Applied Economics

Panel Data

Universidad Carlos III de Madrid

Introduction: Panel Data

- See also Wooldridge (chapter 13) and Stock & Watson (chapter 10).
- Panel or longitudinal data consist of a sample of several observations in subsequent time periods on the same individual units.
  - Each individual unit can refer to an individual, a household, a company, a province, a country, etc.
- Unlike repeated cross sections, in which we have several time periods but observations in each time period correspond to different individuals, with panel data we follow the same individuals over several time periods.
- In general, we consider panel data applications in which the number of individuals (n) is very large while the number of time periods we observe them (T) is shorter.

Introduction: Panel Data (cont.)

- To characterize a population of interest with panel data, we start with a survey to a random sample of individuals in a certain period and collect the same information again in subsequent periods (month, quarter, year).
- An example for Spain is the "Encuesta de Condiciones de Vida" (Living Conditions Survey) from INE. The first survey started on 2004 with a sample size around 15,000 households, and each household is interviewed on an annual basis over four consecutive years.
- With panel data we can follow the evolution over time of particular variables for the same individual units.
- As we have observations for the same individual units over several time periods, such observations may show time dependence.
    - For example, unobserved individual factors that affected an individual's wage in 2014 might also affect her wage in 2015.

Advantages of Panel Data

- Panel data has two major advantages with respect to repeated cross sections.
- These advantages arise from the fact panel data provides several observations of the same individuals over time.
- First, we can control for the potential bias due to **unobserved (individual) heterogeneity**.
- Second, we can estimate **dynamic models**.
  - However, in this course we will focus on **static** models, and estimation of dynamic models will not be discussed

Advantages of Panel Data (cont.)

- **Unobserved (individual) heterogeneity** consist of omitted individual characteristics that differ among individuals but are constant over time.
  - With several observations of the same individual units over time we can overcome **unobserved heterogeneity bias** using special methods and model transformations.
  - This problem cannot be addressed with cross sections and repeated cross sections, as we only observe each individual once.
- **Dynamic models** (well-known in time-series econometrics) include the **lagged dependent variable as explanatory variable**.
  - Panel data provides a (usually short) time series for each individual.
  - With panel data, denoting i for individual unit and t time period, given a sample for two variables $\{X_{it}, Y_{it}\}$, $(i = 1, \ldots, n;\ t = 1, \ldots, T)$, we can consider specifications like

$$Y_{it} = \beta_0 + \beta_1 Y_{i,t-1} + \beta_2 X_{it} + v_{it}.$$

The linear model with panel data

- Consider a sample for two variables X and Y, $\{X_{it}, Y_{it}\}$, $i = 1, \ldots, n$; $t = 1, \ldots, T$.
  - Subscript i refers to the individual unit (individual, household, firm, etc.)
  - Subscript t refers to the time period.
- Consider, for $i = 1, \ldots, n$; $t = 1, \ldots, T$, the linear model

$$Y_{it} = \beta_0 + \sum_{s=2}^{T} \delta_s ds_t + \beta_1 X_{it} + \underbrace{a_i + u_{it}}_{v_{it}}$$

where

- ds_t is a binary variable taking value 1 at period $t = s$ and 1 at $t \neq s$ (common to all units, so it doesn't have subscript i).
  It captures aggregate shocks common to all individuals.
- $v_{it} = a_i + u_{it}$ is the unobserved random error (composed of two terms).

The linear model with panel data (cont.)

- The random **composite** error $v_{it} = a_i + u_{it}$ is broken down into:
  - a_i: unobserved random variable capturing all unobserved individual factors affecting Y_{it} that do not change over time (so it has no subscript t).
  - u_{it}: usual (idiosynchratic) random error that includes any unanticipated shock or unobserved factor affecting Y_{it} that differs both among individuals and over time.
- a_i captures unobserved time-invariant differences among individuals (called unobserved heterogeneity, individual heterogeneity, or individual fixed effect).
- The properties of the model will depend on the relation of the explanatory variable X_{it} with the unobserved error term v_{it}.
  - We will have to consider such relation with each component, a_i and u_{it}.

Exogeneity in the panel data context

- We first start with the relation of X_{it} with the (usual) idiosynchratic shock u_{it}.
- We have **contemporaneous exogeneity** when $C(X_{it}, u_{it}) = 0$.
  - In the cross section framework (or even with repeated cross sections), **contemporaneous exogeneity** suffices to get consistent estimates of the parameters of interest.
- We have **strict exogeneity** when $C(X_{it}, u_{is}) = 0$, $\forall t, s = 1, \ldots, T$.

Exogeneity in the panel data context (cont.)

- Strict exogeneity is much more stringent than contemporaneous exogeneity, as it requires:
  - (i) X_{it} uncorrelated with future shocks, i.e. $C\left(X_{it}, u_{is}\right) = 0,\ s > t$.
  - (ii) X_{it} uncorrelated with current shocks, i.e. $C\left(X_{it}, u_{it}\right) = 0$ (contemporaneous exogeneity).
  - (iii) X_{it} uncorrelated with past shocks, i.e. $C\left(X_{it}, u_{is}\right) = 0,\ s < t$.
- Condition (i) of no correlation of the explanatory variable(s) with future shocks is expected to hold, as unanticipated shocks are not expected to affect realizations of X happened before.
- But condition (iii) will be very demanding in many instances, as we require X to be unaffected by shocks occurred in the past.

