2: Discrete Choice

GECO 6281 Advanced Econometrics 1 (Lab)

Patrick Mokre

Fall 2019

Limited Dependent Variables

- **What exactly does an OLS estimation coefficient capture?**
- **Under which technical conditions is that estimation BLUE?**
- **Under which conditions does that kind of modeling make sense intuitively/in a modeling context?**
- **Bonus: What is the difference between consistency and unbiasedness**?

Limited Dependent Variables 2

$\hat{\beta}_{OLS}$ is an approximation to $\frac{\partial y}{\partial X}$. Inituitively, this makes the most sense with a **continuous** dependent variable and covariates.

$\hat{\beta}_{OLS}$ is **consistent** and **efficient** und the Gauss-Markov conditions.

- $E\epsilon_i = 0 \quad \forall i \in N$
- $\epsilon_1, ..., \epsilon_n$ and $x_1, ..., x_n$ are *independent*
- $Var(\epsilon_i) = \sigma^2 \quad \forall i \in N$
- $cov(\epsilon_i, \epsilon_j) = 0 \quad \forall i, j \in N, \forall j \neq i$

Limited Dependent Variables 3

Often microeconomic data is presented in discrete or discrete mixed continuous form.

Problem 1: If one estimates binary data using OLS, $x'\beta$ must be read as a probability, which by definition can only be between 0 and 1. This is only possible if either x or β are artificially restricted.

Problem 2: Usually the error term is **not normally distributed** and suffers from **heteroskedasticity**:

$$\begin{aligned} P(y_i = 1 \mid x_i) &= x_i'\beta \\ P(\epsilon = -x_i'\beta \mid x_i) &= P(y_i = 0 \mid x_i) = 1 - x_i'\beta \\ P(\epsilon = 1 - x_i'\beta \mid x_i) &= x_i'\beta \\ \Rightarrow V(\epsilon \mid x_i) &= x_i'\beta(1 - x_i'\beta) \neq V(\epsilon) \end{aligned}$$

Clearly, a bipolar distribution is not Gaussian Normal, and the variance depends on the value of the covariates.

Binary Choice Modeling

$$P(y_i = 1 \mid x_i) = G(x_i, \beta)$$

If you choose for the function $G(x_i, \beta)$ the Gaussian Normal distribution $\Phi(x_i'\beta)$, this is called a **Probit** model:

$$\frac{\partial \Phi(x_i'\beta)}{\partial x_{ik}} = \phi(x_i'\beta)\beta_k$$

The logistical distribution $\frac{exp(x_i'\beta)}{1+exp(x_i'\beta)}$ gives a **Logit model**.

$$\frac{\partial L(x_i'\beta)}{\partial x_{ik}} = \frac{exp(x_i'\beta)}{(1+exp(x_i'\beta))^2}\beta_k$$

Latent Variable Models

One can also model a bivariate outcome as the result of a censoring process. For this, one makes behavioural assumptions on why a variable never materializes.

Let y_i^* be an underlying (latent) variable. As an example, think of a reservation wage: If a person is offered less than \$ 1500, they may not enter employment but choose to do domestic labor instead.

$$
\begin{aligned}
y_i^* = x_i'\beta + \epsilon, \quad \epsilon \sim N(0, \sigma^2) \\
y_i = 1 \quad if \quad y_i^* > 0 \\
y_i = 0 \quad if \quad y_i^* < 0
\end{aligned}
$$

Latent Variable Models

The model can be estimated using a simple likelihood formulation.

$$L(\beta) = \prod_{i}^{N} P(y_i = 1 \mid x_i; \beta)^{y_i} P(y_i = 0 \mid x_i; \beta)^{1-y_i}$$

Since the natural logarithm is a monotonous function, the value β that maximizes the likelihood also maximizes the log-likelihood $LL(\beta)$. Since Log-Likelihoods can be summed up rather than multiplied the procedure becomes **computationally more efficient** and does less often run into problems with **floating digits**.

$$LL(\beta) = \sum_{i}^{N} y_i log(P(y_i = 1 \mid x_i; \beta)) + (1 - y_i) log(P(y_i = 0 \mid x_i; \beta))$$

Both Logit and Probit models can be estimated using Maximum (Log-) Likelihood routines.

Goodness of Fit

Goodness of Fit in probabilistic models mostly measure either precision in **calculated probabilities compared to observed frequencies** or **prediction of observed data**.

