Advanced Quantitative Methods: Multilevel and panel data

Johan A. Elkink

University College Dublin

22 April 2014

Outline

Multilevel: motivations

- Clustered sampling

Cluster sampling

To reduce costs, clusters are (randomly) sampled first, before lower levels are clustered.

E.g. selecting schools before selecting students, so that fewer schools need to be visited.

Individual observations from a clustered sample are *not independent*.

Multilevel: motivations

- Clustered sampling
- Inherent structure
- Ability to model context dependency

Examples

schools	teachers
classes	pupils
firms	employees
countries	political parties
doctors	patients
subjects	measurements
interviewers	respondents
judges	suspects

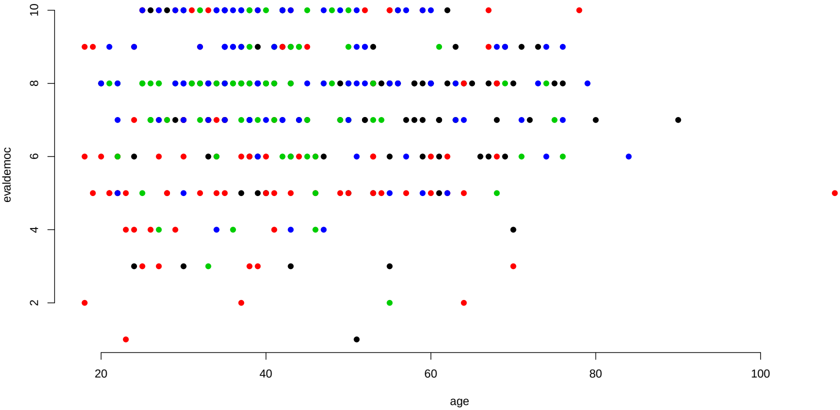
Multilevel: motivations

- Clustered sampling
- Inherent structure
- Ability to model context dependency
- Panel data

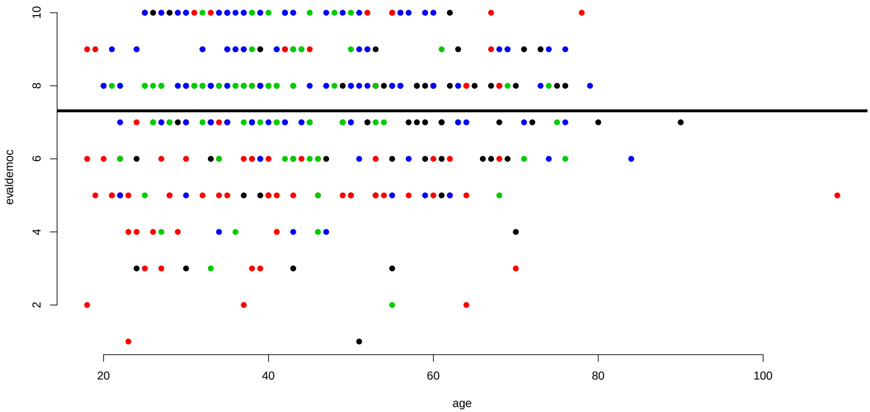
Multilevel characteristics

- Observations are not independent due to within-group correlation
- Variance can be divided in *between-group* and *within-group* variances
- Variables can be measured at either micro- or marco-level, or both
- Micro- and macro-level variables can interact, e.g. context-specific effects