Unobserved heterogeneity

- Under which conditions does pooled OLS provide a consistent estimate of the parameter of interest β_1?
- Let's start assuming contemporaneous exogeneity of X: $C(X_{it}, u_{it}) = 0$.
  - But this condition does not ensure the pooled OLS estimator of β_1 is consistent.
- Under contemporaneous exogeneity, the relevant issue is whether a_i is uncorrelated with X_{it}.
  - If in addition to $C(X_{it}, u_{it}) = 0$ we have $C(X_{it}, a_i) = 0$, then

$$C(X_{it}, v_{it}) = C(X_{it}, a_i + u_{it}) = \underbrace{C(X_{it}, a)}_{0} + \underbrace{C(X_{it}, u_{it})}_{0} = 0$$

    so pooled OLS will yield consistent estimates.
  - But if on the contrary $C(a_i, X_{it}) \neq 0 \Rightarrow C(X_{it}, v_{it}) \neq 0$ so pooled OLS will be inconsistent.
- The bias for omitting a_i is called **unobserved heterogeneity bias**. It results from ignoring a_i when $C(a_i, X_{it}) \neq 0$.

Unobserved heterogeneity: Examples

- Production function of a crop field
  - $Y_{it} = \log\left(\text{Output}/\text{m}^2\right)$, $X_{it} = \log\left(\text{Fertilizer}/\text{m}^2\right)$; a_i = Land's quality (time-invariant, known by the farmer but unkown by the analyst); u_{it} = shocks affecting output crop (rainfall, drought, landslides, etc.)
  - The amount of fertilizer by m^2 is expected to be correlated with land's quality.
- Effect of training on workers' earnings
  - $Y_{it} = \log(\text{Annual earnings})$, $X_{it} = \log(\text{Hours of training})$; a_i = Individual's ability; u_{it} = shocks affecting annual earnings (firm's profitability, labor conflicts, etc.)
  - If the hours of training are chosen by each worker, its amount is expected to be correlated with individual ability.

The false dichotomy of random effects vs fixed effects

- The traditional, old-fashion, approach presents two mutually exclusive points of view about the individual effects a_i:
  - "**Random effects**": the a_i's $(i = 1, \ldots, n)$ are random variables, assumed to be uncorrelated with the explanatory variable(s) (so $C(a_i, X_{it}) = 0$).
  - "**Fixed effects**": the a_i's $(i = 1, \ldots, n)$ are n unknown (fixed) parameters that characterize a different intercept for each individual. Disregarding them will lead to inconsistent estimates.
- However, this is a false dichotomy: a_i is **always a random variable**.
- The **key issue** is whether the individual effects are uncorrelated or not with the explanatory variable(s).
  - If $C(a_i, X_{it}) = 0$, the pooled OLS estimator (that ignores a_i) will be consistent.
  - If $C(a_i, X_{it}) \neq 0$, the pooled OLS estimator will be inconsistent.
    We'll need a transformation of the original model that removes the individual effects a_i to obtain consistent estimates.

Panel data with two time periods

- Consider the simplest panel with $T = 2$,

$$\begin{aligned} Y_{it} &= \beta_0 + \delta_2 d2_t + \beta_1 X_{it} + v_{it} \\ v_{it} &= a_i + u_{it} \end{aligned}$$

- For pooled OLS be consistent for β_1, we need $C(X_{it}, v_{it}) = 0$, for which contemporaneous exogeneity of X, $C(X_{it}, u_{it}) = 0$ is a necessary but not sufficient condition.
  - Even so, if $C(X_{it}, a_i) \neq 0$, OLS will be inconsistent.
- Lack of controlling for omitted a_i in pooled OLS leads to inconsistency due to **heterogeneity bias**.

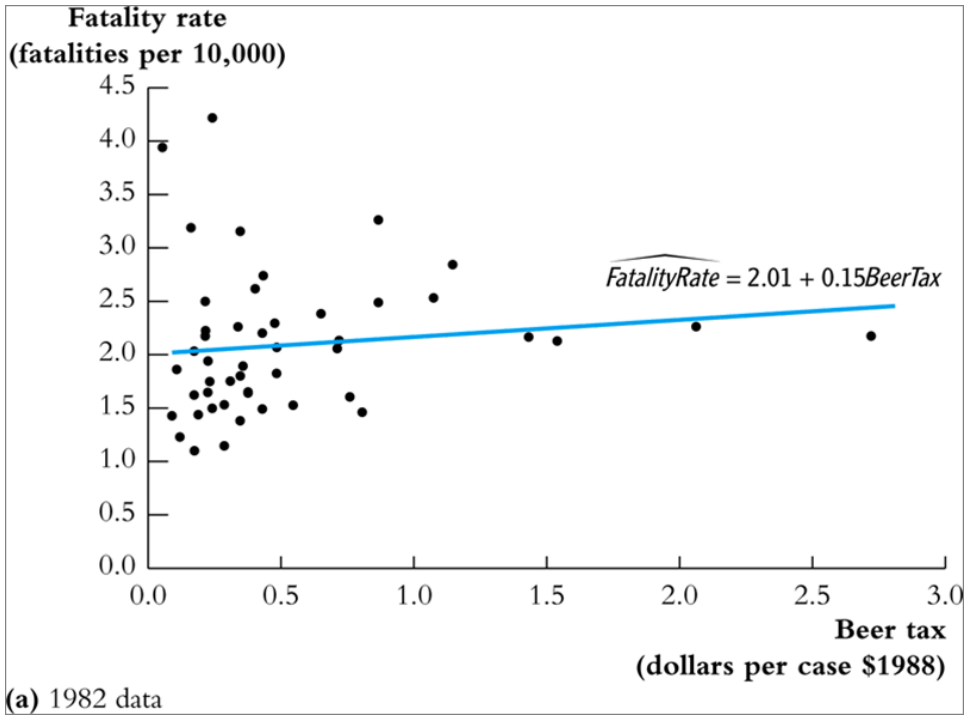
Example: Alcohol Taxation and Traffic Deaths

- Goal: Analyze policies about alcohol consumption to reduce traffic mortality.
- Data `fatality.gdt` in `gretl`: 48 US states, 7 years (1982 to 1988).
- We use only 1982 and 1988.
- Variables:
  - Traffic Mortality: rate of traffic deaths per 10,000 residents (variable *TM*).
  - Tax on a case of beers in 1988 dollars (variable *beertax*).

