Regression 2: Panel data and interaction models

LQRPS

Frederik Hjorth
fh@ifs.ku.dk
fghjorth.github.io
@fghjorth

Department of Political Science
University of Copenhagen

February 7th, 2017

1 Larsen et al. (2016): did Denmark's housing bubble re-elect the government?

2 Paneldata

3 Clustered standard errors

4 Benjamin's paper

5 Multilevel models
- Partial pooling
- Group-level predictors

6 Interactions

7 Limited dependent variables

1 Larsen et al. (2016): did Denmark's housing bubble re-elect the government?

2 Paneldata

3 Clustered standard errors

4 Benjamin's paper

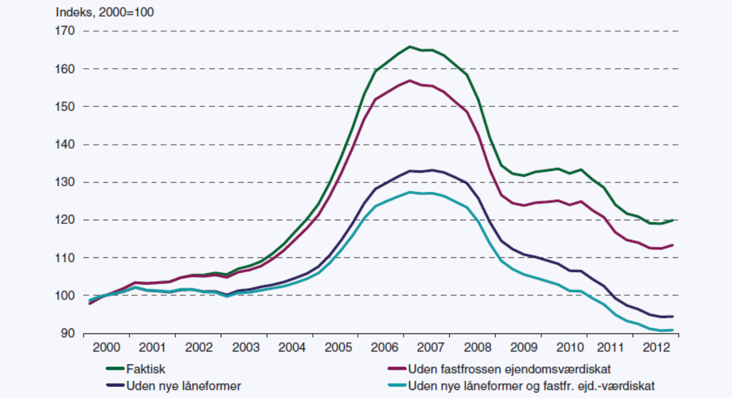
5 Multilevel models

6 Interactions

7 Limited dependent variables

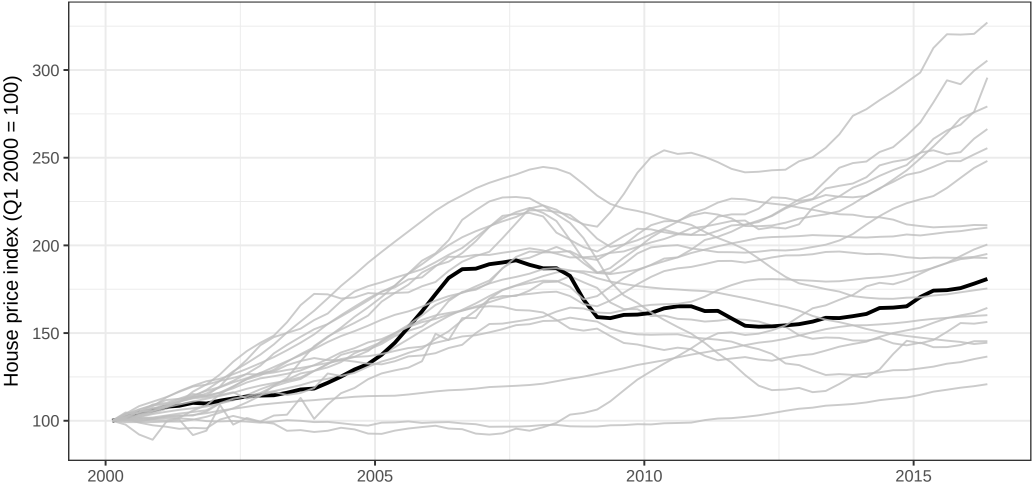
Larsen et al. (2016): did Denmark's housing bubble re-elect the government?

Figur B Huspriser med og uden nye låneformer og fastfrosset ejendomsværdiskat



Anm.: Kontrafaktiske forløb baseret på estimeret efterspørgselsrelation. Variabelt forrentede lån antages først at slå igennem i 1. kvartal 2000. I fravær af den fastfrosne ejendomsværdiskat antages den i MONAs databank imputerede ejendomsværdiskat at blive holdt konstant som andel af boligbeholdningen opgjort til markedsværdi. Fastfrysningen af ejendomsværdiskatten er modelleret som en permanent nedsættelse af ejendomsværdiskatten med 63 pct. i 1. kvartal 2002, svarende til nutidsværdien i ændringen af det fremtidige skatteprovenu under antagelse af at skattestoppet er permanent. Den grønne linje (det faktiske forløb) angiver den samlede effekt af henholdsvis nye låneformer (rød linje) og fastfrossen ejendomsværdiskat (blå linje). Beregningerne er nærmere dokumenteret i Dam m.fl. (2011). Serierne er genberegnet og forlænget frem til udgangen af 2012.

Kilde: Danmarks Nationalbank.


House price index (Q1 2000 = 100)
300
250
200
150
100
2000
2005
2010
2015

»Third, the panel set-up of data enables us to rule out time-invariant structural differences between local contexts as explanations of any observed relationship between local house prices and support for incumbents by using only within-precinct/individual variation in local housing prices (by means of fixed effects). This is particularly important given the strong urban-rural gradient in local property values, which would very likely confound any observed cross-sectional relationship with support for the sitting government.«

»Third, the panel set-up of data enables us to rule out time-invariant structural differences between local contexts as explanations of any observed relationship between local house prices and support for incumbents by using only within-precinct/individual variation in local housing prices (by means of fixed effects). This is particularly important given the strong urban-rural gradient in local property values, which would very likely confound any observed cross-sectional relationship with support for the sitting government.«

»Third, the panel set-up of data enables us to rule out time-invariant structural differences between local contexts as explanations of any observed relationship between local house prices and support for incumbents by using only within-precinct/individual variation in local housing prices (by means of fixed effects). This is particularly important given the strong urban-rural gradient in local property values, which would very likely confound any observed cross-sectional relationship with support for the sitting government.«

Table 1: Estimated effects of house prices on electoral support for governing parties.