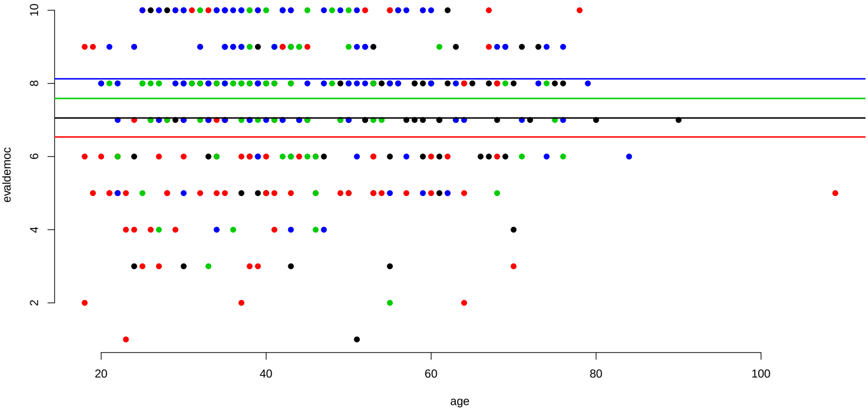
Example



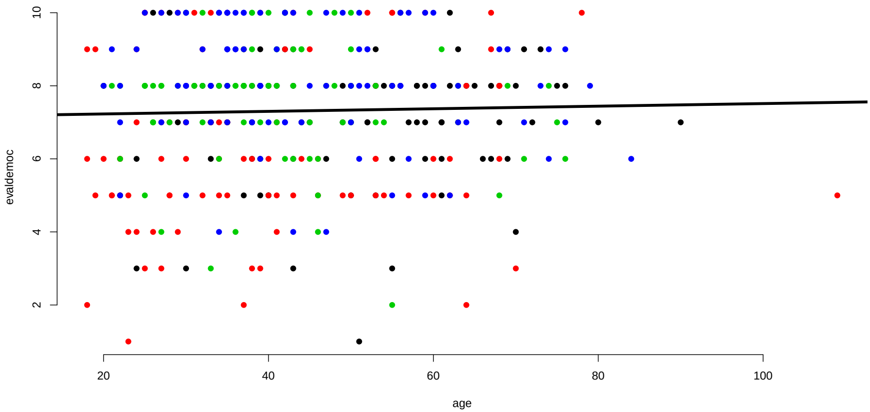
Overall mean



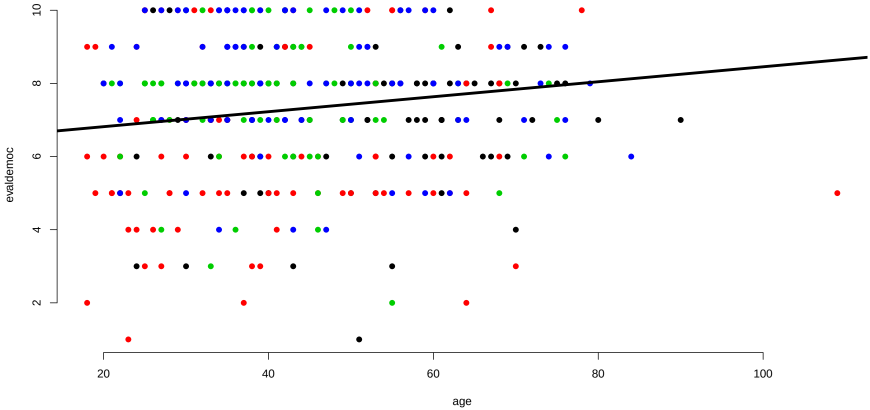
Group means



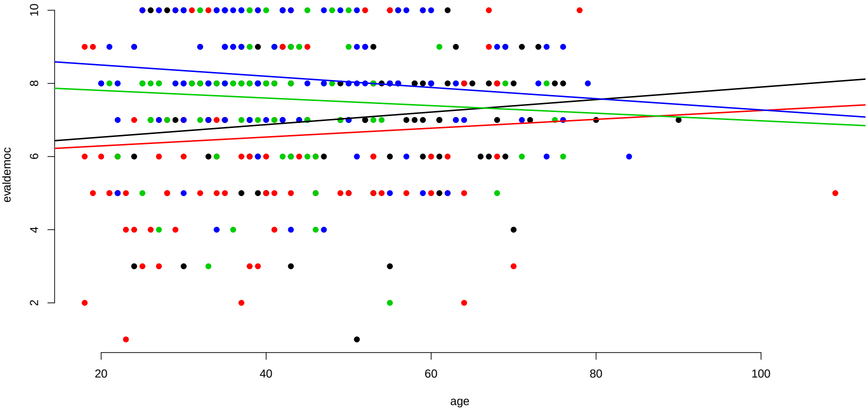
Pooled model



Between model



Within model



Panel data

Country	*Year*	*GDP*	*Population*
Ireland	1990	80	3.4
Ireland	1991	84	3.5
Ireland	1992	82	3.5
...	...	...	...
Germany	1990	132	80.1
Germany	1991	156	80.2
...	...	...	...
Netherlands	1990	102	16.5
Netherlands	1991	103	16.6
...	...	...	...

Panel data

$$y_{it} = \alpha + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$
$$i = \{1, 2, \ldots, N\}$$
$$t = \{1, 2, \ldots, T\}$$

with i indicating the unit (country, individual, etc.) and t indicating the time period (year, month, etc.).

It thus combines concerns with

- Cross-sectional autocorrelation (cf. multilevel)
- Temporal autocorrelation (cf. time-series)
- Spatial autocorrelation

Panel data

Common both at micro-level (usually $N >> T$) or macro-level (usually $T > N$). In political science, the latter is sometimes called **Time-Series Cross-Section** (TSCS) data.

Consistency behaviour differs, for different estimators, for $N \to \infty$ or $T \to \infty$.