Example: Alcohol Taxation and Traffic Deaths

Using a single cross section

- Simple model for year 1982: $TM_i = \beta_0 + \beta_1 beertax_i + u_i$



- The OLS estimated coefficient for the 1982 cross section $\widehat{\beta}_1 = 0.15$ is positive (though statistically non significant).

Example: Alcohol Taxation and Traffic Deaths

Potential explanations to the estimation results

- Using pooled cross sections for 1982 and 1988, the OLS estimated coefficient is positive and significant: Check with `gretl`!!
  - Can we conclude that taxing beer increases the number of traffic fatalities?
- There are many (omitted) state-specific factors behind traffic fatalities, like distribution of automobiles' age, quality of roads, culture of drinking and driving, traffic density...
- If some of these factors are correlated with tax on beer, pooled OLS estimates will be inconsistent.
- Although we could pick data on some of these factors, some other (e.g. culture) are difficult to get.
- These omitted state-specific factors vary little (or not at all) in a 10-year time span.
- If such factors remain constant in the sample period, panel data allows to obtain consistent estimators.

The FD transformation

- Most likely, unobserved individual effects a_i are correlated with the explanatory variable(s).
- We can exploit the fact that whereas Y_{it}, X_{it} vary over time, a_i doesn't.
- When $T = 2$, we can write for each of the two periods

$$\begin{aligned} Y_{i2} &= \beta_0 + (\delta_2 \times 1) + \beta_1 X_{i2} + a_i + u_{i2} \\ Y_{i1} &= \beta_0 + (\delta_2 \times 0) + \beta_1 X_{i1} + a_i + u_{i1} \end{aligned}$$

  where we have used the fact that $d2_t$ takes value 1 when $t = 2$ and 0 when $t = 1$.
- If we substract the first-period equation to the second period equation, we get the transformed first-differences equation:

$$(Y_{i2} - Y_{i1}) = \delta_2 + \beta_1 (X_{i2} - X_{i1}) + (u_{i2} - u_{i1}).$$

The First-Differences (FD) estimator

- Equivalently, we can use the Δ operator to denote time change $\Delta Z_{i2} = Z_{i2} - Z_{i1}$.
- When $T = 2$, this equation in first differences (FD) is a single cross-sectional equation, with each original variable transformed in FD.

$$\Delta Y_{i2} = \delta_2 + \beta_1 \Delta X_{i2} + \Delta u_{i2}$$

- Notice that we have got rid of a_i thanks to the FD transformation: $\Delta a_i = (a_i - a_i) = 0$.
- **Key condition**: $C(\Delta X_{i2}, \Delta u_{i2}) = 0$.
  - If Δu_{i2} is uncorrelated with ΔX_{i2}, OLS applied to the FD equation will yield a consistent estimate of β_1.
  - **Question:** For the key condition to hold, does it suffice $C(X_{it}, u_{it}) = 0$?
- The OLS estimator in the FD equation is called the First-Differences (FD) estimator.

About the FD estimator

- Provided the earlier **key condition** is true, the FD estimator is consistent under any form of correlation between a_i and the controls.
- Under general regularity conditions, $p\lim_{N\to\infty}\left(\widehat{\beta}_1^{FD}\right) = \beta_1$ for T fixed.
- There is no need to estimate the unobserved individual effects a_i $(i = 1, \ldots, n)$.
  - They are dropped when taking FD to the original equation.
- The FD estimator requires ΔX_{i2} to vary across units.
  - If the variability of ΔX_{i2} is small, the FD estimator will yield imprecise estimates (because the variance of $\widehat{\beta}_1^{FD}$ will be high).
- Sometimes, it may make sense to take differences between periods farther away to get enough variability over time in X_{it}.

Example: Alcohol Taxation and Traffic Deaths

Exploiting panel data

- Consider years 1982 and 1988:

$$TM_{it} = \beta_0 + \delta_2 d2_t + \beta_1 beertax_{it} + a_i + u_{it}, \quad t = 1982, 1988.$$

- a_i represents unobserved state-specific time-invariant variables affecting fatality rate in state i.
  - It may include local attitudes towards driving while drunk.
  - Of course, a_i might capture other unobserved time-invariant state-specific features affecting fatalities. **Can you suggest any?**
  - We can consider them as constant between 1982 and 1988 if they change very slowly.
- Most likely, states with a less favorable attitude towards driving while drunk will have on average less traffic fatalities... and maybe higher alcohol taxes too.
- This is a clear case of omitted variable bias: *beertax* partly captures the (indirect) effect of the dominant local attitude towards driving while drunk.