	(1)	(2)	(3)	(4)
Δ housing price	0.104**	0.048**	0.053**	0.030**
	(0.008)	(0.007)	(0.008)	(0.007)
Unemployment rate				-1.904**
				(0.221)
Log(Median income)				-0.887**
				(0.064)
Year FE		✓	✓	✓
Precinct FE			✓	✓
Observations	4197	4197	4197	4177
RMSE	8.407	6.751	5.716	5.326

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$

Panel data: units are observed *multiple times*

Remember the formula from before lunch:

$$Y_i = \alpha + \beta P_i + \gamma A_i + e_i \tag{1}$$

Assume we observe income (Y_i) and private school education (P_i) over time t:

$$Y_{it} = \alpha + \beta P_{it} + \gamma A_i + e_{it} \tag{2}$$

note: A_i does not depend on t, i.e. is *time-invariant*

Panel data: units are observed *multiple times*

Remember the formula from before lunch:

$$Y_i = \alpha + \beta P_i + \gamma A_i + e_i \tag{1}$$

Assume we observe income (Y_i) and private school education (P_i) over time t:

$$Y_{it} = \alpha + \beta P_{it} + \gamma A_i + e_{it} \tag{2}$$

note: A_i does not depend on t, i.e. is *time-invariant*

Panel data: units are observed *multiple times*

Remember the formula from before lunch:

$$Y_i = \alpha + \beta P_i + \gamma A_i + e_i \tag{1}$$

Assume we observe income (Y_i) and private school education (P_i) over time t:

$$Y_{it} = \alpha + \beta P_{it} + \gamma A_i + e_{it} \tag{2}$$

note: A_i does not depend on t, i.e. is *time-invariant*

Panel data: units are observed *multiple times*

Remember the formula from before lunch:

$$Y_i = \alpha + \beta P_i + \gamma A_i + e_i \tag{1}$$

Assume we observe income (Y_i) and private school education (P_i) over time t:

$$Y_{it} = \alpha + \beta P_{it} + \gamma A_i + e_{it} \tag{2}$$

note: A_i does not depend on t, i.e. is *time-invariant*

As long as A_i is time-invariant, panel data lets us estimate β without bias even without observing A_i:

$$Y_i = \alpha_i + \lambda_t + \beta P_i + e_i \tag{3}$$

- 'fixed effects' (FE) model
- α_i = unit fixed effects → captures *time-invariant unobserved unit-level heterogeneity*
- λ_t = time fixed effects → captures *unit-invariant unobserved time-level heterogeneity*
- estimates based only on remaining variation 'within' units → also referred to as *within-estimator*

As long as A_i is time-invariant, panel data lets us estimate β without bias even without observing A_i:

$$Y_i = \alpha_i + \lambda_t + \beta P_i + e_i \tag{3}$$

- 'fixed effects' (FE) model
- α_i = unit fixed effects → captures *time-invariant unobserved unit-level heterogeneity*
- λ_t = time fixed effects → captures *unit-invariant unobserved time-level heterogeneity*
- estimates based only on remaining variation 'within' units → also referred to as *within-estimator*

As long as A_i is time-invariant, panel data lets us estimate β without bias even without observing A_i:

$$Y_i = \alpha_i + \lambda_t + \beta P_i + e_i \tag{3}$$

- 'fixed effects' (FE) model
- α_i = unit fixed effects → captures *time-invariant unobserved unit-level heterogeneity*
- λ_t = time fixed effects → captures *unit-invariant unobserved time-level heterogeneity*
- estimates based only on remaining variation 'within' units → also referred to as *within-estimator*

As long as A_i is time-invariant, panel data lets us estimate β without bias even without observing A_i:

$$Y_i = \alpha_i + \lambda_t + \beta P_i + e_i \tag{3}$$

- 'fixed effects' (FE) model
- α_i = unit fixed effects → captures *time-invariant unobserved unit-level heterogeneity*
- λ_t = time fixed effects → captures *unit-invariant unobserved time-level heterogeneity*
- estimates based only on remaining variation 'within' units → also referred to as *within-estimator*

As long as A_i is time-invariant, panel data lets us estimate β without bias even without observing A_i:

$$Y_i = \alpha_i + \lambda_t + \beta P_i + e_i \tag{3}$$

- 'fixed effects' (FE) model
- α_i = unit fixed effects $\rightarrow$ captures *time-invariant unobserved unit-level heterogeneity*
- λ_t = time fixed effects $\rightarrow$ captures *unit-invariant unobserved time-level heterogeneity*
- estimates based only on remaining variation 'within' units $\rightarrow$ also referred to as *within-estimator*

As long as A_i is time-invariant, panel data lets us estimate β without bias even without observing A_i:

$$Y_i = \alpha_i + \lambda_t + \beta P_i + e_i \tag{3}$$

- 'fixed effects' (FE) model
- α_i = unit fixed effects → captures *time-invariant unobserved unit-level heterogeneity*
- λ_t = time fixed effects → captures *unit-invariant unobserved time-level heterogeneity*
- estimates based only on remaining variation 'within' units → also referred to as *within-estimator*

As long as A_i is time-invariant, panel data lets us estimate β without bias even without observing A_i:

$$Y_i = \alpha_i + \lambda_t + \beta P_i + e_i \tag{3}$$

- 'fixed effects' (FE) model
- α_i = unit fixed effects → captures *time-invariant unobserved unit-level heterogeneity*
- λ_t = time fixed effects → captures *unit-invariant unobserved time-level heterogeneity*
- estimates based only on remaining variation 'within' units → also referred to as *within-estimator*