Outline

Pooled model

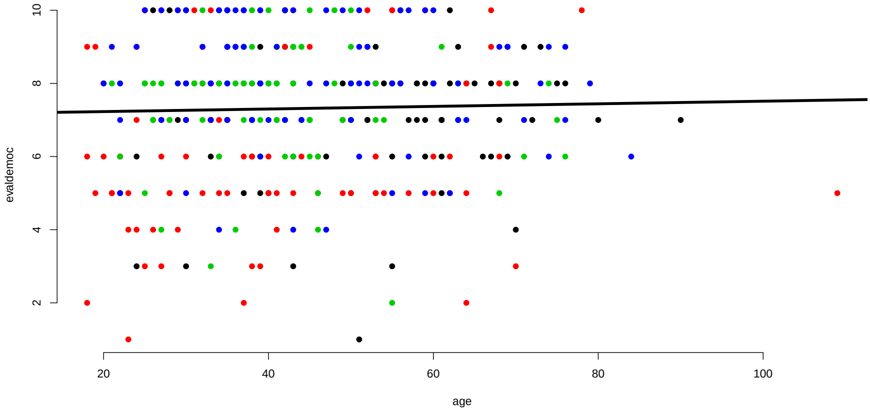
When we simply run a regression using all micro-level data, ignoring the multilevel structure, we call this a **pooled model**.

If we have some observations at the macro-level, we are artificially increasing the number of observations, so we will be **overconfident** in our results.

If there is positive within-group correlation, the effect is similar to positive autocorrelation in time-series, and we will underestimate our standard errors.

E.g. characteristics of judges in explaining the severity of court rulings.

Pooled model



Fixed effects model

With a fixed effects model we explain the *within-group* variation, removing the *between-group* variation by:

- Adding dummy variables for each group
- Subtracting the group means from *all* variables

The two are equivalent, except for the latter the degrees of freedom need to be adjusted.

The term "fixed effects" leads to confusion, as most independent variables are considered fixed. A more appropriate term is **Least Squares Dummy Variable** (LSDV) model.

Fixed effects model

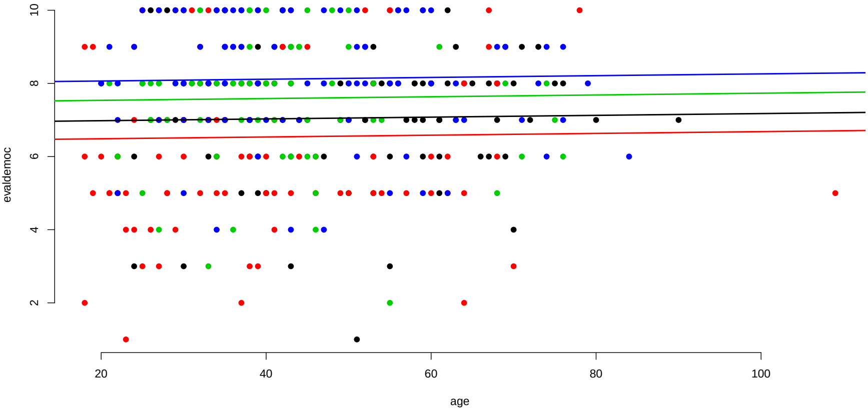
In essence, we thus have different **intercepts** for each group.

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + \mu_{j[i]} + \varepsilon_i,$$

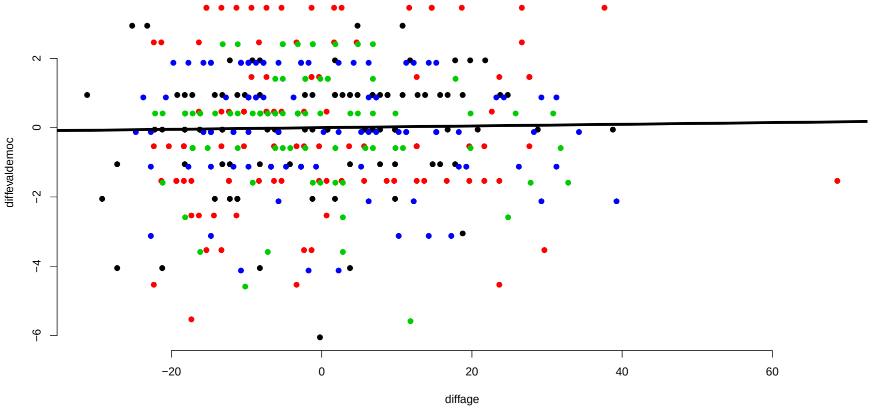
whereby i denotes the individual unit, j the group, and $j[i]$ the group of i.

If the fixed effects model is the true model, pooled estimates are biased and inconsistent.

