Panel data. Between and within variation. Random and fixed effects models. Between regression. Hausman-Taylor estimator

Jakub Mućk
SGH Warsaw School of Economics

Introduction

Panel of data consists of a group of cross-section units (people, firms, states, countries) that are observed over the time:

- Cross-section:
  y_i where
  $i \in \{1, \ldots, N\}$.
- Time series:
  y_t where
  $t \in \{1, \ldots, T\}$.
- Panel data:
  y_{it} where
  $i \in \{1, \ldots, N\}$
  $t \in \{1, \ldots, T\}$.

  In general,
- N - the cross-sectional dimension.
- T - the time dimension.

We might describe panel data using T and N:

- **long/short** describes the time dimension (T);
- **wide/narrow** describes the cross-section dimension (N);

For example: panel with relatively large N and T: long and wide panel.

- In a **balanced** panel, each individuals(unit) has the same number of observation.
- **Unbalanced** panel is a panel in which the number of time series observations is different across units.

Advantages of Panel Data (Baltagi, 2014)

- Controlling for **individual heterogeneity**.
- Panel data offer more informative data, more variability, less collinearity among the dependent variables, more degrees of freedom and more efficiency in estimation.
- Identification and measurement of effects that are simply not detectable in pure cross-section or pure time-series data.
- Testing more complicated behavioral models than purely cross-section or time-series data.
- Reduction in biases resulting from aggregation over firms or individuals.
- Overcome the problem of nonstandard distributions typical of unit roots tests $\Longrightarrow$ macro panels.

Limitations of Panel Data

SGH

- **Design and data collection problems:**
  - coverage;
  - nonresponse;
  - frequency of interviewing;
- **Distortions of measurement errors**
- **Selectivity problems:**
  - self-selectivity;
  - nonresponse;
  - attrition;
- **Short T.**
- **Cross-sectional dependence.**

- **Classical example**
  Agricultural Cobb-Douglas production function. Consider the following model:

$$y_{it} = \beta x_{it} + u_{it} + \eta_i \tag{1}$$

  - y_{it} – the log output.
  - x_{it} – the log of a variable input;
  - η_i – an farm-specific input that is constant over time, e.g., soil quality.
  - u_{it} – a stochastic input that is outside framer's control, e.g., rainfalls.
  - β - the technological parameter.

- **Classical example**
  Agricultural Cobb-Douglas production function. Consider the following model:

$$y_{it} = \beta x_{it} + u_{it} + \eta_i \tag{1}$$

  - y_{it} – the log output.
  - x_{it} – the log of a variable input;
  - η_i – an farm-specific input that is constant over time, e.g., soil quality.
  - u_{it} – a stochastic input that is outside framer's control, e.g., rainfalls.
  - β - the technological parameter.

- **An example in which panel data does not work**
  Returns to education. Consider the following model:

$$y_{it} = \alpha + \beta x_{it} + u_{it} \tag{2}$$

  - y_{it} – the log wage;
  - x_{it} – years of the full-time education;
  - β – returns to education.

  In addition:

$$u_{it} = \eta_i + \varepsilon_{it} \tag{3}$$

  where η_i stands for the **unobserved individual** *ability*.
  Problem: x_{it} lacks of time variation.

Pooled OLS estimator

- **Pooled model** is one where the data on different units are pooled together with **no assumption on individual differences**:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it} \quad (4)$$

  where
  - y_{it} – the dependent variable;
  - x_{kit} – the $k - th$ explanatory variable;
  - u_{it} – the error/disturbance term;
  - β_0 – the intercept;
  - $\beta_1, \ldots, \beta_k$ – the structural parameters;
- Note that the coefficients $\beta_0, \beta_1, \ldots, \beta_k$ are the same for all unit (do not have i or t subscript).

- **Pooled model** is one where the data on different units are pooled together with **no assumption on individual differences**:

$$y_{it} = \beta_0 + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it} \tag{4}$$

where

  - y_{it} – the dependent variable;
  - x_{kit} – the $k-th$ explanatory variable;
  - u_{it} – the error/disturbance term;
  - β_0 – the intercept;
  - β_1, ..., β_k – the structural parameters;

- Note that the coefficients β_0, β_1, ..., β_k are the same for all unit (do not have i or t subscript).