- note: the FE model does not account for non-independent error terms
- specifically: observations (and hence error terms) are 'lumped' within units
- leads to underestimated standard errors → sad!
- models need to account for this 'lumpiness'
- → more on this later

- note: the FE model does not account for non-independent error terms
- specifically: observations (and hence error terms) are 'lumped' within units
- leads to underestimated standard errors → sad!
- models need to account for this 'lumpiness'
- → more on this later

- note: the FE model does not account for non-independent error terms
- specifically: observations (and hence error terms) are 'lumped' within units
- leads to underestimated standard errors → sad!
- models need to account for this 'lumpiness'
- → more on this later

- note: the FE model does not account for non-independent error terms
- specifically: observations (and hence error terms) are 'lumped' within units
- leads to underestimated standard errors → sad!
- models need to account for this 'lumpiness'
- → more on this later

- note: the FE model does not account for non-independent error terms
- specifically: observations (and hence error terms) are 'lumped' within units
- leads to underestimated standard errors $\rightarrow$ sad!
- models need to account for this 'lumpiness'
- $\rightarrow$ more on this later

- note: the FE model does not account for non-independent error terms
- specifically: observations (and hence error terms) are ‘lumped’ within units
- leads to underestimated standard errors $\rightarrow$ sad!
- models need to account for this ‘lumpiness’
- $\rightarrow$ more on this later

Questions or comments?

Observations' 'lumpiness' (clustering) can be expressed by the *intraclass correlation coefficient* (ICC):

$$ICC = \rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_y^2} \tag{4}$$

interpretation: what proportion of total variation reflects between-group differences?
→ ICC also used as reliability measure in psychometry

Observations' 'lumpiness' (clustering) can be expressed by the *intraclass correlation coefficient* (ICC):

$$ICC = \rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_y^2} \tag{4}$$

interpretation: what proportion of total variation reflects between-group differences?
→ ICC also used as reliability measure in psychometry

Observations' 'lumpiness' (clustering) can be expressed by the *intraclass correlation coefficient* (ICC):

$$ICC = \rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_y^2} \tag{4}$$

interpretation: what proportion of total variation reflects between-group differences?
→ ICC also used as reliability measure in psychometry

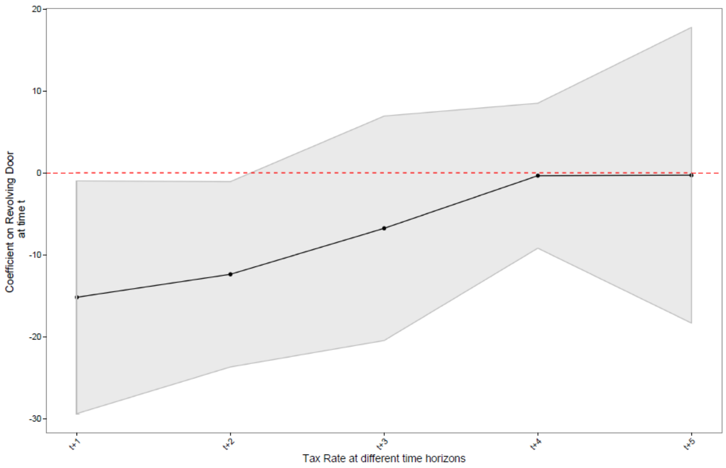
Observations' 'lumpiness' (clustering) can be expressed by the *intraclass correlation coefficient* (ICC):

$$ICC = \rho = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_y^2} \tag{4}$$

interpretation: what proportion of total variation reflects between-group differences?
→ ICC also used as reliability measure in psychometry

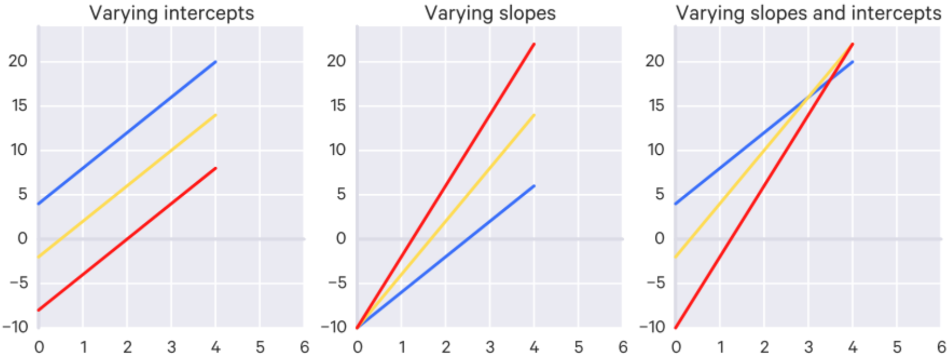
With clustered s.e.'s, increase depends on magnitude of ρ:

$$\sigma^2_{\hat{\beta}} = [1 + (k-1)\rho]\sigma^2_{\hat{\beta}_{OLS}}$$

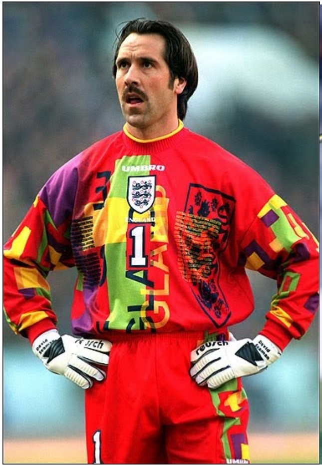

20
10
0
-10
-20
-30
Coefficient on Revolving Door
at time t
t+1
t+2
t+3
t+4
t+5
Tax Rate at different time horizons

Questions or comments?