Fixed effects model (LSDV)



Fixed effects model (demeaning)



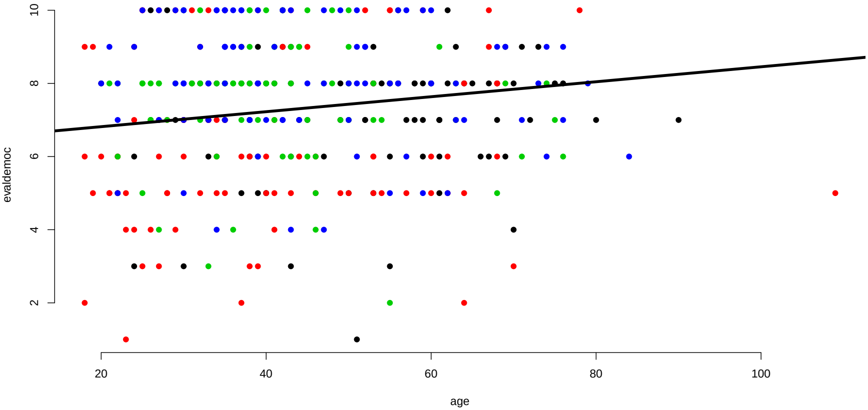
Between effects model

Another way of dealing with clustered data is looking at the between model:

$$\bar{y}_j = \bar{\mathbf{x}}_j \boldsymbol{\beta} + \varepsilon_j$$

Typical mistake: conclusions about individuals from aggregate data - **ecological fallacy**.

Between effects model



Exercise

Using the `asiabaro.dta` data file:

1. Estimate a pooled model explaining `evaldemoc` by `evalcorrupt`.
2. Estimate a country fixed effects model using dummy variables, including an intercept.
3. Estimate a country fixed effects model using dummy variables, excluding the overall intercept.
4. Estimate a country fixed effects model using demeaned variables.

Demeaning:

```
mux <- apply(x, country, mean, na.rm = TRUE)
xdemeaned <- x - mux[country]
```

Group-level variables

Note that fixed effects models cannot deal with group-level variables.

The effect would be *perfect multicollinearity*.

High multicollinearity also arises from variables with low variance - e.g. political institutions.

In most cases with group-level variables, a **random effects** or **random intercept** model is more appropriate.

Fixed effects

A panel data structure is similar to a multilevel structure, but with a time element.

Thus, we can think of the variation between units and within one unit over time.

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

Therefore, fixed unit effects are one way of dealing with the between variation. E.g. controlling for "culture" in a macro-level analysis.

Fixed effects

As with multilevel data,

- unit fixed effects cannot be combined with (near-)constant unit-level variables (e.g. institutions);
- unit fixed effects take away $N - 1$ degrees of freedom - T needs to be sufficiently large;
- unit fixed effects make it difficult to make out-of-sample predictions.

F-test

An F-test is possible to test whether the intercepts indeed vary:

$$\frac{(R^2_{FE} - R^2_{Pooled})/(N-1)}{(1-R^2_{FE})/(NT-N-k)} \sim F(N-1, NT-N-k)$$

(Greene, 2002, 289)

Time fixed effects

When common shocks are of concern, time fixed effects are possible:

$$y_{it} = \gamma_t + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$

And one can control for both at once:

$$y_{it} = \alpha_i + \gamma_t + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$

This of course requires sufficiently large N and T.

Unbalanced panels

While the algebra is complicated when fixed effects are applied to an unbalanced panel (not all units are observed for all time periods), the application is simple and the regular regression valid.

(Greene, 2002, 293)

Random effects

For the random effects model we still have:

$$y_i = \mathbf{x}_i \boldsymbol{\beta} + \mu_{j[i]} + \varepsilon_i.$$

However, this time we assume $\mu_j \sim N(0, \sigma_\mu^2)$.

By assuming that μ_j comes from a normal distribution, we have fewer parameters to estimate (only one σ_μ^2 instead of J μ's).

Variance components

In the population, the variance of the dependent variable can be split in *within-group* and *between-group* variance:

$$\sigma_y^2 = \sigma_{between}^2 + \sigma_{within}^2$$

Random effects

When we can assume α_i and $\mathbf{x}_{it}$ are uncorrelated, we can use random effects.

$$y_{it} = \alpha_i + \mathbf{x}_{it}'\boldsymbol{\beta} + \varepsilon_{it}$$
$$\alpha_i \sim N(0, \sigma_\alpha^2)$$

Breusch-Pagan LM test

Again, a test exists whether the intercepts indeed vary ($H_0 : \sigma_\mu^2 = 0$), the Breusch-Pagan LM test:

$$\frac{NT}{2(T-1)}\left[\frac{\sum_{i=1}^{N}(\sum_{t=1}^{T} e_{it})^2}{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2} - 1\right]^2 \sim \chi(1)$$

(Greene, 2002, 299)

Hausman test

The random effects model assumes that $E(\mu_i \mathbf{x}_{it}) = \mathbf{0}$. Whether this is valid can be tested by looking at the differences in coefficients between the random and fixed models, using a Hausman test:

```
fe <- plm(inv ~ value + capital, data = Grunfeld,
model = "within")

re <- plm(inv ~ value + capital, data = Grunfeld,
model = "random")

phtest(fe, re)
```

With H_0: both models are consistent.

