference in coefficients not systematic  
chi2(6) = (b-B)'[(V_b-V_B)^(-1)](b-B)  
= 728.09  
Prob > chi2 = 0.0000
```

```
. test `mean_predictors'  
( 1) [docvis]healthy_m = 0  
( 2) [docvis]working_m = 0  
( 3) [docvis]univ_m = 0  
( 4) [docvis]hhkids_m = 0  
chi2( 4) = 528.47  
Prob > chi2 = 0.0000
```

Final remarks

- Distinguish assumptions between FE and RE. E.g. Individual panel effects:
 - Are assumed independent for RE
 - Could be correlated to the regressors in the case of FE.
- Be cautious about sample reduction in testing FE vs RE
- Dynamic panel estimators account for endogeneity between the lagged dependent variable and the idiosyncratic error term.
- Panel VAR models are particularly useful to analyze the impact of shocks in the endogenous variables.
- Non-linear panel data models require special treatment for FE, which may only be feasible for some specific models.

References

Anderson and Hsiao 1981. *Estimation of dynamic models with error components*. Journal of the American Statistical Association 76:598—606

Arellano, M. and S. Bond. 1991. *Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations*. The Review of Economic Studies 58: 277—297.

Arellano, M. and O. Bover. 1995. *Another look at the instrumental variable estimation of error-components models*. Journal of Econometrics 68: 29—51.

Blundell, R. and S. Bond. 1998. *Initial conditions and moment restrictions in dynamic panel data models*. Journal of Econometrics 87: 115—43.

Drukker, D. M. 2003. *Testing for serial correlation in linear panel-data models*. Stata Journal 3: 168—177.

Poi and Wiggins 2001. <http://www.stata.com/support/faqs/stat/panel.html>

Wooldridge, J. M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press

World Bank DataBank <http://databank.worldbank.org/data/Home.aspx>