Often GOF statistics implicitly compare the model with one that includes only a constant by comparing the calculated likelihoods. Let

Amemiya Pseudo-R^2:

$$1 - \frac{1}{1 + 2(logL_1 - logL_0)/N}$$

McFadden statistic:

$$1 - \frac{logL_1}{logL_0}$$

Restricted Dependent Variables: TOBIT

When dependent variables are continuous, but constrained, more problems arise. Examples are when a variable is zero for a large part of the population and positive for the rest (eg. expenditures, income from a certain type of activity or asset, work hours).

Tobit models are well-suited for such latent variable problems. It applies conditional probabilities ot the problem, usually introducing a Gaussian Normal density function.

$$P(y_i = 0) = P(y_i^* < 0) = P(\epsilon_i < -x_i'\beta) = 1 - \Phi(\frac{x_i'\beta}{\sigma})$$

$$E(y_i \mid y_i > 0) = x_i'\beta + E(\epsilon_i \mid \epsilon_i > -x_i'\beta) = x_i'\beta + \sigma \frac{\phi(x_i'\beta/\sigma)}{\Phi(x_i'\beta/\sigma)}$$

Restricted Dependent Variables: TOBIT 2

The parameters obtained in a Maximum-Likelihood procedure can be interpreted in two ways. Note that the ML procedure has to simultaneously estimate β and σ.

Marginal impact on the probability to observe a zero value in the dependent variable:

$$\frac{\partial P(y_i = 0)}{\partial x_{ik}} = -\phi(\frac{x_i'\beta}{\sigma})\frac{\beta_k}{\sigma}$$

Marginal impact on th expected value of the dependent variable, conditional on a positive realization:

$$\begin{aligned} E(y_i) &= x_i'\beta\Phi(x_i'\beta/\sigma) + \sigma\phi(x_i'\beta/\sigma) \\ \frac{\partial E(y_i)}{\partial x_{ik}} &= \beta_k\Phi(x_i'\beta/\sigma) \\ \frac{\partial E(y_i^*)}{\partial x_{ik}} &= \beta_k \end{aligned}$$

Specification/Estimation Issues

Violations of the distributional assumptions on ϵ_i (e.g. non-normality and heteroskedasticity) will lead to inconsistent parameter estimations.

Pagan and Vella (1989) propose a moment-based test for **normality**, as for normally distributed errors it should hold that $E(\epsilon^3/\sigma^3 \mid x_i) = 0$ and $E(\epsilon^4/\sigma^4 - 3 \mid x_i) = 0$ (absence of skewness and kurtosis.

Selection Bias: Tobit 2-Model/Heckman 2-Step Selection Model

One can argue that underlying the restriction of a continuous variable y (say: wages) lies a binary outcome h (say: to seek employment or not).

$$
\begin{aligned}
y_i^* &= x_{1i}'\beta_1 + \epsilon_1 \\
h_i^* &= x_{2i}'\beta_2 + \epsilon_2 \\
y_i = y*_i, h_i = 1 \quad if \quad & h_i^* > 0 \\
y_i = 0, h_i = 0 \quad if \quad & h_i^* \leq 0
\end{aligned}
$$

Selection Bias: Tobit 2-Model/Heckman 2-Step Selection Model 2

Under the assumption that $\epsilon_2 \sim N(0,1) \Rightarrow \sigma_2^2 = 1$:

$$E(w_i \mid h_i = 1) = x'_{1i}\beta_1 + \sigma_{12}\frac{\phi(x'_{2i}\beta_2)}{\Phi(x'_{2i}\beta_2)}$$

$$\sigma_{12} = \rho_{12}\sigma_1$$

$$\rho_{12} = Corr(\epsilon_1, \epsilon_2)$$

Selection Bias: Tobit 2-Model/Heckman 2-Step Selection Model 3

The model can be denoted as a maximum likelihood estimation.

$$\begin{aligned} logL_3(\beta, \sigma_1^2, \sigma_{12}) &= \sum_{i \in I_0} logP(h_1 = 0) + \sum_{i \in I_1}[logf(y_i \mid h_1 = 1) + logP(h_i = 1)] \\ &= \sum_{i \in I_0} logP(h_1 = 0) + \sum_{i \in I_1}[logf(y_i) + logP(h_i = 1 \mid y_i)] \end{aligned}$$

Selection Bias: Tobit 2-Model/Heckman 2-Step Selection Model 4

Heckman provides a two step estimation technique which is often applied in research.

$$y_i = x'_{1i}\beta_1 + \sigma_{12}\lambda_i + \eta_i$$
$$\lambda_i = \frac{\phi(x'_{2i}\beta_2)}{\Phi(x'_{2i}\beta_2)}$$

The only unknown in λ_i is β_2, which can be estimated in a Tobit routine to be then plugged into a linear regression for the upper equation.