Assumptions (for linear pooled model):

$$\mathbb{E}(u) = 0 \quad (5)$$
$$\mathbb{E}(uu') = \sigma_u^2 I \quad (6)$$
$$rank(X) = K + 1 < NT \quad (7)$$
$$\mathbb{E}(u|X) = 0 \quad (8)$$

- (8): X is nonstochastic and is not correlated with u.
- (6): the error term (u) is not autocorrelated and homoscedastic.
- (8) $\Longrightarrow$ strictly exogeneity of independent variables.

Gauss-Markov Theorem

If (5)-(8)and are satisfied then $\hat{\beta}^{POOLED}$ is BLUE (the best linear unbiased estimator).

- The general assumption in pooled regression on the error terms are very strong or even unrealistic.
- **The lack of correlation between errors corresponding to the same individuals**.
- Let us relax the above assumption:

$$\text{cov}(u_{i,t}, u_{i,s}) \neq 0 \tag{9}$$

- Then we have problem of both autocorrelation and heteroskedasticity.
- The OLS estimator is still consistent but the standard errors are incorrect.
- We might use the **clustered/robust standard errors**. Here, the time series for each individual are clusters.

Fixed effects model

- In the model:

$$y_{it} = \alpha + X_{it}'\beta + u_{it} \quad i \in \{1, \ldots, N\}, t \in \{1, \ldots, T\} \tag{10}$$

  it is assumed that all units are homogeneous. Why?

- **One-way error component model:**

$$u_{i,t} = \mu_i + \varepsilon_{i,t} \tag{11}$$

  where:
  - μ_i – **the unobservable individual-specific effect**;
  - $\varepsilon_{i,t}$ – the remainder disturbance.

- We can relax assumption that all individuals have the same coefficients

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it} \tag{12}$$

- Note that an i subscript is added to only intercept α_i but the slope coefficients, β_1, . . ., β_k are constant for all individuals.
- **An individual intercept (α_i)** are include to **control** for individual-specific and **time-invariant characteristics. That intercepts are called fixed effects**.
- Fixed effects capture the **individual heterogeneity.**
- The estimation:
  i) The least squares dummy variable estimator
  ii) The fixed effects estimator

- The natural way to estimate fixed effect for all individuals is to include an indicator variable. For example, for the first unit:

$$\mathcal{D}_{1i} = \begin{cases} 1 & i = 1 \\ 0 & \text{otherwise} \end{cases} \tag{13}$$

  The number of dummy variables equals N. **It is not feasible to use the least square dummy variable estimator when N is large**
- We might rewrite the fixed regression as follows:

$$y_{it} = \sum_{j=1}^{N} \alpha_i \mathcal{D}_{ji} + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{i,t} \tag{14}$$

- Why α is missing?

The Fixed Effects (Within Group) Estimator

- Let us start with simple fixed effects specification for individual i:

$$y_{it} = \alpha_i + \beta_1 x_{1it} + \ldots + \beta_k x_{kit} + u_{it} \qquad t = 1, \ldots, T \tag{15}$$

- **Average the observation across time** and using the assumption on time-invariant parameters we get:

$$\bar{y}_i = \alpha_i + \beta_1 \bar{x}_{1i} + \ldots + \beta_k \bar{x}_{ki} + \bar{u}_i \tag{16}$$

where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$, $\bar{x}_{1i} = \frac{1}{T}\sum_{t=1}^{T} x_{1it}$, $\bar{x}_{ki} = \frac{1}{T}\sum_{t=1}^{T} x_{kit}$ and $\bar{u}_i = \frac{1}{T}\sum_{t=1}^{T} u_i$

- Now we substract (16) from (15)

$$(y_{it} - \bar{y}_i) = \underbrace{(\alpha_i - \alpha_i)}_{=0} + \beta_1(x_{1it} - \bar{x}_{1i}) + \ldots + \beta_k(x_{itk} - \bar{x}_{ki}) + (u_{it} - \bar{u}_i) \tag{17}$$

Using notation: $\tilde{y}_{it} = (y_{it} - \bar{y}_i)$, $\tilde{x}_{1it} = (x_{1it} - \bar{x}_{1i})$, $\tilde{x}_{kit} = (x_{kit} - \bar{x}_{ki})$ $\tilde{u}_{it} = (u_{it} - \bar{u}_i)$,
we get

$$\tilde{y}_{it} = \beta_1 \tilde{x}_{1t} + \ldots + \beta_k \tilde{x}_{kt} + \tilde{u}_{i,t} \tag{18}$$

Note that we **do not estimate directly fixed effects**.