Key distinction: variation in *slopes* ctr. *intercepts*



Example: two goalkeepers



Important KPI for goalkeepers: saving propensity π_i

- $\bar{\pi} = .1$
- $\pi_S = \frac{150}{1000}$
- $\pi_C = \frac{2}{5}$

$\rightarrow$ which goalie should we prefer?

Important KPI for goalkeepers: saving propensity π_i

- $\bar{\pi} = .1$
- $\pi_S = \frac{150}{1000}$
- $\pi_C = \frac{2}{5}$

$\rightarrow$ which goalie should we prefer?

Important KPI for goalkeepers: saving propensity π_i

- $\bar{\pi} = .1$
- $\pi_S = \frac{150}{1000}$
- $\pi_C = \frac{2}{5}$

$\rightarrow$ which goalie should we prefer?

Important KPI for goalkeepers: saving propensity π_i

- $\bar{\pi} = .1$
- $\pi_S = \frac{150}{1000}$
- $\pi_C = \frac{2}{5}$

$\rightarrow$ which goalie should we prefer?

Important KPI for goalkeepers: saving propensity π_i

- $\bar{\pi} = .1$
- $\pi_S = \frac{150}{1000}$
- $\pi_C = \frac{2}{5}$

$\rightarrow$ which goalie should we prefer?

- complete-pooling: $\pi_S = \pi_C = \bar{\pi} \approx .15 \rightarrow$ indifferent
- no-pooling: $\pi_S = .15$, $\pi_C = .4 \rightarrow$ prefer Campos
- are these satisfactory? why/why not?

- complete-pooling: $\pi_S = \pi_C = \bar{\pi} \approx .15 \rightarrow$ indifferent
- no-pooling: $\pi_S = .15$, $\pi_C = .4 \rightarrow$ prefer Campos
- are these satisfactory? why/why not?

- complete-pooling: $\pi_S = \pi_C = \bar{\pi} \approx .15 \rightarrow$ indifferent
- no-pooling: $\pi_S = .15$, $\pi_C = .4 \rightarrow$ prefer Campos
- are these satisfactory? why/why not?

- complete-pooling: $\pi_S = \pi_C = \bar{\pi} \approx .15 \rightarrow$ indifferent
- no-pooling: $\pi_S = .15$, $\pi_C = .4 \rightarrow$ prefer Campos
- are these satisfactory? why/why not?

- complete-pooling: $\pi_S = \pi_C = \bar{\pi} \approx .15 \rightarrow$ indifferent
- no-pooling: $\pi_S = .15$, $\pi_C = .4 \rightarrow$ prefer Campos
- are these satisfactory? why/why not?

- complete-pooling: $\pi_S = \pi_C = \bar{\pi} \approx .15 \rightarrow$ indifferent
- no-pooling: $\pi_S = .15$, $\pi_C = .4 \rightarrow$ prefer Campos
- are these satisfactory? why/why not?

- complete-pooling: $\pi_S = \pi_C = \bar{\pi} \approx .15 \rightarrow$ indifferent
- no-pooling: $\pi_S = .15$, $\pi_C = .4 \rightarrow$ prefer Campos
- are these satisfactory? why/why not?

$$\hat{\alpha}_j^{multilevel} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \tag{5}$$

- key insight: partial pooling $\rightarrow \hat{\alpha}_j^{multilevel}$ estimated as weighted avg. of $\bar{y}_j$ og $\bar{y}_{all}$
- utilized in *multilevel models* (a.k.a. random effects models)
- note: in this ex., multilevel modeling only w.r.t. α, i.e. the constant $\rightarrow$ *varying intercepts*

$$\hat{\alpha}_j^{multilevel} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \tag{5}$$

- key insight: partial pooling $\rightarrow$ $\hat{\alpha}_j^{multilevel}$ estimated as weighted avg. of $\bar{y}_j$ og $\bar{y}_{all}$
- utilized in *multilevel models* (a.k.a. random effects models)
- note: in this ex., multilevel modeling only w.r.t. α, i.e. the constant $\rightarrow$ *varying intercepts*

$$\hat{\alpha}_j^{multilevel} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \qquad (5)$$

- key insight: partial pooling $\rightarrow \hat{\alpha}_j^{multilevel}$ estimated as weighted avg. of $\bar{y}_j$ og $\bar{y}_{all}$
- utilized in *multilevel models* (a.k.a. random effects models)
- note: in this ex., multilevel modeling only w.r.t. α, i.e. the constant $\rightarrow$ *varying intercepts*

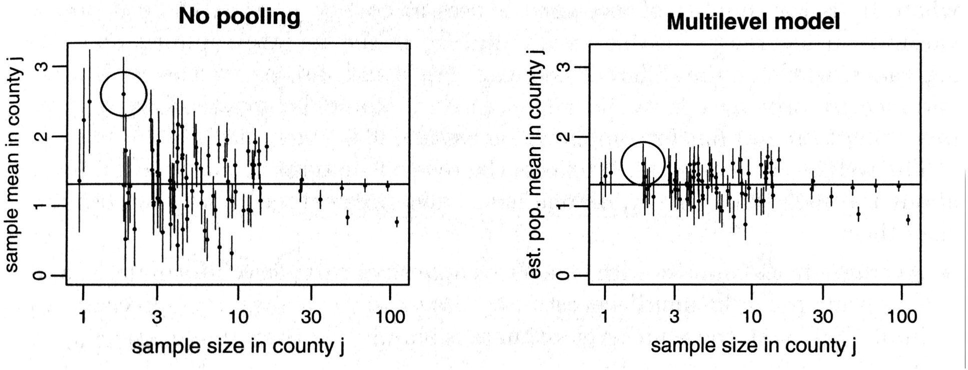
$$\hat{\alpha}_j^{multilevel} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \tag{5}$$