Panel Data

Panel Data combines aspects of time series and cross-sectional econometric analysis.

We will have to deal with multi-dimensional group-specific effects.

Popular examples:

- Longitudinal Surveys
- Cross-Country Macro analysis
- Experiments rolled out in multiple waves (Why do researchers do that?)

Panel Data Econometrics is among the most popular methods is economic research.

Panel Data: Big Questions

- Does the time component matter? Why/Why Not?
- Which groups of observations come to mind? Do groups matter? Why/Why Not?
- How does the panel structure of data change economic modeling questions? What additional knowledge is there to find? Which additional, non-statistical difficulties arise?
- What can count as an observation?

Panel Data: Pro and Con

+ Data allows for **more complicated** and **more realistic** economic models. Example: What is the insight won from observing (a) the average rate of profit in one year, (b) the average rate of profit over 20 years and (c) the industrial average rate of profit over 20 years?

+ Estimate changes on an individual (observation) level

- Independence of observations no longer holds

- Missing observations

Panel Data: Representation

You have observations (y_{it}, x_{it}) for individuals $i \in I$ and periods $t \in T$. Estimate the impact of x on y.

The most general formulation of a model is:

$$y_{it} = \alpha_{it} + x_{it}'\beta_{it} + \epsilon_{it}$$

What is the insurmountable weakness of this model? How is it located between **descriptive** and **inference statistics**?

Panel Data: OLS estimation

Estimate the impact of x on y by simple OLS:

$$y_{it} = \alpha + x'_{it}\beta + \epsilon_{it}$$

Note: $\hat{\beta}_{OLS}$ is the best linear unbiased estimator (BLUE) only if the Gauss-Markov properties are fulfilled. With regard to **independent observations** and **homoskedasticity**, this is problematic.

Panel Data: Clustered Errors

You cannot assume that ϵ_{it} is i.i.d., and specifically that $\epsilon_{it} \sim N(0, \sigma)$.

Introduce clustered errors:

$$y_{it} = \alpha + x_{it}'\beta + \epsilon_{it}$$
$$\epsilon_{it} \sim N(0, \sigma_i)$$

Panel Data: OLS versus cluster-robust standard errors

OLS:

```
reg lwage ed exp ind

      Source |       SS           df       MS      Number of obs   =     4,165
-------------+----------------------------------   F(3, 4161)      =    484.01
       Model |  229.434018         3  76.478006   Prob > F        =    0.0000
    Residual |  657.470884     4,161  .158007903   R-squared       =    0.2587
-------------+----------------------------------   Adj R-squared   =    0.2582
       Total |  886.904902     4,164  .212993492   Root MSE        =     .3975

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0803785   .0023154    34.72   0.000     .0758391    .0849179
         exp |     .01251   .0005793    21.59   0.000     .0113743    .0136458
         ind |   .1070021   .0130489     8.20   0.000     .0814193    .1325849
       _cons |   5.353169   .0355112   150.75   0.000     5.283549     5.42279
------------------------------------------------------------------------------
```

Panel Data: OLS versus cluster-robust standard errors 2

Cluster-Robust Errors:

```
 reg lwage ed exp ind, vce(cluster id)

Linear regression                               Number of obs     =      4,165
                                                F(3, 594)         =      83.29
                                                Prob > F          =     0.0000
                                                R-squared         =     0.2587
                                                Root MSE          =      .3975

                                    (Std. Err. adjusted for 595 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          ed |   .0803785   .0053806    14.94   0.000     .0698112    .0909458
         exp |     .01251   .0014803     8.45   0.000     .0096029    .0154172
         ind |   .1070021   .0272083     3.93   0.000     .0535659    .1604383
       _cons |   5.353169   .0856763    62.48   0.000     5.184904    5.521435
------------------------------------------------------------------------------
```

Spot a difference?

Clustered Standard Errors: What Happened?