- We can test estimates of intercept to verify whether the fixed effects are different among units:

$$\mathcal{H}_0 : \alpha_1 = \alpha_2 = \ldots = \alpha_N. \tag{19}$$

- The poolability test is design to test joint significance of individual-specific intercepts.
- To test (19) we estimate: i) **unrestricted model** (the least squares dummy variable estimator) and ii) **restricted model** (pooled regression). Then we calculate sum of squared errors for both models: SSE_U and SSE_R.

$$\mathcal{F} = \frac{(SSE_R - SSE_U)/(N-1)}{SSE_U/(NT-K)} \tag{20}$$

if null is true then $\mathcal{F} \sim \mathcal{F}_{(N-1,NT-K)}$.

Random Effects model

- Let's assume the following model

$$y_{it} = \alpha + X'_{it}\beta + u_{it}, \quad i \in \{1,\dots,N\}, t \in \{1,\dots,T\}. \tag{21}$$

- The error component (u_t) is the sum of the individual specific random component (μ_i) and idiosyncratic disturbance $(\varepsilon_{i,t})$:

$$u_{it} = \mu_i + \varepsilon_{it}, \tag{22}$$

where $\mu_i \sim \mathcal{N}(0, \sigma^2_\mu)$;
and $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$.

- Let's assume the following model

$$y_{it} = \alpha + X'_{it}\beta + u_{it}, \quad i \in \{1, \dots, N\}, t \in \{1, \dots, T\}. \tag{21}$$

- The error component (u_t) is the sum of the individual specific random component (μ_i) and idiosyncratic disturbance $(\varepsilon_{i,t})$:

$$u_{it} = \mu_i + \varepsilon_{it}, \tag{22}$$

where $\mu_i \sim \mathcal{N}(0, \sigma^2_\mu)$;
and $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$.

- Note that independent variables can be time invariant.
- Individual (random) effects are independent:

$$\mathbb{E}\left(\mu_i, \mu_j\right) = 0 \qquad \text{if} i \neq j. \tag{23}$$

- Estimation method: GLS (*generalized least squares*).

- The error component (u_t):

$$u_{it} = \mu_i + \varepsilon_{it}, \tag{24}$$

where $\mu_i \sim \mathcal{N}(0, \sigma^2_\mu)$ and $\varepsilon_{i,t} \sim \mathcal{N}(0, \sigma^2_\varepsilon)$.

- Diagonal elements of the variance covariance matrix of the error term:

$$\begin{aligned} \mathbb{E}\left(u^2_{it}\right) &= \mathbb{E}\left(\mu^2_i\right) + \mathbb{E}\left(\varepsilon^2_{it}\right) + 2cov\left(\mu_i, \varepsilon_{it}\right) \\ &= \sigma^2_\mu + \sigma^2_\varepsilon \end{aligned}$$

- Non-diagonal elements of the variance covariance matrix of the error term $(t \neq s)$:

$$cov\left(u_{it}, u_{is}\right) = \mathbb{E}\left(u_{it}u_{is}\right) = \mathbb{E}\left[\left(\mu_i + \varepsilon_{it}\right)\left(\mu_i + \varepsilon_{is}\right)\right].$$

- After manipulation:

$$cov\left(u_{it}, u_{is}\right) = \underbrace{\mathbb{E}\left(\mu^2_i\right)}_{\sigma^2_\mu} + \underbrace{\mathbb{E}\left(\mu_i\varepsilon_{it}\right)}_{0} + \underbrace{\mathbb{E}\left(\mu_i\varepsilon_{is}\right)}_{0} + \underbrace{\mathbb{E}\left(\varepsilon_{it}\varepsilon_{is}\right)}_{0} = \sigma^2_\mu$$

- Finally, variance covariance matrix of the error term for given individual (i):

$$\mathbb{E}\left(u_{i.}u_{i.}'\right) = \Sigma_{u,i} = \left(\sigma_\mu^2 + \sigma_\varepsilon^2\right) \begin{bmatrix} 1 & \rho & \dots & \rho \\ \rho & 1 & \vdots & \rho \\ \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \dots & 1 \end{bmatrix}, \tag{25}$$

  where

$$\rho = \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma_\varepsilon^2}.$$

- Note that Σ_u is block diagonal with equicorrelated diagonal elements $\Sigma_{u,i}$ but not spherical.
- Although disturbances from different (cross-sectional) units are independent presence of the time invariant random effects (μ_i) leads to equi-correlations among regression errors belonging to the same (cross-sectional) unit.