- key insight: partial pooling $\rightarrow \hat{\alpha}_j^{multilevel}$ estimated as weighted avg. of $\bar{y}_j$ og $\bar{y}_{all}$
- utilized in *multilevel models* (a.k.a. random effects models)
- note: in this ex., multilevel modeling only w.r.t. α, i.e. the constant $\rightarrow$ *varying intercepts*

$$\hat{\alpha}_j^{multilevel} \approx \frac{\frac{n_j}{\sigma_y^2}\bar{y}_j + \frac{1}{\sigma_\alpha^2}\bar{y}_{all}}{\frac{n_j}{\sigma_y^2} + \frac{1}{\sigma_\alpha^2}} \tag{5}$$

- key insight: partial pooling $\rightarrow \hat{\alpha}_j^{multilevel}$ estimated as weighted avg. of $\bar{y}_j$ og $\bar{y}_{all}$
- utilized in *multilevel models* (a.k.a. random effects models)
- note: in this ex., multilevel modeling only w.r.t. α, i.e. the constant $\rightarrow$ *varying intercepts*

Illustration in Gelman & Hill:



complete-pooling formalized:

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{6}$$

no-pooling formalized:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i \tag{7}$$

partial pooling: like no-pooling, but $\alpha_{j[i]}$ is modeled like in (5)
$\rightarrow$ group-level intercepts are modeled to fit a normal distribution, variance of which is estimated, allowing for calculating ICC

complete-pooling formalized:

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{6}$$

no-pooling formalized:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i \tag{7}$$

partial pooling: like no-pooling, but $\alpha_{j[i]}$ is modeled like in (5)
$\rightarrow$ group-level intercepts are modeled to fit a normal distribution, variance of which is estimated, allowing for calculating ICC

complete-pooling formalized:

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{6}$$

no-pooling formalized:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i \tag{7}$$

partial pooling: like no-pooling, but $\alpha_{j[i]}$ is modeled like in (5)
$\rightarrow$ group-level intercepts are modeled to fit a normal distribution, variance of which is estimated, allowing for calculating ICC

complete-pooling formalized:

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{6}$$

no-pooling formalized:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i \tag{7}$$

partial pooling: like no-pooling, but $\alpha_{j[i]}$ is modeled like in (5)

$\rightarrow$ group-level intercepts are modeled to fit a normal distribution, variance of which is estimated, allowing for calculating ICC

complete-pooling formalized:

$$y_i = \alpha + \beta x_i + \epsilon_i \tag{6}$$

no-pooling formalized:

$$y_i = \alpha_{j[i]} + \beta x_i + \epsilon_i \tag{7}$$

partial pooling: like no-pooling, but $\alpha_{j[i]}$ is modeled like in (5)
$\rightarrow$ group-level intercepts are modeled to fit a normal distribution, variance of which is estimated, allowing for calculating ICC

Questions or comments?

Example (Steenbergen & Jones, 2002): how is country-level trade exposure to Europe associated with public support for the EU?

Table 4 Determinants of EU Support

Parameter	Multilevel Estimate	Regression Estimate
Fixed Effects		
Constant	5.504**	5.016**
	(.220)	(.124)
Tenure	0.014	0.011**
	(.014)	(.002)
Trade	0.032	0.039**
	(.025)	(.003)
Party Cue	0.233**	0.275**
	(.028)	(.018)
Lowest Income Quartile	−.106+	−.181**
	(.064)	(.068)
Highest Income Quartile	0.048	−.001
	(.059)	(.062)
Ideology	0.019	0.023+
	(.015)	(.013)
Opinion Leadership	0.152**	0.166**
	(.028)	(.030)
Male	0.088+	0.093+
	(.050)	(.053)
Age	−.013**	−.014**
	(.002)	(.002)

- data: 6,354 respondents from 15 countries in Eurobarometer 1996
- focus here: *Trade* = trade with other EU countries
- why does significance level change when going from simple OLS (i.e., no-pooling) to multilevel model?

Example (Steenbergen & Jones, 2002): how is country-level trade exposure to Europe associated with public support for the EU?

Table 4 Determinants of EU Support

Parameter	Multilevel Estimate	Regression Estimate
Fixed Effects		
Constant	5.504**	5.016**
	(.220)	(.124)
Tenure	0.014	0.011**
	(.014)	(.002)
Trade	0.032	0.039**
	(.025)	(.003)
Party Cue	0.233**	0.275**
	(.028)	(.018)
Lowest Income Quartile	–.106+	–.181**
	(.064)	(.068)
Highest Income Quartile	0.048	–.001
	(.059)	(.062)
Ideology	0.019	0.023+
	(.015)	(.013)
Opinion Leadership	0.152**	0.166**
	(.028)	(.030)
Male	0.088+	0.093+
	(.050)	(.053)
Age	–.013**	–.014**
	(.002)	(.002)

- data: 6,354 respondents from 15 countries in Eurobarometer 1996
- focus here: *Trade* = trade with other EU countries
- why does significance level change when going from simple OLS (i.e., no-pooling) to multilevel model?

Example (Steenbergen & Jones, 2002): how is country-level trade exposure to Europe associated with public support for the EU?

