PUCP

DEPARTAMENTO DE ECONOMÍA
LABORATORIO DE ECONOMETRÍA: STATA
1ECO31

Sesión 17
Econometría de Datos de Panel

Docente: Juan Palomino

PUCP

índice

1. Características de Datos de Panel

Características de Datos de Panel

Los datos de panel consisten en observaciones de un corte transversal de unidades individuales (hogares, empresas, países, etc.) repetidas sobre el tiempo.

Datos de Panel incluye N individuos observados en T periodos de tiempo regulares.

Ejemplos de modelos de datos de panel

- **Economía Laboral**: efecto de educación sobre ingreso, con data a través del tiempo e individuos.
- **Crecimiento Económico**: efectos del capital físico y humano sobre el crecimiento económico, con data a través de los años y países.

Clasificación de datos de panel

- **Paneles Cortos**: muchos individuos y pocos periodos de tiempo.
- **Paneles Largos**: muchos periodos de tiempo y pocos individuos.
- **Ambos**: muchos periodos de tiempo y muchos individuos.

Tipos de Datos de Panel: Panel Balanceado

Nos referimos a **datos de panel balanceado** cuando para cada observación debe conocerse el individuo i y el periodo temporal t al que se refiere.

Individuo	Año	Edad	Ingreso
1	2018	28	1800
1	2019	29	1950
1	2020	30	2200
2	2018	20	800
2	2019	21	850
2	2020	22	900
500	2018	53	2200
500	2019	54	2400
500	2020	55	2550

Tipos de Datos de Panel: Panel No Balanceado

Nos referimos a **datos de panel no balanceado** cuando para cada observación no falta el individuo i en un año o dos del periodo temporal t al que se refiere.

Individuo	Año	Edad	Ingreso	Sexo
1	2018	28	1800	Femenino
1	2019	29	1950	Femenino
2	2018	20	800	Masculino
2	2019	21	850	Masculino
2	2020	22	900	Masculino
500	2019	54	2400	Femenino
500	2020	55	2550	Femenino

2. Regresores y Variable Dependiente

Regresores

- **Regresores que varían x_{it}**
  - Ingreso anual de una persona, consumo anual de un producto

- **Regresores invariantes en el tiempo $x_{it} = x_i$ para todo t**
  - Sexo, raza, lugar de nacimiento

- **Regresores invariantes de individuos $x_{it} = x_t$ para todo i**
  - Tendencia de tiempo, tendencia de economía tales como tasa de desempleo

Variación para la variable dependiente y regresores

- **Variación Overall**: variación sobre el tiempo e individuos.
- **Variación Between**: variación entre individuos
- **Variación Within**: variación dentro de individuos (sobre tiempo).

id	Time	Variable	Individual mean	Overall mean	Overall deviation	Between deviation	Within deviation	Within deviation (modified)
i	t	x_{it}	$\bar{x}_i$	$\bar{x}$	$x_{it} - \bar{x}$	$\bar{x}_i - \bar{x}$	$x_{it} - \bar{x}_i$	$x_{it} - \bar{x}_i + \bar{x}$
1	1	9	10	20	-11	-10	-1	19
1	2	10	10	20	-10	-10	0	20
1	3	11	10	20	-9	-10	1	21
2	1	20	20	20	0	0	0	20
2	2	20	20	20	0	0	0	20
2	3	20	20	20	0	0	0	20
3	1	25	30	20	5	10	-5	15
3	2	30	30	20	10	10	0	20
3	3	35	30	20	15	10	5	25

Variación para la variable dependiente y regresores

- **Individual mean**: $\bar{x}_i = \frac{1}{T}\sum_t x_{it}$
- **Overall mean:** $\bar{x} = \frac{1}{NT}\sum_i \sum_t x_{it}$
- **Overall variance:** $s_0^2 = \frac{1}{NT-1}\sum_i \sum_t (x_{it} - \bar{x})^2$
- **Between variance:** $s_B^2 = \frac{1}{N-1}\sum_i (\bar{x}_i - \bar{x})^2$
- **Within variance:** $s_W^2 = \frac{1}{NT-1}\sum_i \sum_t (x_{it} - \bar{x})^2 = \frac{1}{NT-1}\sum_i \sum_t (x_{it} - \bar{x}_i + \bar{x})^2$
- La variación overall puede ser descompuesta en variación between y variación within:

$$s_0^2 \approx s_B^2 + s_W^2$$

- Regresores de tiempo invariante (raza, sexo) tienen variación within cero.
- Regresores invariantes de individuos (tiempo, tendencia económica) tiene variación between cero.
- Nosotros necesitamos revisar la data para ver si la variación between o within es más grande para cada variable.

Estadísticas Descriptivas

- Existe una distribución de cada uno de ellos que caracterizar: su máximo, mínimo, percentiles, varianza, etc.
- Para variables discretas, una tabulación de valores puede ofrecer:
  - **Overall**: observaciones que toman ese valor x_{it}
  - **Between**: individuos para los que alguna vez toma ese valor $\bar{x}_i$
  - **Within**: porcentaje de individuos que nunca cambia de valor $x_{it} - (\bar{x}_i) + \bar{x}$

3. Modelos de Datos de Panel

Modelos de Datos de Panel

- Modelos de datos de panel describe el comportamiento individual tanto a través del tiempo y a través de individuos.