- Using the GLS estimator we have:

$$\hat{\beta}^{RE} = \left(X'\Sigma^{-1}X\right)^{-1} X'\Sigma^{-1}y, \tag{26}$$

$$Var\left(\hat{\beta}^{RE}\right) = \left(X'\Sigma^{-1}X\right)^{-1}, \tag{27}$$

but we don't know Σ !

- We know that $\Sigma = \mathbb{E}\left(uu'\right)$ is block-diagonal and:

$$cov\left(u_{it}, u_{js}\right) = \begin{cases} 0 & \text{if} \quad i \neq j, \\ \sigma_\mu^2 + \sigma_\varepsilon^2 & \text{if} \quad i = j \text{ and } s = t, \\ \rho\left(\sigma_\mu^2 + \sigma_\varepsilon^2\right) & \text{if} \quad i = j \text{ and } s \neq t, \end{cases}$$

where $\rho = \sigma_\mu^2 / \left(\sigma_\mu^2 + \sigma_\varepsilon^2\right)$.

- But we do not know σ_μ and σ_ε.
- There are many different strategies to estimates σ_μ and σ_ε.

- We use the GLS transforms of the independent (x^*_{jit}) and dependent variable (y^*_{it}):

$$x^*_{jit} = x_{jit} - \hat{\theta}_i \bar{x}_{ji}$$
$$y^*_{it} = y_{it} - \hat{\theta}_i \bar{y}_i$$

where $\bar{x}_{ji}$ and $\bar{y}_i$ are the individuals means.

- Estimates of the transforming parameter:

$$\hat{\theta}_i = 1 - \sqrt{\frac{\hat{\sigma}^2_\varepsilon}{T_i \hat{\sigma}^2_\mu + \hat{\sigma}^2_\varepsilon}}.$$

- The estimates of the idiosyncratic error component σ_ε:

$$\hat{\sigma}^2_\varepsilon = \frac{\sum_i^n \sum_t^{T_i} \hat{u}^2_{it}}{NT - N - K + 1}$$

where

$$\hat{u}^2_{it} = (y_{it} - \bar{y}_i + \bar{y}) - \hat{\alpha}^{Within} - (x_{it} - \bar{x}_i + \bar{x})\,\hat{\beta}^{Within},$$

where $\hat{\alpha}^{Within}$ and $\hat{\beta}^{Within}$ stand for the within estimates.

- The error variance of individual specific random component (σ^2_μ):

$$\hat{\sigma}^2_\mu = \frac{SSR^{Between}}{N-K} - \frac{\hat{\sigma}^2_\varepsilon}{\bar{T}}$$

where $\bar{T}$ is the harmonic mean of T_i, i.e., $\bar{T} = n/\sum_i^n (1/T_i)$, and $SSR^{Between}$ stands for the sum of squared residuals from the between regression (details on further lectures):

$$SSR^{Between} = \sum_i^n \left(\bar{y}_i - \hat{\alpha}^{Between} - \bar{x}_i\hat{\beta}^{Between}\right)$$

where $\hat{\beta}^{Between}$ and $\hat{\alpha}^{Between}$ stand for the coefficient estimates from the between regression.

- Method which gives more precise estimates in small samples and unbalanced panels.
- The estimation of σ_ε (the idiosyncratic error component) is the same $\Longrightarrow$ it bases on the residuals from the within regression.
- The variance or the individual error term:

$$\hat{\sigma}^2_{\mu,SA} = \frac{SSR^{Between} - (N-K)\,\hat{\sigma}^2_\varepsilon}{NT - tr},$$

where $SSR^{Between}$ is the sum of the squared residuals from the between regression and

$$tr = \text{trace}\left\{\left(X'PX\right)^{-1} X'ZZ'X\right\}$$

$$P = \text{diag}\left\{\frac{1}{T_i} e_{T_i} e'_{T_i}\right\}$$

$$Z = \text{diag}\left\{e_{T_i}\right\}$$

where e_{T_i} is a $T_i \times 1$ vector of ones.