Table 4 Determinants of EU Support

Parameter	Multilevel Estimate	Regression Estimate
Fixed Effects		
Constant	5.504**	5.016**
	(.220)	(.124)
Tenure	0.014	0.011**
	(.014)	(.002)
Trade	0.032	0.039**
	(.025)	(.003)
Party Cue	0.233**	0.275**
	(.028)	(.018)
Lowest Income Quartile	−.106+	−.181**
	(.064)	(.068)
Highest Income Quartile	0.048	−.001
	(.059)	(.062)
Ideology	0.019	0.023+
	(.015)	(.013)
Opinion Leadership	0.152**	0.166**
	(.028)	(.030)
Male	0.088+	0.093+
	(.050)	(.053)
Age	−.013**	−.014**
	(.002)	(.002)

- data: 6,354 respondents from 15 countries in Eurobarometer 1996
- focus here: *Trade* = trade with other EU countries
- why does significance level change when going from simple OLS (i.e., no-pooling) to multilevel model?

Example (Steenbergen & Jones, 2002): how is country-level trade exposure to Europe associated with public support for the EU?

Table 4 Determinants of EU Support

Parameter	Multilevel Estimate	Regression Estimate
Fixed Effects		
Constant	5.504**	5.016**
	(.220)	(.124)
Tenure	0.014	0.011**
	(.014)	(.002)
Trade	0.032	0.039**
	(.025)	(.003)
Party Cue	0.233**	0.275**
	(.028)	(.018)
Lowest Income Quartile	−.106+	−.181**
	(.064)	(.068)
Highest Income Quartile	0.048	−.001
	(.059)	(.062)
Ideology	0.019	0.023+
	(.015)	(.013)
Opinion Leadership	0.152**	0.166**
	(.028)	(.030)
Male	0.088+	0.093+
	(.050)	(.053)
Age	−.013**	−.014**
	(.002)	(.002)

- data: 6,354 respondents from 15 countries in Eurobarometer 1996
- focus here: *Trade* = trade with other EU countries
- why does significance level change when going from simple OLS (i.e., no-pooling) to multilevel model?

Example (Steenbergen & Jones, 2002): how is country-level trade exposure to Europe associated with public support for the EU?

Table 4 Determinants of EU Support

Parameter	Multilevel Estimate	Regression Estimate
Fixed Effects		
Constant	5.504**	5.016**
	(.220)	(.124)
Tenure	0.014	0.011**
	(.014)	(.002)
Trade	0.032	0.039**
	(.025)	(.003)
Party Cue	0.233**	0.275**
	(.028)	(.018)
Lowest Income Quartile	−.106+	−.181**
	(.064)	(.068)
Highest Income Quartile	0.048	−.001
	(.059)	(.062)
Ideology	0.019	0.023+
	(.015)	(.013)
Opinion Leadership	0.152**	0.166**
	(.028)	(.030)
Male	0.088+	0.093+
	(.050)	(.053)
Age	−.013**	−.014**
	(.002)	(.002)

- data: 6,354 respondents from 15 countries in Eurobarometer 1996
- focus here: *Trade* = trade with other EU countries
- why does significance level change when going from simple OLS (i.e., no-pooling) to multilevel model?

»These differences arise precisely because the OLS standard errors are too small. This attenuation is caused by ignoring the clustering of the data. The OLS analysis assumes that we have 6354 independent observations in our data. (...) The problem, of course, is that we do not have 6354 independent observations. (...) To pretend that they are independent is to assume that one has more information than really exists. Thus, the OLS analysis presents too optimistic a view about the significance of the predictor«

– Steenbergen, M. R., & Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science*, 46(1), 218-237.

> »These differences arise precisely because the OLS standard errors are too small. This attenuation is caused by ignoring the clustering of the data. The OLS analysis assumes that we have 6354 independent observations in our data. (...) The problem, of course, is that we do not have 6354 independent observations. (...) To pretend that they are independent is to assume that one has more information than really exists. Thus, the OLS analysis presents too optimistic a view about the significance of the predictor«

– Steenbergen, M. R., & Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science*, 46(1), 218-237.

> »These differences arise precisely because the OLS standard errors are too small. This attenuation is caused by ignoring the clustering of the data. The OLS analysis assumes that we have 6354 independent observations in our data. (...) The problem, of course, is that we do not have 6354 independent observations. (...) To pretend that they are independent is to assume that one has more information than really exists. Thus, the OLS analysis presents too optimistic a view about the significance of the predictor«

– Steenbergen, M. R., & Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science*, 46(1), 218-237.

»These differences arise precisely because the OLS standard errors are too small. This attenuation is caused by ignoring the clustering of the data. The OLS analysis assumes that we have 6354 independent observations in our data. (...) The problem, of course, is that we do not have 6354 independent observations. (...) To pretend that they are independent is to assume that one has more information than really exists. Thus, the OLS analysis presents too optimistic a view about the significance of the predictor«

– Steenbergen, M. R., & Jones, B. S. (2002). Modeling Multilevel Data Structures. *American Journal of Political Science*, 46(1), 218-237.

Advantage in multilevel models: in cross-sectional data, group-level coefficients can be estimated along with group-varying intercepts ($\sim$ group FE)

- in OLS: group FE and group-level predictor are collinear $\rightarrow$ can't be estimated simultaneously
- in multilevel models: α_j's estimated w. partial pooling

Advantage in multilevel models: in cross-sectional data, group-level coefficients can be estimated along with group-varying intercepts ($\sim$ group FE)

- in OLS: group FE and group-level predictor are collinear $\rightarrow$ can't be estimated simultaneously
- in multilevel models: α_j's estimated w. partial pooling

Advantage in multilevel models: in cross-sectional data, group-level coefficients can be estimated along with group-varying intercepts ($\sim$ group FE)

- in OLS: group FE and group-level predictor are collinear $\rightarrow$ can't be estimated simultaneously
- in multilevel models: α_j's estimated w. partial pooling

Questions or comments?