- Hay tres tipos de modelos:

  - ❖ Modelo agrupado (pooled)
  - ❖ Modelo de efectos fijos
  - ❖ Modelo de efectos aleatorios

Modelo Pooled

El modelo pooled (agrupado) específica coeficientes constantes, los supuestos habituales para el análisis transversal:

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

Este es el modelo más restrictivo de datos de panel y es usado como punto de partida en la literatura.

Modelo de Efectos específicos individuales

- Nosotros asumimos que hay heterogeneidad no observada a través de individuos capturados por α_i.
    - **Ejemplo**: habilidad no observada de un individuo que afecta los salarios.

- La principal pregunta es si los efectos específicos individuales α_i están correlacionados con los regresores:
    - Si ellos están correlacionados, usamos el modelo de efectos fijos.
    - Si ellos no están correlacionados, usamos el modelo de efectos aleatorios.

Modelo de Efectos Fijos (FE)

- Este modelo permite que los efectos específicos individuales α_i esten correlacionados con los regresores x.
- Incluimos α_i como interceptos, en donde cada individuo tiene un diferente intercepto y los mismos parámetros de pendientes.

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

- Nosotros podemos recuperar los efectos específicos individuales después de la estimación como:

$$\hat{\alpha}_i = \bar{y}_i - \bar{x}_i'\hat{\beta}$$

- En otras palabras, los efectos específicos individuales son las variaciones restantes en la variable dependiente que no pueden ser explicadas por los regresores.
- Dummies temporales pueden ser incluidas en los regresores x.

Modelo de Efectos Aleatorios (RE)

- El modelo de efectos aleatorios asume que los efectos específicos individuales α_i están distribuidos independientemente de los regresores.

- Cada individuo tiene los mismos parámetros de pendiente y un término de error compuesto:

$$\epsilon_{it} = \alpha_i + u_{it}$$

- Entonces:

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

- Aquí $var(\epsilon_{it}) = \sigma_\alpha^2 + \sigma_u^2$ y $cov(\epsilon_{it}, \epsilon_{is}) = \sigma_\alpha^2$, de tal manera que $\rho_\epsilon = corr(\epsilon_{it}, \epsilon_{is}) = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \sigma_u^2}$

- ρ es la correlación interclase del error. ρ es la fracción de la varianza en el error debido a los efectos individuales específicos. Este se aproxima a 1 si los efectos individuales dominan los errores idiosincráticos.

4. Test Breusch-Pagan y Hausman

Test de Multiplicador de Lagrange Breusch-Pagan

- ¿Cómo podemos saber si es necesario usar el modelo de efectos aleatorios o el modelo agrupado?
- Test de Breusch-Pagan LM para efectos aleatorios vs modelo agrupado.
- Testea si σ_ϵ^2 o equivalentemente $corr(\epsilon_{it}, \epsilon_{is})$ es significativamente diferente de cero.
- Si el test LM es significativo, usamos el modelo de efectos aleatorios en vez del modelo OLS.

Test de Hausman

- Test de Hausman evalúa si hay una diferencia significativa entre los estimadores de efectos fijos y aleatorios.
- El estadístico de test de Hausman puede ser calculado solo para los regresores de tiempo variante.
- El estadístico de test de Hausman es:

$$H = \left(\hat{\beta}_{RE} - \hat{\beta}_{FE}\right)' (V\left(\hat{\beta}_{RE}\right) - V(\hat{\beta}_{FE}))(\hat{\beta}_{RE} - \hat{\beta}_{FE})$$

- Se distribuye chi-cuadrado con grados de libertad igual al número de parámetros para los regresores de tiempo variante.
- Si el test de Hausman es no significativo, usamos efectos aleatorios.
- Si el test de Hausman es significativo usamos efectos fijos

5. Ejemplos

Ejemplo: Modelo de crecimiento tradicional de Solow

- La ecuación a estimar del modelo es el siguiente:

$$\frac{\ln y_{it} - \ln y_{it-1}}{T} = \beta_0 + \beta_1 \ln y_{it-1} + \beta_2 \ln S_{K,it} + \beta_3 \ln S_{H,it} + \beta_4 \ln(n_{it} + g + \delta)$$

- Utilizando los datos del Sistema de Información Regional para la Toma de Decisiones "SIRTOD", se tienen las siguientes variables:

  - $\ln(PBIpc_{it})$: Logaritmo del PBI per cápita regional
  - $\ln(tpet_{it} + 0.05)$: Logaritmo de la tasa de crecimiento de la población en edad de trabajar más la depreciación y crecimiento tecnológico
  - $\ln(sk)$: Logaritmo del capital físico
  - $\ln(sh)$: Logaritmo del capital humano (años de escolaridad)

- Se tiene una muestra de 24 regiones del Perú para el periodo 2021-2020.

PUCP