- In the standard RE model, the variance of the random effect is assumed to be σ_μ^2 ($\mu \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$).
- We can test for the presence of heterogeneity:

$$\begin{aligned} \mathcal{H}_0 : \quad & \sigma_\mu = 0 \\ \mathcal{H}_1 : \quad & \sigma_\mu \neq 0 \end{aligned}$$

- If the null hypothesis is rejected, then we conclude that there are random individual differences among sample members, and that the random effects model is appropriate.
- If we fail to reject the null hypothesis, then we have no evidence to conclude that random effects are present.
- We construct the Lagrange multiplier statistic:

$$LM = \sqrt{\frac{NT}{2(T-1)}} \left(\frac{\sum_{i=1}^{N} \left(\sum_{t=1}^{T} \hat{u}_{it}\right)^2}{\sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it}^2} - 1 \right), \tag{28}$$

where $\hat{u}_{i,t}$ stands for the residuals, i.e., $\hat{u}_{it} = y_{it} - \hat{\alpha}_0 - \hat{\beta}_1 x_{1it} - \ldots - \hat{\beta}_k x_{kit}$.
- Conventionally, $LM \sim \chi^2(1)$. In large samples, $LM \sim \mathcal{N}(0,1)$.

Comparison of Fixed and Random Effects model

If the **true DGP** includes random effect then RE model is preferred:

- The random effects estimator takes into account the random sampling process by which the data were obtained.
- The random effects estimator permits us to estimate the effects of variables that are individually time-invariant.
- The random effects estimator is a generalized least squares estimation procedure, and the fixed effects estimator is a least squares estimator.

Endogenous regressors

If the **error term is correlated with any explanatory variable** then both OLS and GLS estimators of parameters are **biased and inconsistent**.

	Random effects	Fixed effects
Individual effects	$\mu_i \sim \mathcal{N}\left(0, \sigma_\mu^2\right)$ drawn from the random sample $\Longrightarrow$ we can estimate the parameter of distribution, i.e., σ_μ^2	α_i α_i are assumed to be constant over time
Assumptions:	(i) $\mathbb{E}\left(\mu_i \mid \varepsilon_{it}\right) = 0$ (ii) $\mathbb{E}\left(\mu_i \mid x_{it}\right) = 0$ individual effects are independent of the explanatory variable x_{it}	(i) $\mathbb{E}\left(\alpha_i \mid \varepsilon_{it}\right) = 0$
Estimation	GLS	OLS (within or LSDV)
Efficiency	higher	lower
Additional:		impossible to use time invariant regressors (collinearity with α_i)

- The null $\mathcal{H}_0$ in the Hausman test is that **both the random and fixed effect estimates are consistent**.
- If the alternative hypothesis $\mathcal{H}_1$ holds then **the random effect estimates are inconsistent**.
- Test statistics:

$$H = \left[\hat{\beta}^{FE} - \hat{\beta}^{RE}\right]' \left(Var\hat{\beta}^{FE} - Var\hat{\beta}^{RE}\right)^{-1} \left[\hat{\beta}^{FE} - \hat{\beta}^{RE}\right] \tag{29}$$

- The statistics H is distributed χ^2 with degrees of freedom determined by K, i.e., the dimension of the coefficient vector β.

- In general, the Hausman test asks whether the fixed effects and random effects estimates of β are **significantly** different.
- We can test only models with the same set of explanatory variables:
  - We cannot compare the random effects estimates corresponding to time-invariant regressors due to their collinearity with individual intercept.
- The rejection of the null hypothesis indicates that the random effect estimates of β are not consistent or **that the model is wrongly specified (misspecification error)**.
- It is assumed that the fixed effect model is consistent under both null and alternative.
  - What if regressors are not strictly exogenous?
- The Hausman may be used in more general context.

Between Estimator

- **The between estimator** uses **just cross-sectional variation**.
- Averaging over time yields:

$$\bar{y}_i = \alpha^{Between} + \beta_1^{Between}\bar{x}_{1i} + \ldots + \beta_k^{Between}\bar{x}_{ki} + \bar{u}_i, \tag{30}$$

where $\bar{y}_i = T^-1\sum_t y_{it}$, $\bar{x}_{1i} = T^-1\sum_t x_{1it}$, ..., $\bar{x}_{ki} = T^-1\sum_t x_{kit}$, $\bar{u}_i = T^-1\sum_t u_{it}$.