R^2

Variance of a multilevel model has different components:

$$\sigma_y^2 = \sigma_{between}^2 + \sigma_{within}^2$$

Estimated two models, one with and one without explanatory variables (A and B, respectively).

$$R_{within}^2 = \frac{\hat{\sigma}_{\mu,A}^2 + \hat{\sigma}_{\varepsilon,A}^2}{\hat{\sigma}_{\mu,B}^2 + \hat{\sigma}_{\varepsilon,B}^2}$$

$$R_{between}^2 = \frac{\hat{\sigma}_{\mu,A}^2 + \hat{\sigma}_{\varepsilon,A}^2/n}{\hat{\sigma}_{\mu,B}^2 + \hat{\sigma}_{\varepsilon,B}^2/n},$$

whereby n is the typical group size *in the population*.

Predicted random effects

With a fixed effects model, we have the coefficients on the group dummies which we can interpret as group-level predictors.

In a random effects model, we do not have these predictions, as we only estimated σ_μ^2 and $\boldsymbol{\beta}$.

The predicted group levels can be estimated using:

$$\hat{\beta}_{0,j} = \lambda_j \bar{y}_j + (1 - \lambda_j)\hat{\beta}_0$$
$$\lambda_j = \frac{\hat{\sigma}_\mu^2}{\hat{\sigma}_\mu^2 + \hat{\sigma}_\varepsilon^2 / n_j},$$

whereby $\bar{y}_j$ is the mean on y of group j.

Predicted random effects

In R, you get the *estimated* fixed effects with:

```
model.random <- lmer(y ~ x1 + x2 + (1|group))
fixef(model.random)
```

and the *predicted* random effects with:

```
ranef(model.random)
```

Fixed vs random

When to use random effects?

- A group effect is random if we can think of the levels we observe in that group to be samples from a larger population.
- When making out-of-sample inferences.
- When there are group-level variables.
- When the sizes of groups are small.

Fixed vs random

When to use random effects?

Alternatively, one can primarily look at n_j and N:

N small	fixed effects
N not small, n_j small	random effects
n_j larger	not as important

But this is only a preliminary quick judgment!

Fixed vs random

When to use random effects?

Gelman and Hill (2007): "Our advice (...) is to *always* use multilevel modeling ('random effects')."

Johnston and DiNardo (1997): choose random effects when you can assume that $\mathbf{x}_i$ and $\mu_{j[i]}$ are uncorrelated; fixed effects otherwise.

Exercise

Using the `asiabaro.dta` data file:

1. Estimate a country fixed effects model explaining `evaldemoc` by `evalcorrupt` using dummy variables, excluding the overall intercept.
2. Estimate a country random effects model explaining `evaldemoc` by `evalcorrupt`.

Outline

Random coefficients

In the random effects model, we assume that group intercepts vary according to a normal distribution.

But what about the coefficients?

I.e. what about group slopes that vary following a normal distribution?

Random coefficients

$$
\begin{aligned}
y_i &= \mathbf{x}_i \boldsymbol{\beta} + \mathbf{x}_i \boldsymbol{\gamma}_{j[i]} + \mu_{j[i]} + \varepsilon_i \\
\mu_j &\sim N(0, \sigma_\mu^2) \\
\gamma_j &\sim N(0, \sigma_\gamma^2)
\end{aligned}
$$

Note that a model with random coefficients, but a constant intercept across groups rarely makes sense, especially because of the often arbitrary location if $x = 0$.