In a standard OLS regression, you assume that the error terms ϵ_i are independent and Gaussian Normal distributed $\sim N(0, \sigma^2)$

In a panel setup, you cannot just assume that: Characteristics of individuals are not independent over time (your wage in 2018 cannot be modeled as the outcome of an experiment independent of your wage in 2017).

Furthermore, observations might be **clustered**: individuals might live in the same city, work a similar job, and so on. You cannot just assume that the impact of x_i on y_i is independent although you know of this clustering, even if you cannot observe the clusters directly.

In a panel setup, you try to "catch" these unobserved effects using a **fixed effects** indicator (more on that later). In the above case, the indicator is ther personal identification number `id`.Software packages can automatically search for clusters. We then re-calculate **coefficient standard errors** taking into account that $\epsilon_{it}\ N(0, \sigma_c^2)$, where $c \in C$ denotes the cluster.

Note: Calculating cluster robust standard errors allows you to not specify a model of **how** clusters affect the outcome. However, you need to assume that **the number of clusters approaches infinity** (Ibragimov and Müller, 2016)

Clustered Standard Errors 2: How did it happen?

In OLS, the degrees-of-freedom corrected estimator $s^2 = \frac{1}{N-K}\sum_i (e_i)^2$ with e_i the forecast residuals can be used to efficiently estimate a **coefficient standard error**. (Verbeek 2004, 18f)

$$\sqrt{\tilde{V}(b_k)} = \sqrt{(s^2(\sum_i x_{i,k}^2)^{-1}}$$

In a clustered design, you choose clusters or let a software choose it by some efficiency properties for you.

$$\hat{V}(\hat{b}_k) = [X'X]^{-1}[\sum_c^C x_c' \hat{\epsilon}_c' \hat{\epsilon}_c x_c][X'X]^{-1}$$

You retrieve the coefficient standard error by taking the square root of the variance.

Panel Data: Representations

Enough about errors, more about predicitions.

Models: Generalizations of an assumed structure of the data. Start at the beginning. Note that u_{it} is the observed residual, and not necessarily the model error term.

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it}$$

Unit-Specific Representation (in stacked form, i.e. T equations)

$$\begin{array}{ccccc} y_i & = & \alpha_i & \tau_t + X_i & \beta + u_i \\ (T \times 1) & = & (1 \times 1) & (T \times 1) + (T \times k) & (k \times 1) + (T \times 1) \end{array}$$

Time-Specific Representation (N equations)

$$\begin{array}{ccccc} y_t & = & \alpha & X_t \beta + & u_t \\ (N \times 1) & = & (N \times 1) & (N \times k) + (k \times 1) + & (N \times 1) \end{array}$$

Panel Data Estimation: Pooled OLS

Under assumption of **homogenous intercept** $\alpha_i = \alpha \quad \forall i \in N$ and strictly exogenous covariates x_i, the panel can be estimated using **ordinary least squares** OLS.

STATA:

```
reg lwage ed exp ind, vce(cluster id)
```

Fixed Effect estimation

In a fixed effects estimation, you allow for individual effects, formalized in **heterogenous intercepts** α_i.

$$y_{it} = \alpha_i + \beta' x_{it} + u_{it}$$

Stochastically, we can say that α_i are drawn from a **joint distribution of** α_i, x_{it}, u_{it} with the parameters of the distribution allowed to increase with the same speed as the number of cross-sectional observations.

Increasing the number of regression coefficients α_i, β by N strongly decreases the degrees of freedom.

Methodologically estimating a fixed effects model amounts to **eliminating the fixed effects** from the regression (e.g. by using first difference $x_{it} - x_{it-1}$ or mean difference $x_{it} - \bar{x_i}$ as covariates), then calculate them from the estimated coefficients.

Degrees of Freedom

The degrees of freedom (DF) indicate the number of independent values that can vary in an analysis without breaking any constraints. it increases in independent information you can use for parameter estimation, and decreases in parameters you **have to** estimate due to your modeling choices.

In **frequentist** statistics, hypothesis testing is based in the assumption that **coefficient estimates** (such as $\hat{\beta}$) foolow some distribution, where the shape is co-determined by the degrees of freedom (Student T, χ^2, ...).

For low degrees of freedom, these distributions become very narrow, making hypothesis testing difficult. Coefficietn estimates become unreliable, and the hypothesis tests lose testing power.

Fixed Effects: Intuition

You want to estimate β after eliminating individual effects α_i.