Example: Alcohol Taxation and Traffic Deaths

First differences

- The unobserved time-invariant individual effects a_i can be removed taking FD

$$\Delta TM_{it} = \delta_2 + \beta_1 \Delta beertax_{it} + \Delta u_{it}, \quad t = 1988.$$

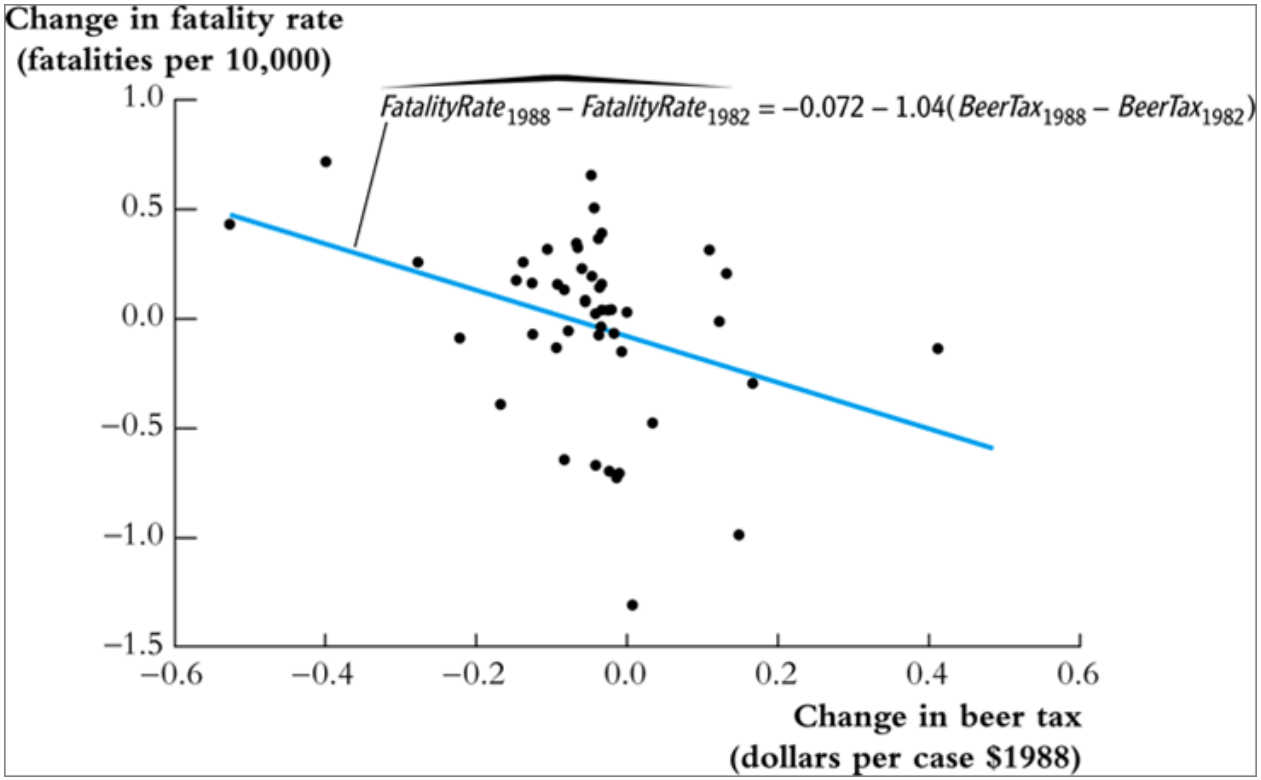
  Notice that the FD transformation removes any time-invariant term, like the constant term β_0 and the individual effect a_i.
  Also, $\Delta \delta_2 d2_t = \delta_2 \times (1 - 0) = \delta_2$.

- Summarizing:
  - Attitudes about driving while drunk (contained in a_i) affect the proportion of drunk drivers and thus the rate of traffic fatalities in the state.
  - If such attitudes barely change in a moderately large time span (e.g., 1982 to 1988), **they could not affect the change** in the state's traffic fatality rate.
  - Hence, the time change in TM_{it} (between 1982 and 1988), ΔTM_{it}, must be due to other factors, like changes in state beer taxes $beertax_{it}$ and in idyosynchratic shocks u_{it}.

Example: Alcohol Taxation and Traffic Deaths

FD estimation results

- FD model for year 1988: $\Delta TM_{i2} = \delta_2 + \beta_1 \Delta beertax_{i2} + \Delta u_{i2}$.



- Unlike the pooled OLS estimates, the FD estimation shows that increasing alcohol taxes reduces traffic fatalities.

Example: Alcohol Taxation and Traffic Deaths

FD estimation results. Comments

- The intercept measures the effect on the fatality rates of aggregate effects common to all states.
- A 1$ increase in the beer tax per case of beers reduces average traffic mortality by 1.04 deaths per $10,000$ residents.
- The estimated effect of the beer tax is significant and relevant:
  - The sample average annual fatality rate is 2 deaths per $10,000$ residents.
  - Hence, the estimated reduction of 1.04 deaths suggests a 1$ tax increase per beer case might reduce the fatality rate by half (50%).

Extensions and limitations of the FD estimator

- The extension of the FD estimator with several explanatory variables is straightforward.
- As with cross section data, **heteroskedasticity** (non-constant conditional variance of the idiosynchratic error u_{it}) is very likely.
  Hence, we should compute **heteroskedasticity-robust standard errors**.
- The FD estimator removes any time-invariant factor, either unobserved (a_i) or observed.
  - Hence, the FD estimator cannot provide the estimated effects of any observed time-invariant explanatory variable.
- The FD estimator requires sufficient time variation in explanatory variables.
  - In the traffic fatalities example, we need time variation in *beertax* in several states to estimate its effect with sufficient precision.

Extensions and limitations of the FD estimator (cont.)