Exercise

Using the `asiabaro.dta` data file:

1. Estimate a country random coefficients model explaining `evaldemoc` by `evalcorrupt`.

General multilevel model

$$
\begin{aligned}
y_i &= \mu_{j[i]} + \mathbf{x}_i \gamma_{j[i]} + \varepsilon_i, & \varepsilon_i &\sim N(0, \sigma^2) \\
\mu_j &= \beta_0^\mu + \mathbf{z}_j \beta^\mu + \varepsilon_j^\mu, & \varepsilon_j^\mu &\sim N(0, \sigma_\mu^2) \\
\gamma_j &= \beta_0^\gamma + \mathbf{z}_j \beta^\gamma + \varepsilon_j^\gamma, & \varepsilon_j^\gamma &\sim N(0, \sigma_\gamma^2)
\end{aligned}
$$

General multilevel model

$$
\begin{aligned}
y_i &= \mu_{j[i]} + \mathbf{x}_i \gamma_{j[i]} + \varepsilon_i, & \varepsilon_i &\sim N(0, \sigma^2) \\
\mu_j &= \beta_0^\mu + \mathbf{z}_j \beta^\mu + \varepsilon_j^\mu, & \varepsilon_j^\mu &\sim N(0, \sigma_\mu^2) \\
\gamma_j &= \beta_0^\gamma + \mathbf{z}_j \beta^\gamma + \varepsilon_j^\gamma, & \varepsilon_j^\gamma &\sim N(0, \sigma_\gamma^2)
\end{aligned}
$$

This model can thus incorporate both individual-level variables ($\mathbf{X}$) and group-level variables ($\mathbf{Z}$), including implicitly interactions, without assuming a perfect model at either level ($\sigma_\mu^2, \sigma_\gamma^2 > 0$).

General multilevel model

$$y_i = \mu_{j[i]} + \mathbf{x}_i \gamma_{j[i]} + \varepsilon_i, \qquad \varepsilon_i \sim N(0, \sigma^2)$$
$$\mu_j = \beta_0^\mu + \mathbf{z}_j \beta^\mu + \varepsilon_j^\mu, \qquad \varepsilon_j^\mu \sim N(0, \sigma_\mu^2)$$
$$\gamma_j = \beta_0^\gamma + \mathbf{z}_j \beta^\gamma + \varepsilon_j^\gamma, \qquad \varepsilon_j^\gamma \sim N(0, \sigma_\gamma^2)$$

Assumptions:

- $E(\varepsilon_i) = E(\varepsilon_j^\mu) = E(\varepsilon_j^\gamma) = 0$
- $V(\varepsilon_i) = \sigma^2$, $V(\varepsilon_j^\mu) = \sigma_\mu^2$, $V(\varepsilon_j^\gamma) = \sigma_\gamma^2$
- $Cov(\varepsilon_j^\mu, \varepsilon_j^\gamma) = \tau$
- $Cov(\varepsilon_j^\mu, \varepsilon_i) = Cov(\varepsilon_j^\gamma, \varepsilon_i) = 0$

Intraclass correlation

The general model also allows to calculate the **intraclass correlation**:

$$\rho_{intra} = \frac{Cov(\varepsilon_i, \varepsilon_k)}{\sqrt{V(\varepsilon_i)}\sqrt{V(\varepsilon_k)}},$$

with $i \neq k$, $j[i] = j[k]$.

Outline

Panel-Corrected Standard Errors (PCSE)

A fixed effects model assumes that the residual variance is constant across units and contains no autocorrelation.

Common problems:

- $E(\varepsilon_{it}^2) = \sigma_i^2$, $E(\varepsilon_{it}^2) \neq E(\varepsilon_{jt}^2)$ (panel heteroscedasticity)
- $E(\varepsilon_{it}\varepsilon_{jt}) = \sigma_{ij}$, with $E(\varepsilon_{it}\varepsilon_{js}) = 0 \forall t \neq s$ (contemporaneously correlated errors)

Beck and Katz (1995) suggest to first estimate the OLS fixed effects model and then correct the standard errors to take these into account.

This is similar to White's robust standard errors, but more efficient due to stronger assumptions on panel structure.

Panel-Corrected Standard Errors (PCSE)

Default variance-covariance matrix:

$$\mathbf{\Omega} = \sigma^2 \mathbf{I}_{NT} = \begin{bmatrix} \sigma^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}$$

Panel-Corrected Standard Errors (PCSE)

White's robust variance-covariance matrix:

$$\mathbf{\Omega} = \begin{bmatrix} \hat{\sigma}_1^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_2^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \hat{\sigma}_3^2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \hat{\sigma}_4^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & \hat{\sigma}_{NT}^2 \end{bmatrix}$$

With the HCC0, HCC1, HCC2, HCC3 variations in estimating $\hat{\sigma}_i^2$.