One approach is to calculate averages over time:

$$\bar{y_i} = \alpha_i + \beta' \bar{x}_i + \bar{u}_i$$

Then, for each observation $i \in N$:

$$y_{it} - \bar{y_i} = \beta'(x_{it} - \bar{x}_i) + (u_{it} - \bar{u}_i)$$

This is called the **within transformation** of a fixed effect model, and can be efficiently estimated by **pooled OLS**.

Note that x_{it} needs to be **time-varying** for the within estimator to be meaningful. Furthermore note that expected values $E(y_{it} \mid x_{it}) = E(a_i \mid x_{it}) + \beta' x_{it}$ cannot be estimated, as we have no estimate way for estimating the intercept in **short panels**. (Cameron and Trivedi 2009, 231)

FE Estimation in STATA 1

For panel data estimation, STATA has special commands like `xtreg`, `xtline`, and so on. Here, the `x` denotes the cross-sectional and `t` the time dimension.

Load data:

```
use mus08psidextract.dta, clear
```

Set the panel indicators using `xtset`.

```
xtset t id

       panel variable:  t (strongly balanced)
        time variable:  id, 1 to 595
                delta:  1 unit
```

Perform a within regression including fixed effects using the `xtreg` command including the `, fe` specification.

FE Estimation in STATA 2

```
. xtreg lwage ed exp ind, fe vce(robust)

Fixed-effects (within) regression               Number of obs     =      4,165
Group variable: t                               Number of groups  =          7

R-sq:                            . xtreg lwage exp ind, fe

Fixed-effects (within) regression               Number of obs     =      4,165
Group variable: id                              Number of groups  =        595

R-sq:                                           Obs per group:
     within  = 0.6507                                         min =          7
     between = 0.0251                                         avg =        7.0
     overall = 0.0439                                         max =          7

                                                F(2,3568)         =    3322.89
corr(u_i, Xb)  = -0.9145                        Prob > F          =     0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .0969207   .0011893    81.50   0.000      .094589    .0992524
         ind |    .022139   .0155742     1.42   0.155    -.0083963    .0526743
       _cons |   4.743349   .0244748   193.81   0.000     4.695363    4.791335
-------------+----------------------------------------------------------------
     sigma_u |  1.0592693
```

Fixed Effects: Retrieve the intercept

$\hat{\alpha}_i = \bar{y_i} - \hat{\beta}'_{FE}\bar{x_i}$

Note: In short and wide panels (small N, large T) the intercept cannot be efficiently retrieved.

Fixed Effects: Stacked LSDV Estimation

The Within-Estimator is equivalent to a **stacked** estimation with N dummy variables α_i. This procedure is called the **least-squares dummy variable (LSDV)** estimator. It cannot estimate α_i consistently in short panels, but consistently estimates β. (Cameron and Trivedi 2009, 253)

In STATA, this can be estimated using the `areg` command and specifying the fixed effects dimension in the `absorb` specification.

Fixed Effects: Stacked LSDV Estimation in STATA

```
areg lwage exp ind, absorb(id) vce(cluster id)

Linear regression, absorbing indicators         Number of obs     =      4,165
Absorbed variable: id                           No. of categories =        595
                                                F(   2,    594)   =    1282.85
                                                Prob > F          =     0.0000
                                                R-squared         =     0.9052
                                                Adj R-squared     =     0.8894
                                                Root MSE          =     0.1535

                                 (Std. Err. adjusted for 595 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lwage |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .0969207    .001914    50.64   0.000     .0931616    .1006798
         ind |    .022139   .0245714     0.90   0.368    -.0261185    .0703965
       _cons |   4.743349    .039158   121.13   0.000     4.666444    4.820254
------------------------------------------------------------------------------
```

Random Effects: Properties

The random effects estimator assumes that the individual effects α_i are drawn from a joint probabilistic distribution. Often, this is modeled as part of the error term, which in turn allows for introducing a general intercept term, a primitive form of **hierarchical modeling**.