- Consistency of the FD estimator requires the **key condition** $C(\Delta X_{it}, \Delta u_{it}) = 0$.
  - Notice that contemporanous exogeneity does not suffice.
- There may be other individual time-varying factors, correlated with the explanatory variable(s), which affect Y_{it}.
  If such factors are omitted, then $C(X_{it}, u_{it}) \neq 0$, and $C(\Delta X_{it}, \Delta u_{it}) \neq 0$ too.
  - In such case, the FD estimator will suffer from endogeneity bias and will be inconsistent.
  - Solutions to this problem would require the use of IV/2SLS FD estimation, which are not discussed in this course.

FD with more periods

$$Y_{it} = \beta_0 + \sum_{s=2}^{T} \delta_s ds_t + \beta_1 X_{it} + a_i + u_{it}, \quad i = 1, \ldots, n; \; t = 1, \ldots, T.$$

- It is possible to apply the earlier FD strategy when $T > 2$, just by taking differences from adjacent periods (between t and $t-1$).
- As with $T = 2$, the **key condition** is $C(\Delta X_{it}, \Delta u_{it}) = 0 \; \forall t \geq 2$.
  - Notice $C(\Delta X_{it}, \Delta u_{it}) = C(X_{it} - X_{i,t-1}, u_{it} - u_{i,t-1}) =$
    $C(X_{it}, u_{it}) - C(X_{it}, u_{i,t-1}) - C(X_{i,t-1}, u_{it}) + C(X_{i,t-1}, u_{i,t-1})$.
  - Then, the **key condition** requires $C(X_{it}, u_{it}) = 0$, $C(X_{it}, u_{i,t-1}) = 0$,
    $C(X_{it}, u_{i,t+1}) = 0$,
    or equivalently $C(X_{it}, u_{i,t+s}) = 0$, $\forall t, s = -1, 0, 1$.

FD with more periods (cont.)

- For the sake of simplicity, consider $T = 3$,

$$Y_{it} = \beta_0 + \delta_2 d2_t + \delta_3 d3_t + \beta_1 X_{it} + a_i + u_{it}, \qquad t = 1, 2, 3.$$

- We are interested in estimating β_1.
- If $C(X_{it}, a_i) \neq 0$, pooled OLS will be inconsistent, even though $C(X_{it}, u_{it}) = 0$.
- **Key condition** for consistency of the FD estimator: $C(\Delta X_{it}, \Delta u_{it}) = 0 \Leftrightarrow C(X_{it}, u_{i,t+s}) = 0, \ \forall t, s = -1, 0, 1.$

FD with more periods (cont.)

- Applying FD to get rid of a_i,

$$\Delta Y_{it} = \delta_2 \Delta d2_t + \delta_3 \Delta d3_t + \beta_1 \Delta X_{it} + \Delta u_{it}, \quad t = 2, 3.$$

  which is equivalent to include a constant term and a single time dummy:

$$\Delta Y_{it} = \alpha_0 + \alpha_3 d3_t + \beta_1 \Delta X_{it} + \Delta u_{it}, \quad t = 2, 3.$$

- In general, when $T > 3$ we will have $T - 1$ observations per individual unit, and include $T - 2$ time dummies

$$\Delta Y_{it} = \alpha_0 + \sum_{s=3}^{T} \alpha_s ds_t + \beta_1 \Delta X_{it} + \Delta u_{it}, \quad t = 2, \ldots, T.$$

Other transformation to remove individual effects

- There are alternative transformations to get rid of the unobserved individual heterogeneity term.
- Consider the simple linear model, abstracting from aggregate time effects

$$Y_{it} = \beta_0 + \beta_1 X_{it} + a_i + u_{it}, \quad i = 1, \ldots, n; \; t = 1, \ldots, T.$$

- We can consider,for each individual, the average over time of the expression above

$$\overline{Y}_i = \beta_0 + \beta_1 \overline{X}_i + a_i + \overline{u}_i, \quad i = 1, \ldots, n,$$

where $\overline{Y}_i = \frac{1}{T}\sum_{t=1}^{T} Y_{it}$, $\overline{X}_i = \frac{1}{T}\sum_{t=1}^{T} X_{it}$, $\overline{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_{it}$.

- If we substract the average equation to the original one at each period t, we get

$$\left(Y_{it} - \overline{Y}_i\right) = \beta_1 \left(X_{it} - \overline{X}_i\right) + \left(u_{it} - \overline{u}_i\right), \quad i = 1, \ldots, n; \; t = 1, \ldots, T.$$

The WG estimator

- The original variables for each individual unit i have been transformed in deviation with respect to the individual means $\widetilde{Y}_{it} = Y_{it} - \overline{Y}_i$, $\widetilde{X}_{it} = X_{it} - \overline{X}_i$, etc.

$$\widetilde{Y}_{it} = \beta_1 \widetilde{X}_{it} + \widetilde{u}_{it}, \quad i = 1, \ldots, n; \; t = 1, \ldots, T.$$

- This transformation is called within-group (WG) transformation.
- The individual effect a_i is removed, since its WG transformation is equal to 0.
- The pooled OLS estimation of the WG transformation of the model yields the WG estimator.
- The **key condition** for consistency of the WG estimator is $C\left(\widetilde{X}_{it}, \widetilde{u}_{it}\right) = 0$, for which we need **strict exogeneity**!!