Panel-Corrected Standard Errors (PCSE)

PCSE variance-covariance matrix ($N = 4$):

$$\mathbf{\Omega} = \begin{bmatrix} \hat{\sigma}_1^2 & 0 & 0 & 0 & \hat{\sigma}_{12}^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_1^2 & 0 & 0 & 0 & \hat{\sigma}_{12}^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \hat{\sigma}_1^2 & 0 & 0 & 0 & \hat{\sigma}_{12}^2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \hat{\sigma}_1^2 & 0 & 0 & 0 & \hat{\sigma}_{12}^2 & \cdots & \hat{\sigma}_{1N}^2 \\ \hat{\sigma}_{12}^2 & 0 & 0 & 0 & \hat{\sigma}_2^2 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \hat{\sigma}_{12}^2 & 0 & 0 & 0 & \hat{\sigma}_2^2 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \hat{\sigma}_{12}^2 & 0 & 0 & 0 & \hat{\sigma}_2^2 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \hat{\sigma}_{12}^2 & 0 & 0 & 0 & \hat{\sigma}_2^2 & \cdots & \hat{\sigma}_{2N}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \hat{\sigma}_{1N}^2 & 0 & 0 & 0 & \hat{\sigma}_{2N}^2 & \cdots & \hat{\sigma}_N^2 \end{bmatrix}$$

Limited dependent variables

Multilevel models can also be estimated with limited dependent variables.

In R, the extension can actually be quite straightforward:

```
require(lme4)
glmer(y ~ x1 + x2 + (1|group),
     family=binomial(link="logit"))
glmer(y ~ x1 + x2 + (1|group), family=poisson)
```

Logit and probit

When T fixed, $N \to \infty$, a fixed effects specification like

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

is infeasible, as the number of parameters increases with N.

With linear models, this can be handled by taking first-differences.

In logit or probit, all coefficients interact, so this cannot work. Solutions are beyond the scope of this class.

With N fixed, $T \to \infty$, problems are not major.

Markov chains

When the dependent variable is categorical, there are different ways of modeling this (e.g. survival analysis). One approach is that of a Markov chain.

Transition matrix:

$$\begin{vmatrix} \pi_{11} & \pi_{01} \\ \pi_{10} & \pi_{00} \end{vmatrix} = \begin{vmatrix} Pr(y_{it} = 1 | y_{i,t-1} = 1) & Pr(y_{it} = 1 | y_{i,t-1} = 0) \\ Pr(y_{it} = 0 | y_{i,t-1} = 1) & Pr(y_{it} = 0 | y_{i,t-1} = 0) \end{vmatrix}$$

Markov chains

$$Pr(y_{it} = 1) = \mathcal{F}(y_{i,t-1}(\mathbf{X}_t\beta) + (1 - y_{i,t-1})(\mathbf{X}_t\alpha))$$

$\mathcal{F}()$ can be a logistic or probit transformation.

For examples, see Przeworski and Limongi (1997); Gleditsch and Ward (2006); Elkink (2013).

Lagged dependent variable

A straightforward way of modeling time dependence is adding a lagged dependent variable:

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \rho y_{i,t-1} + \varepsilon_{it}$$

In this case, ε_{it} is by definition correlated with one of the independent variables ($y_{i,t-1}$).

The OLS estimation is biased, but consistent as $T \rightarrow \infty$. Hence, this is problematic with large N, but small T.

(Greene, 2002, 307)

Arellano-Bond GMM estimator

First, the data is first-differenced to get rid of unit fixed effects:

$$y_{it} = \alpha_i + \mathbf{x}'_{it}\boldsymbol{\beta} + \rho y_{i,t-1} + \varepsilon_{it}$$
$$\Delta y_{it} = \Delta \mathbf{x}'_{it}\boldsymbol{\beta} + \rho \Delta y_{i,t-1} + \Delta \varepsilon_{it}$$

OLS is inconsistent because $E(\Delta\varepsilon_{it}\Delta y_{i,t-1}) \neq 0$. An **instrumental variable** model can fix this, with lags depending on t:

$$t = 3 : y_1$$
$$t = 4 : y_1, y_2$$
$$t = 5 : y_1, y_2, y_3$$

(GMM and IV estimators are beyond the scope of this course)

Arellano-Bond GMM estimator

This estimator is primarily useful for $N >> T$, such that it is consistent for T fixed, $N \rightarrow \infty$.

It is implemented in R using the `pgmm()` function - see `plm` documentation for an example.

In political science, we more often deal with $T > N$.

Error correction model

(From time-series class)

$$
\begin{aligned}
y_i &= \alpha + \beta_1 y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t \\
\Delta y_t &= \alpha + (\beta_1 - 1) y_{t-1} + \gamma_0 x_t + \gamma_1 x_{t-1} + u_t \\
&= \alpha + (\beta_1 - 1)(y_{t-1} + \frac{\gamma_0 + \gamma_1}{\beta_1 - 1} x_{t-1}) + \gamma_0 \Delta x_t + u_t \\
&= \alpha + (\beta_1 - 1)(y_{t-1} - \lambda x_{t-1}) + \gamma_0 \Delta x_t + u_t,
\end{aligned}
$$

with $\lambda = \frac{\gamma_0 + \gamma_1}{1 - \beta_1}$, the long-run multiplier.