- or in the matrix form:

$$\bar{y} = \alpha^{Between} + \bar{x}\beta^{Between} + u. \tag{31}$$

- The parameters $\alpha^{Between}$, $\beta_1^{Between}$, ..., $\beta_k^{Between}$ can be estimated with the OLS estimator.

Hausman-Taylor estimator

- Let's consider the following one-way RE model:

$$y_{it} = x_{1it}\beta_1 + x_{2it}\beta_2 + z_{1i}\gamma_1 + z_{2i}\gamma_2 + \mu_i + u_{it} \tag{32}$$

where:
x_{1it} are **time-varying** variables; **not correlated with** μ_i
x_{2it} are **time-varying** variables; **correlated with** μ_i
z_{1i} are **time-invariant** variables; **not correlated with** μ_i
z_{2i} are **time-invariant** variables; **correlated with** μ_i

- The RE model estimates on γ_2 (and β_2) are inconsistent.
- The estimator proposed by Hausman and Taylor (1981) takes into account the above correlation.

- **First step**: Within regression for the model including only time-variable regressors, both x_{1it} and x_{2it}. Here, the usual differences from the *temporal* mean are used:

$$(y_{it} - \bar{y}_i) = \beta_1(x_{1i,} - \bar{x}_{1i}) + \beta_2(x_{2it} - \bar{x}_{2i}) + (u_{it} - \bar{u}_i) \tag{33}$$

- Based on the expression above we can estimate variance of the idiosyncratic error, i.e., $\hat{\sigma}^2_{\varepsilon}$.

- **Second step**: construct the *intra-temporal* mean of the residuals from (33):

$$\bar{e} = [\underbrace{(\bar{e}_1, \bar{e}_1, \dots, \bar{e}_1)}_{T}, \dots, \underbrace{(\bar{e}_N, \bar{e}_N, \dots, \bar{e}_N)}_{T}]' \quad (34)$$

- Then make TSLS for $\bar{e}_i$ using:
  variables: z_{1it} (time invariant, not correlated with μ_i), z_{2it} (time invariant, correlated with μ_i)
  instruments: z_{1it}, x_{1it} (time invariant, not correlated with μ_i)
  Specifically,
  1. Regress z_{2it} on z_{1it} as well as x_{1it}.
  2. Use the predicted value from the above regression and create new matrix, i.e., $Z = [z_{1it}, \hat{z}_{21t}]$.
  3. Regress $\bar{e}_i$ on Z to get estimates of γ_1 and γ_2.
  4. Calculate $\sigma^2_{TSLS,\bar{e}}$ the variance of the error components from the above regression.

- Now, we can calculate the variation of the individual-specific error component:

$$\sigma^2_\mu = \sigma^2_{TSLS,\bar{e}} - \frac{\sigma^2_\varepsilon}{T}. \quad (35)$$

- Based on the estimates of σ_μ^2 and σ_ε^2 calculate the conventional in the FGLS regression scale parameter θ:

$$\theta = \sqrt{\frac{\sigma_\varepsilon^2}{\sigma_\varepsilon^2 + T^{-1}\sigma_\mu^2}} \tag{36}$$

- Finally, do a TSLS regression of y^* on X^* with instruments described by V:

$$\begin{aligned} y^* &= y_{it} - \theta y_{it}, & (37)\\ X^* &= [x_{1it}, x_{2it}, z_{1i}, z_{2i}] - \theta[x_{1it}, x_{2it}, z_{1i}, z_{2i}], & (38)\\ V &= [(x_{1it} - \bar{x}_{1i}), (x_{2it} - \bar{x}_{2i}), z_{1i}, \bar{x}_{1i}], & (39) \end{aligned}$$

  more specifically:

  1. Regress X^* on the instruments (V) and obtain fitted values, i.e., $\hat{X}^*$,
  2. Regress y^* on the predicted values from the previous step, i.e, $\hat{X}^*$, in order to get the estimates of $[\beta, \gamma]$.

- The estimates of the variance-covariance of the structural parameters are a little bit more complicated.