Assume a classic regression model:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i \tag{8}$$

but: effects of X_i og Z_i depend on value of each other $\rightarrow$ they *interact*:

$$\beta_1 = \delta_1 + \delta_2 Z_i \tag{9}$$

$$\beta_2 = \delta_3 + \delta_4 X_i \tag{10}$$

hence:

$$Y_i = \alpha + \beta_x X_i + \beta_z Z_i + \beta_{xz} X_i Z_i \tag{11}$$

note: $\beta_x \neq \beta_1$, $\beta_z \neq \beta_2$

Assume a classic regression model:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i \tag{8}$$

but: effects of X_i og Z_i depend on value of each other $\rightarrow$ they *interact*:

$$\beta_1 = \delta_1 + \delta_2 Z_i \tag{9}$$

$$\beta_2 = \delta_3 + \delta_4 X_i \tag{10}$$

hence:

$$Y_i = \alpha + \beta_x X_i + \beta_z Z_i + \beta_{xz} X_i Z_i \tag{11}$$

note: $\beta_x \neq \beta_1$, $\beta_z \neq \beta_2$

Assume a classic regression model:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i \tag{8}$$

but: effects of X_i og Z_i depend on value of each other $\rightarrow$ they *interact*:

$$\beta_1 = \delta_1 + \delta_2 Z_i \tag{9}$$

$$\beta_2 = \delta_3 + \delta_4 X_i \tag{10}$$

hence:

$$Y_i = \alpha + \beta_x X_i + \beta_z Z_i + \beta_{xz} X_i Z_i \tag{11}$$

note: $\beta_x \neq \beta_1$, $\beta_z \neq \beta_2$

Assume a classic regression model:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i \tag{8}$$

but: effects of X_i og Z_i depend on value of each other $\rightarrow$ they *interact*:

$$\beta_1 = \delta_1 + \delta_2 Z_i \tag{9}$$

$$\beta_2 = \delta_3 + \delta_4 X_i \tag{10}$$

hence:

$$Y_i = \alpha + \beta_x X_i + \beta_z Z_i + \beta_{xz} X_i Z_i \tag{11}$$

note: $\beta_x \neq \beta_1$, $\beta_z \neq \beta_2$

Assume a classic regression model:

$$Y_i = \alpha + \beta_1 X_i + \beta_2 Z_i \tag{8}$$

but: effects of X_i og Z_i depend on value of each other $\rightarrow$ they *interact*:

$$\beta_1 = \delta_1 + \delta_2 Z_i \tag{9}$$

$$\beta_2 = \delta_3 + \delta_4 X_i \tag{10}$$

hence:

$$Y_i = \alpha + \beta_x X_i + \beta_z Z_i + \beta_{xz} X_i Z_i \tag{11}$$

note: $\beta_x \neq \beta_1$, $\beta_z \neq \beta_2$

Implementation in R:

```
lm(y~x+z+x:z,data=df)
```

or

```
lm(y~x*z,data=df)
```

Implementation in R:

```
lm(y~x+z+x:z,data=df)
```

or

```
lm(y~x*z,data=df)
```

Implementation in R:

```
lm(y~x+z+x:z,data=df)
```

or

```
lm(y~x*z,data=df)
```

Questions or comments?

Imagine a dichotomous $Y \in \{0, 1\}$, e.g.

- voting
- referendum choice
- going to war
- treaty ratification
- etc.

$\rightarrow$ we can interpret the CEF as a *probability of Y*, i.e.

$$E[Y = 1|X = x] = Pr(Y = 1|X = x)$$

Imagine a dichotomous $Y \in \{0, 1\}$, e.g.

- voting
- referendum choice
- going to war
- treaty ratification
- etc.

$\rightarrow$ we can interpret the CEF as a *probability of Y*, i.e.

$$E[Y = 1|X = x] = Pr(Y = 1|X = x)$$

Imagine a dichotomous $Y \in \{0, 1\}$, e.g.

- voting
- referendum choice
- going to war
- treaty ratification
- etc.

$\rightarrow$ we can interpret the CEF as a *probability of Y*, i.e.

$$E[Y = 1|X = x] = Pr(Y = 1|X = x)$$

Imagine a dichotomous $Y \in \{0, 1\}$, e.g.

- voting
- referendum choice
- going to war
- treaty ratification
- etc.

$\rightarrow$ we can interpret the CEF as a *probability of Y*, i.e.

$$E[Y = 1|X = x] = Pr(Y = 1|X = x)$$

Imagine a dichotomous $Y \in \{0, 1\}$, e.g.

- voting
- referendum choice
- going to war
- treaty ratification
- etc.

→ we can interpret the CEF as a *probability of Y*, i.e.

$$E[Y = 1|X = x] = Pr(Y = 1|X = x)$$

Imagine a dichotomous $Y \in \{0, 1\}$, e.g.

- voting
- referendum choice
- going to war
- treaty ratification
- etc.

→ we can interpret the CEF as a *probability of Y*, i.e.

$$E[Y = 1|X = x] = Pr(Y = 1|X = x)$$

Imagine a dichotomous $Y \in \{0,1\}$, e.g.

- voting
- referendum choice
- going to war
- treaty ratification
- etc.

$\rightarrow$ we can interpret the CEF as a *probability of Y*, i.e.

$$E[Y = 1|X = x] = Pr(Y = 1|X = x)$$

Imagine a dichotomous $Y \in \{0, 1\}$, e.g.

- voting
- referendum choice
- going to war
- treaty ratification
- etc.