WG or Fixed Effects (FE) estimation

- The WG estimator is also known as fixed-effects (FE) estimator.
- The reason is that estimating by OLS the WG transformation of the original model is equivalent to estimate by OLS the original model augmented with a set of additional binary (dummy) variables for each individual unit i:

$$Y_{it} = \sum_{i=1}^{n} \gamma_i D_i + \beta_1 X_{it} + u_{it}, \quad i = 1, \ldots, n; \; t = 1, \ldots, T,$$

  where D_i is a *fixed effect* binary variable, which takes on value 1 if observation corresponds to individual i and 0 otherwise.
- In fact, although each fixed effect $\gamma_i D_i$ is not really a fixed parameter, it is constant in the sample period.
- The fixed effect term is an old-fashioned view that considers unobserved individual effects a_i as nuisance unknown parameters.
- Unobserved individual effects must be viewed as individual-specific random variables that barely change over the sample period.

Properties of the WG estimator

- The WG or FE estimator is consistent if $C\left(X_{it}, u_{is}\right) = 0 \ \forall t, s$ (**strict exogeneity**).
  - This key condition is very strong and it cannot be held in many situations.
- Both the FD and the WG estimator allow for any kind of correlation between the individual effects a_i and the explanatory variables.
  - Also, both FD and WG transformations remove any time-invariant variable in the model. Hence, FD and WG estimators do not identify the coefficient of any observed time-invariant variable.
- In the most usual case of large n and small T, the choice between FD or WG estimators relies on the relative efficiency of each estimator (since both will be consistent under strict exogeneity).
  - When $T > 2$ the WG estimator is generally more efficient than the FD estimator.
  - When $T = 2$, the WG and the FD estimators are numerically identical.

Example: Alcohol Taxation and Traffic Deaths

WG estimation results. Comments

- WG model transformación for year 1988: $\widetilde{TM}_{it} = \beta_1 \widetilde{beertax}_{it} + \widetilde{u}_{it}$ with $t = 1982, 1988$.
- We must first tell `gretl` that we have panel data, declaring the variables identifying the individual unit (in this case, `state`) and the time period (in this case, `year`):
  `setobs state year --panel-vars`
- Then, we use the command `panel`. We can use the option `--time-dummies` to include them in the model estimation.
- We get $\widehat{\beta}_1^{WG} = -1.04097$, with a standard error of 0.3475.
  - The IG estimate is identical to the pooled OLS estimate of the original model augmented with state dummies.
  - **Question**: How do the WG and the FD estimates compare in this case?

"Random Effects" (RE)

- Usually, the individual effects a_i are correlated with the explanatory variables.
- The WG estimator is consistent irrespective on whether $C(a_i, X_{it}) = 0$ or not, provided that the **key condition** $C(X_{it}, u_{is}) = 0 \forall t, s = 1, \ldots, T$ is held.
- Nonetheless, it must be noticed that $C(a_i, X_{it}) = 0$ is very unlikely.
- When $C(a_i, X_{it}) = 0$ and the key condition above holds:
  - The WG estimator is consistent but inefficient.
  - Pooled OLS (in the untransformed model) is consistent too, but inefficient, as it ignores the **structure of the composite error term** $a_i + u_{it}$.
- Traditionally, this framework has been wrongly labelled as "random effects", but it should be called **uncorrelated random effects**.

(Uncorrelated) Random Effects (RE)

- Consider the simplest model

$$Y_{it} = \beta_0 + \beta_1 X_{it} + a_i + u_{it}, i = 1, \ldots, n; t = 1, \ldots, T.$$

- If $C(X_{it}, u_{it}) = 0$ and $C(X_{it}, a_i) = 0$, pooled OLS estimator of the model above will yield a consistent estimate of β_1.
  - In fact, it suffices $T = 1$ (a single cross section) to get a consistent estimate.
- However, the composite error term $v_{it} = a_i + u_{it}$ is serially correlated over time.
- Assuming homoskedasticity, and lack of serial correlation in u_{it} ($C(u_{it}, u_{is}) = 0, \forall t \neq s$), denoting $\sigma_a^2 = V(a_i)$, $\sigma_u^2 = V(u_{it})$:
  - $V(v_{it}) \equiv V(a_i + u_{it}) = \sigma_a^2 + \sigma_u^2$.
  - $C(v_{it}, v_{is}) = C(a_i + u_{it}, a_i + u_{is}) = \sigma_a^2, \forall t \neq s$.
- Pooled OLS will be inefficient as it ignores this serial correlation.

The "Random Effects" (RE) Estimator

- Instead, we can use Feasible Generalized Least Squares (Feasible GLS), which accounts for serial autocorrelation (we don´t see the details here).
- The so-called RE (Random Effects) estimator is a FGLS estimator that consists on estimating by OLS the following transformation of the model

$$Y_{it} - \lambda \overline{Y}_i = \beta_0 (1 - \lambda) + \beta_1 \left(X_{it} - \lambda \overline{X}_i \right) + (v_{it} - \lambda \overline{v}_i)$$

  where $\overline{Z}_i = \frac{1}{T} \sum_{t=1}^{T} Z_{it}$ denotes the average of Z_{it} over time.
- λ is an unknown parameter that lies between 0 and 1, which verifies $(1 - \lambda)^2 = 1 / \left(1 + T\sigma_a^2 / \sigma_u^2\right)$. Hence, λ is increasing with the relative variance σ_a^2 / σ_u^2.
  - The smaller σ_a^2 / σ_u^2 the closer λ will be to 0, and the RE estimator will be alike the pooled OLS estimator.
  - The higher σ_a^2 / σ_u^2 the closer λ will be to 1, and the RE estimator will be alike the WG or FE estimator.

The "Random Effects" (RE) Estimator (cont.)