(Davidson and MacKinnon, 1993, 683); (Greene, 2003, 579)

Error correction model

$$y_{it} = \alpha_i + \rho y_{i,t-1} + \mathbf{x}'_{it}\boldsymbol{\beta} + \mathbf{x}'_{i,t-1}\boldsymbol{\gamma} + \varepsilon_{it},$$

with $\boldsymbol{\lambda} = \frac{1}{1-\rho}(\boldsymbol{\beta} + \boldsymbol{\gamma})$, the long-run multiplier.

This can be estimated by:

- Using the original equation (possibly Arellano-Bond GMM)
- Using a non-linear estimator and estimate the error-correction specification directly

Examples: Fish and Choudhry (2007); Gans-Morse and Nichter (2008).

Exercise

A famous example dataset is the Grunfeld data:

inv	gross investment
value	market value of firm
capital	value of stock of equipment / plant

for 10 firms ($N = 10$) over 20 years ($T = 20$).

```
library(plm); data("Grunfeld", package = "plm"); head(Grunfeld)
```

- Using `plm()`, estimate the pooled, unit fixed, time fixed, random unit effects models.
- Test the fixed and random effects specifications.
- Calculate panel-corrected standard errors.
- Graphically inspect the regression results.

Example

F-test for fixed effects:

```
m.pooled <- plm(inv ~ value + capital, data = Grunfeld,
model = "pooling", index = c("firm", "year"))

m.fe <- plm(inv ~ value + capital, data = Grunfeld,
model = "within", index = c("firm", "year"))

pFtest(m.fe, m.pooled)
```

Example

Random effects:

```
lmer(inv ~ value + firm + (1|firm), data = Grunfeld)

plm(inv ~ value + firm, data = Grunfeld,
model = "random", index = c("firm", "year"))

pbgtest(inv ~ value + firm, data = Grunfeld,
model = "random", index = c("firm", "year"))
```

Example

PCSE:

```
library(pcse)

m.ols <- lm(inv ~ value + capital, data = Grunfeld)
m.pcse <- pcse(m.ols, Grunfeld$firm, Grunfeld$year)
summary(m.pcse)

m.fe <- lm(inv ~ value + capital + factor(firm),
data = Grunfeld)
m.pcse.fe <- pcse(m.fe, Grunfeld$firm, Grunfeld$year)
summary(m.pcse.fe)
```

Further information

A clear, relatively introductory textbook on multilevel modeling is Snijders and Bosker (1999), *Multilevel analysis. An introduction to basic and advanced multilevel modeling*.

Snijders: http://stat.gamma.rug.nl/snijders/

An excellent, modern book on multilevel modeling, using primarily R and Bugs, is Gelman and Hill (2007), *Data analysis using regression and multilevel/hierarchical models*.

Gelman: http://www.stat.columbia.edu/∼gelman/

Beck, Nathaniel L and Jonathan N Katz. 1995. "What to do (and not to do) with time-series cross-section data." *American political science review* 89(3):634–647.

Davidson, Russell and James G. MacKinnon. 1993. *Estimation and inference in econometrics*. Oxford: Oxford University Press.

Elkink, Johan A. 2013. "Spatial, temporal and spatio-temporal clustering of democracy." Paper presented at the American Political Science Association, Annual Meeting, Chicago, August 29-September 1.

Fish, M. Steven and Omar Choudhry. 2007. "Democratization and economic liberalization in the Postcommunist world." *Comparative Political Studies* 40(3):254–282.

Gans-Morse, Jordan and Simeon Nichter. 2008. "Economic reforms and democracy: Evidence of a J-curve in Latin America." *Comparative Political Studies* 41(10):1398–1426.

Gelman, Andrew and Jennifer Hill. 2007. *Data analysis using regression and multilevel/hierarchical models*. Analytical Methods for Social Research Cambridge: Cambridge University Press.

Gleditsch, Kristian S. and Michael D. Ward. 2006. "Diffusion and the international context of democratization." *International Organization* 60(4):911–933.

Greene, William H. 2002. *Econometric analysis*. London: Prentice Hall.

Greene, William H. 2003. *Econometric analysis*. 5th ed. Upper Saddle River: Prentice Hall.

Johnston, Jack and John DiNardo. 1997. *Econometric methods*. New York: McGrawHill.

Przeworski, Adam and Fernando Limongi. 1997. "Modernization: theories and facts." *World Politics* 49(2):155–183.

Snijders, Tom and Roel Bosker. 1999. *Multilevel analysis: an introduction to basic and advanced multilevel modeling*. Sage.