$\rightarrow$ we can interpret the CEF as a *probability of Y*, i.e.

$$E[Y = 1|X = x] = Pr(Y = 1|X = x)$$

Two ways to handle this:

1. model $Pr(Y)$ as a linear function of X, i.e. a *linear probability model* (LPM)
2. model the *logit transformed* Y as a linear function of X

in option 2, the *logistic regression model*, we model

$$log\left(\frac{Pr(Y)}{1+Pr(Y)}\right) = \alpha + \beta X$$

solving for $Pr(Y)$, we get

$$Pr(Y) = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}} = \frac{1}{1+e^{-(\alpha+\beta X)}}$$

Two ways to handle this:

1. model $Pr(Y)$ as a linear function of X, i.e. a *linear probability model* (LPM)
2. model the *logit transformed* Y as a linear function of X

in option 2, the *logistic regression model*, we model

$$log\left(\frac{Pr(Y)}{1+Pr(Y)}\right) = \alpha + \beta X$$

solving for $Pr(Y)$, we get

$$Pr(Y) = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}} = \frac{1}{1+e^{-(\alpha+\beta X)}}$$

Two ways to handle this:

1. model $Pr(Y)$ as a linear function of X, i.e. a *linear probability model* (LPM)
2. model the *logit transformed* Y as a linear function of X

in option 2, the *logistic regression model*, we model

$$log\left(\frac{Pr(Y)}{1+Pr(Y)}\right)=\alpha+\beta X$$

solving for $Pr(Y)$, we get

$$Pr(Y)=\frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}}=\frac{1}{1+e^{-(\alpha+\beta X)}}$$

Two ways to handle this:

1. model $Pr(Y)$ as a linear function of X, i.e. a *linear probability model* (LPM)
2. model the *logit transformed* Y as a linear function of X

in option 2, the *logistic regression model*, we model

$$log\left(\frac{Pr(Y)}{1+Pr(Y)}\right) = \alpha + \beta X$$

solving for $Pr(Y)$, we get

$$Pr(Y) = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}} = \frac{1}{1+e^{-(\alpha+\beta X)}}$$

Two ways to handle this:

1. model $Pr(Y)$ as a linear function of X, i.e. a *linear probability model* (LPM)
2. model the *logit transformed* Y as a linear function of X

in option 2, the *logistic regression model*, we model

$$log\left(\frac{Pr(Y)}{1+Pr(Y)}\right) = \alpha + \beta X$$

solving for $Pr(Y)$, we get

$$Pr(Y) = \frac{e^{\alpha+\beta X}}{1+e^{\alpha+\beta X}} = \frac{1}{1+e^{-(\alpha+\beta X)}}$$

- why would we go through all this trouble?
- probabilities are bounded by $[0, 1]$, whereas linear functions are unbounded
- $\rightarrow$ using the LPM, some $\hat{Y}$ are guaranteed to be wrong
- the logistic regression model fixes this by bounding Pr(Y)
- additionally, the functional form imposed by logistic regression may be more appropriate for probabilities
- but: for any given dichotomous-outcome DGP, this may not hold
- aren't "all models wrong" anyways?

- why would we go through all this trouble?
- probabilities are bounded by $[0, 1]$, whereas linear functions are unbounded
- $\rightarrow$ using the LPM, some $\hat{Y}$ are guaranteed to be wrong
- the logistic regression model fixes this by bounding Pr(Y)
- additionally, the functional form imposed by logistic regression may be more appropriate for probabilities
- but: for any given dichotomous-outcome DGP, this may not hold
- aren't "all models wrong" anyways?

- why would we go through all this trouble?
- probabilities are bounded by $[0, 1]$, whereas linear functions are unbounded
- $\rightarrow$ using the LPM, some $\hat{Y}$ are guaranteed to be wrong
- the logistic regression model fixes this by bounding Pr(Y)
- additionally, the functional form imposed by logistic regression may be more appropriate for probabilities
- but: for any given dichotomous-outcome DGP, this may not hold
- aren't "all models wrong" anyways?

- why would we go through all this trouble?
- probabilities are bounded by $[0, 1]$, whereas linear functions are unbounded
- $\rightarrow$ using the LPM, some $\hat{Y}$ are guaranteed to be wrong
- the logistic regression model fixes this by bounding Pr(Y)
- additionally, the functional form imposed by logistic regression may be more appropriate for probabilities
- but: for any given dichotomous-outcome DGP, this may not hold
- aren't "all models wrong" anyways?

- why would we go through all this trouble?
- probabilities are bounded by $[0, 1]$, whereas linear functions are unbounded
- $\rightarrow$ using the LPM, some $\hat{Y}$ are guaranteed to be wrong
- the logistic regression model fixes this by bounding Pr(Y)
- additionally, the functional form imposed by logistic regression may be more appropriate for probabilities
- but: for any given dichotomous-outcome DGP, this may not hold
- aren't "all models wrong" anyways?

- why would we go through all this trouble?
- probabilities are bounded by $[0, 1]$, whereas linear functions are unbounded
- $\rightarrow$ using the LPM, some $\hat{Y}$ are guaranteed to be wrong
- the logistic regression model fixes this by bounding Pr(Y)
- additionally, the functional form imposed by logistic regression may be more appropriate for probabilities
- but: for any given dichotomous-outcome DGP, this may not hold
- aren't "all models wrong" anyways?

- why would we go through all this trouble?
- probabilities are bounded by $[0, 1]$, whereas linear functions are unbounded
- $\rightarrow$ using the LPM, some $\hat{Y}$ are guaranteed to be wrong
- the logistic regression model fixes this by bounding Pr(Y)
- additionally, the functional form imposed by logistic regression may be more appropriate for probabilities
- but: for any given dichotomous-outcome DGP, this may not hold
- aren't "all models wrong" anyways?

See you tomorrow!