- To obtain the RE estimator we need to estimate λ, for which we have to estimate σ_a^2 and σ_u^2. Such estimates can be obtained from the WG estimator.
- Computation of the RE estimator is straightforward in `gretl`.
- However, this framework is quite unrealistic, as it requires the idiosynchratic error u_{it} to be serially uncorrelated over time and homoskedastic!!
- If we have heteroskedasticity and/or serial correlation, the RE transformation is not optimal. Hence, the RE estimator is no longer the FGLS estimator and therefore is not efficient.
- In practice, it is unlikely that the former assumptions to ensure consistency and efficiency of the RE estimator are true.
- In particular, the individual effects are usually correlated with the explanatory variables, so the RE estimator is inconsistent (and will differ very much from the WG estimator).

"Fixed Effects" vs "Random Effects"

- Suppose that the **key condition** $C(X_{it}, u_{is}) = 0 \forall t, s = 1, \ldots, T$ holds, and the idyosynchratc error is homoskedastic and serially uncorrelated.
- Consider the null hypothesis $H_0 : C(X_{it}, a_i) = 0$ vs the alternative $H_1 : C(X_{it}, a_i) \neq 0$.

Estimation Method	Hypothesis about $C(X_{it}, a_i)$	
	$H_0 : C(X_{it}, a_i) = 0$	$H_1 : C(X_{it}, a_i) \neq 0.$
WG or FE	consistent	consistent
RE	consistent efficient	inconsistent

- The WG/FE estimator is consistent both under the null and under the alternative. But if the null is true, then the RE estimator is consistent too and more efficient than the WG/FE estimator.
- Under no correlation between the unobserved individual effects and the explanatory variables, the RE estimator is preferable to the WG/FE estimator.

Testing "Fixed Effects" vs "Random Effects"

- If the key condition above holds, we can implement a **Hausman test** to assess whether the unobserved individual effects are correlated with the explanatory variables or not. The null hypothesis is $H_0 : C(X_{it}, a_i) = 0$.
- The Hausman test evaluates such hypothesis testing whether the difference between the WG/FE and the RE estimators is statistically significant or not.
- This test is straightforward to implement in `gretl` (we are not going to derive it).

Testing "Fixed Effects" vs "Random Effects" (cont.)

- Notice, however, that the validity of the Hausman test lies on very strong assumptions to guarantee that the WG/FE estimator is consistent anyway.
  - Namely, we need the **key condition** $C\left(X_{it}, u_{is}\right) = 0 \forall t, s = 1, \ldots, T$.
  - If such condition does not hold, the WG/FE estimator will be inconsistent both under H_0 and H_1 (and also the RE estimator).
- Even if the key condition is held, we still need the idiosynchratic error to satisfy conditional homoskedasticity and lack of serial correlation.
  - Otherwise the Hausman test would still be inconsistent.
- Consequently, the strategy based on the Hausman test confronting the FE and the RE estimators is not recommended.
- This strategy is currently overcome among the practitioners.

Example FE vs RE: GMR

- Consider the following model for innovation expenditure (`inn`) with firm-level panel data

$$\texttt{inn}_{it} = \beta_{1j}\texttt{GMR}_{it}\,(j) + \beta_{2k}\texttt{Dsize}_{it}\,(k) + \beta_{3l}\texttt{Dsector}_{it}\,(l) + a_i + u_{it}$$

  where $\texttt{GMR}\,(j)$ is the range of the market where the company operates, with j between 1 and 4 corresponding to local (1), national (2), European (3) or International (4) and $\texttt{Dsize}\,(k)$ and $\texttt{Dsector}(l)$ correspond to company size and sector or industry in which the company operates.
- If `GMR` affects innovation positively, β_{1j} would be higher for higher j.
- We can use the Spanish "Panel de Innovación Tecnológica" (PITEC) or Panel of Technological Innovation, an unbalanced panel with detailed company-level information on innovation from 2003.
  - Our sample runs between 2003 and 2008.

Example FE vs RE: GMR

Pooled OLS

```
ols inn const dummify(gmr) dummify(year) dummify(sector)
----robust
```

```
Model 1: Pooled
Included 5004 cross-sectional units
Time-series length: minimum 1, maximum 5
Dependent variable: inn
Robust (H
```

	coefficient	std. error	t-ratio	p-value	
const	0.504561	0.0472255	10.68	1.40e-26	***
Dgmr_2	0.0928605	0.0174111	5.333	9.73e-08	***
Dgmr_3	0.190534	0.0191665	9.941	3.08e-23	***
Dgmr_4	0.275561	0.0179527	15.35	6.64e-53	***
Dyear_2	-0.0347065	0.00787474	-4.407	1.05e-05	***
Dyear_3	0.0141995	0.00744819	1.906	0.0566	*
Dyear_4	0.0175798	0.00649652	2.706	0.0068	***
Dyear_5	-0.00381661	0.00481397	-0.7928	0.4279	
Dsector_2	0.0528567	0.0528875	0.9994	0.3176	
Dsector_3	0.105232	0.0467047	2.253	0.0243	**
Dsector_4	0.0568898	0.0481531	1.181	0.2374	
Dsector_5	0.0699164	0.0468809	1.491	0.1359	
Dsector_6	0.125310	0.0470796	2.662	0.0078	***
Dsector_7	0.0890264	0.0469188	1.897	0.0578	*
Dsector_8	0.0244243	0.0536698	0.4551	0.6491	