You will not be able to efficiently estimate an RE model using OLS, and will need to specifiy a **GLS model** estimation method

$$\begin{aligned} y_{it} &= \alpha + x'_{it}\beta + u_{it} \\ u_{it} &= \alpha_i + \epsilon_{it} \\ E(\epsilon_{it} \mid x_{it}) = 0 &\Rightarrow E(u_{it} \mid \alpha_i, x_{it}) = 0 \end{aligned}$$

This implies a number of important properties for u_{it}.

$$\begin{aligned} E(u_{it}^2) &= \sigma_\alpha^2 + \sigma^2 + 2Cov(\alpha_i, u_{it}) = \sigma_\alpha^2 + \sigma^2 \\ E(u_{it}u_{is}) &= E[(\alpha_i + u_{it})(\alpha_i + u_{is})] = \sigma_\alpha^2 \end{aligned}$$

Random Effects: Efficiency

A GLS estimation of an RE model is consistent for β with N or T going to infinity if you assume **exogeneity of covariates**, normal distribution of error terms, and a non-singular asymptotical variance-covariance matrix.

Note that for ML estimation in the FGLS, you need to assume that both α_i and ϵ_{it} are i.i.d.

Random Effects: Estimation in STATA

```
xtreg lwage exp ind, re

Random-effects GLS regression                   Number of obs     =      4,165
Group variable: id                              Number of groups  =        595

R-sq:                                           Obs per group:
     within  = 0.6500                                         min =          7
     between = 0.0249                                         avg =        7.0
     overall = 0.0438                                         max =          7

                                                Wald chi2(2)      =    2807.13
corr(u_i, X)   = 0 (assumed)                    Prob > chi2       =     0.0000

------------------------------------------------------------------------------
       lwage |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         exp |   .0612741   .0011572    52.95   0.000      .059006    .0635423
         ind |  -.0123885   .0177007    -0.70   0.484    -.0470812    .0223041
       _cons |   5.464722   .0309463   176.59   0.000     5.404068    5.525375
-------------+----------------------------------------------------------------
     sigma_u |  .38545785
     sigma_e |  .15349733
         rho |  .86312569   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

Relationship between RE and pooled OLS

Including a general intercept term α_i and keeping $u_{it} = \alpha_{it} + \epsilon_{it}$, the RE model has $E(u_{it} \mid X) = 0$. Under the additional assumptions of deterministic and bounded covbariates as well asa asymptotically positive definite variance-covariance matrix (Pesaran 2015, 636), cross-sectional independence of the errors and allowing for serial correlation between errors in the time dimension (all included in RE), **pooled OLS is consistent for RE**.

However, under the RE specifications that ϵ_{it} is serially uncorrelated and homoskedastic, **pooled OLS is inefficient**.

If the last assumption is unlikely to hold, **pooled OLS may be preferrable to FGLS estimation**.

Relationship between RE and FE

The relationship between the RE and FE setup is determined by the **heterogeneity in α_i and** σ_α^2. For maximum heterogeneity, RE converges to FE, for minimum heterogeneity, RE converges to the pooled OLS estimator.

Furthermore, for $T \rightarrow \infty$, RE and FE estimators converge.

Deciding on a model

There are different approaches to choosing between FE and RE setups. These are some:

1 Theoretical determinationn (Pesaran 2015): If we are interested in between-individual heterogeneity, FE makes sense. If N is large and you consider it a random sample from the population, RE is more appropriate. More technically, the decision variable is your **beliefs about the correlation between individual effects and covariates** x_{it}.

2 Hausman Test: The HT tests under the null that effects are random and compares the FE and RE estimators. Under the Null, the estimators converge. in STATA, you need to run both models, store the estimates, and use the `hausman` command.

3 Gelman's Rejection of fixed effects: Andrew Gelman, an important researcher into Bayesian multilevel modeling argues, that the notion of "fixed effects models" makes little effect in and of itself, and one should rather assume all downstream hierarchical coefficients are the product of **some** random distribution. However, this is easier said in Bayesian statistics, as it allows for distributions other than the Gaussian Normal.

Hausman Test: STATA

```
. quietly xtreg lwage exp ind, fe

. estimates store FE

. quietly xtreg lwage exp ind, re

. estimates store RE

. hausman FE RE

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       FE           RE         Difference          S.E.
-------------+----------------------------------------------------------------
         exp |    .0969207     .0612741        .0356466        .0002741
         ind |     .022139    -.0123885        .0345275               .
------------------------------------------------------------------------------
                           b = consistent under Ho and Ha; obtained from xtreg
            B = inconsistent under Ha, efficient under Ho; obtained from xtreg

    Test:  Ho:  difference in coefficients not systematic

                  chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =    15144.30
                Prob>chi2 =      0.0000
                (V_b-V_B is not positive definite)
```