Example FE vs RE: GMR

Fixed Effects FE/WG

```
panel inn const dummify(gmr) dummify(year) dummify(sector)
----robust
```

```
# Fixed-effects
? panel inn const dummify(gmr) dummify(year) dummify(sector) --robust

Model 2: Fixed-effects, using 22548 observations
Included 5004 cross-sectional units
Time-series length: minimum 1, maximum 5
Dependent variable: inn
Robust (H


              coefficient   std. error   t-ratio   p-value
  ---------------------------------------------------------
  const         0.411051     0.110877     3.707    0.0002   ***
  Dgmr_2        0.0189910    0.0163840    1.159    0.2464
  Dgmr_3        0.0565919    0.0188047    3.009    0.0026   ***
  Dgmr_4        0.0848563    0.0194780    4.357    1.33e-05 ***
  Dyear_2      -0.00946958   0.00767044  -1.235    0.2170
  Dyear_3       0.0220273    0.00718667   3.065    0.0022   ***
  Dyear_4       0.0223016    0.00629260   3.544    0.0004   ***
  Dyear_5      -0.00125518   0.00466273  -0.2692   0.7878
  Dsector_2     0.140703     0.130886     1.075    0.2824
  Dsector_3     0.213858     0.117661     1.818    0.0691   *
  Dsector_4     0.246948     0.121435     2.034    0.0420   **
  Dsector_5     0.136018     0.116954     1.163    0.2448
  Dsector_6     0.138575     0.120310     1.152    0.2494
  Dsector_7     0.201261     0.118028     1.705    0.0882   *
  Dsector_8     0.208629     0.134680     1.549    0.1214
  Dsector_9     0.291881     0.148914     1.960    0.0500   *
```

Example FE vs RE: GMR

"Random Effects" (RE)

```
panel inn const dummify(gmr) dummify(year) dummify(sector)
----random-effects
```

```
# Random-effects
? panel inn const dummify(gmr) dummify(year) dummify(sector) --random-effects

Model 4: Random-effects (GLS), using 22548 observations
Included 5004 cross-sectional units
Time-series length: minimum 1, maximum 5
Dependent variable: inn


              coefficient   std. error   t-ratio    p-value
  ------------------------------------------------------------
  const         0.529804    0.0402741    13.15      2.23e-39 ***
  Dgmr_2        0.0625103   0.0108095     5.783     7.44e-09 ***
  Dgmr_3        0.137182    0.0125538    10.93      1.00e-27 ***
  Dgmr_4        0.192742    0.0122588    15.72      2.08e-55 ***
  Dyear_2      -0.0142156   0.00652366   -2.179     0.0293   **
  Dyear_3       0.0192957   0.00619107    3.117     0.0018   ***
  Dyear_4       0.0214864   0.00612074    3.510     0.0004   ***
  Dyear_5      -0.00224029  0.00616410   -0.3634    0.7163
  Dsector_2     0.0563393   0.0483025     1.166     0.2435
  Dsector_3     0.122983    0.0430382     2.858     0.0043   ***
```

Example FE vs RE: GMR

The parameter λ and the Hausman test

- 'Within-variance' = 0.0837569 'Between-variance' = 0.130407
- Hausman test - Null hypothesis: GLS estimates are consistent
  - Asymptotic test statistic: Chi-square(29) = 374.727 with p-value = 9.52973e-62
- Notice the Hausman test is valid only if the WG estimator is consistent, what requires the strong **key condition** $C\left(X_{it}, u_{is}\right) = 0 \forall t, s = 1, \ldots, T$.

Advantages of panel data

- Unlike single cross sections or repeated cross sections, panel data allows to control for individual unobserved heterogeneity (unobserved individual effects).
- Individual effects a_i $(i = 1, \ldots, n)$ are unobserved individual-specific, time-invariant, random variables that capture unobserved differences among individual units that do not change over time (at least, along the sample time period).
- Such individual effects are potentially correlated with the explanatory variables, what make pooled OLS estimators inconsistent.
- In order to get consistent estimators, we must consider transformations of the original model that remove the (time-invariant) individual effects, like FD or WG.
- If **strict exogeneity** $C(X_{it}, u_{is}) = 0 \forall t, s$ holds, OLS estimation of the FD or WG transformation of the model of interest yields consistent estimates of the parameters of interest.

Panel data in practice

- Unobserved individual effects are random variables that are usually correlated with the explanatory variables.
- This advises against the old-fashion strategy of using the Hausman test to choose between FE/WG vs RE estimators.
  - The RE estimator relies on very unrealistic assumptions.
  - If such assumptions do not hold, the Hausman test is ill-defined, so it can lead to misleading conclusions.
- Also, as it happens with cross sections, we should expect heteroskedasticity, so heteroskedasticity-robust standard errors are recommended.
- We should also allow for common aggregate shocks by introducing time dummies.
- In general, some individual units will stop answering iat some time period and leave the sample. Hence, for the sake of sample representativeness, the sample must be refreshed with new individual units. Consequently, not all individuals are observed in each sample period. Also, panel data can be unbalanced, with different individual units being observed in a different time span (T_i).

Limitations of panel data

- Very often, the **key condition** $C(X_{it}, u_{is}) = 0 \forall t, s = 1, \ldots, T$ (**strict exogeneity**) does not hold.
- If some explanatory variable is not strictly exogenous, then the FD and WG estimators will be inconsistent.
  - Obviously, this condition is not held with dynamic models (not studied here), which include the lagged endogenous variable among the explanatory variables.
- Notice that the WG estimator requires much more stringent conditions than the FD estimator, making the WG estimator less robust than the FD estimator.
- If the key condition doesn't hold, we should consider instrumental-variable methods for the FD transformation of the model (not studied here).