

Lección 1.
Análisis de Datos de panel: una
metodología útil para la investigación
aplicada



Teodosio Pérez Amaral,
Universidad Complutense de Madrid
Telefónica, mayo 2005

1



-
- 1. Introducción.
 - 2. Ventajas e inconvenientes.
 - 3. Dos Modelos habituales.
 - 4. Conclusiones.

2



■ 1. Introducción.

■ ¿Qué son datos de panel?

■ Muestra de individuos a lo largo del tiempo.

■ Disponibilidad de

- datos de panel,
- ordenadores y
- programas comerciales.

■ ¿Merece la pena el esfuerzo de aprender la metodología de datos de panel?

3



Ejemplos de bases de datos de panel

- NLS, National Longitudinal Surveys of Labor Market Experience. EEUU
- PSID, Panel Study of Income Dynamics, U Michigan.
- SEP, Socioeconomic Panel, Holanda
- ECHP, European Community Household Panel,
- Varios en países en desarrollo, Banco Mundial.

4



España

- Central de Balances, B. España.
- Encuesta Industrial, SEPI.
- Empresas de Capital Riesgo. José Martí.
- Panel de consumidores y empresas, red.es.
- Panel de abonados, Telefónica.
- Panel de usuarios, Consorcio de Transportes de Madrid.

5



2. Ventajas e inconvenientes

- Ventajas:
 - Mayor número de observaciones.
 - Más variabilidad en las variables.
 - Mayor número de grados de libertad.
 - Más y mejor información.

6

Más ventajas

- Menos multicolinealidad entre los datos de las variables explicativas,
- Mayor eficiencia en la estimación.
- Permite responder a preguntas que no pueden ser respondidas con otros datos.
- Alivia el problema de variables omitidas. Por ej, si no varían en el tiempo se pueden eliminar tomando diferencias.

7

Tratamiento de la heterogeneidad

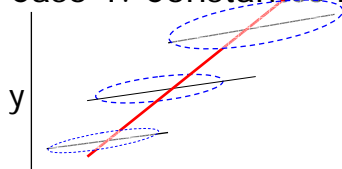
$$y_{it} = \alpha^*_i + \beta_i x_{it} + u_{it} \quad i = 1, \dots, N, \\ t = 1,$$

..., T.

modelo conjunto

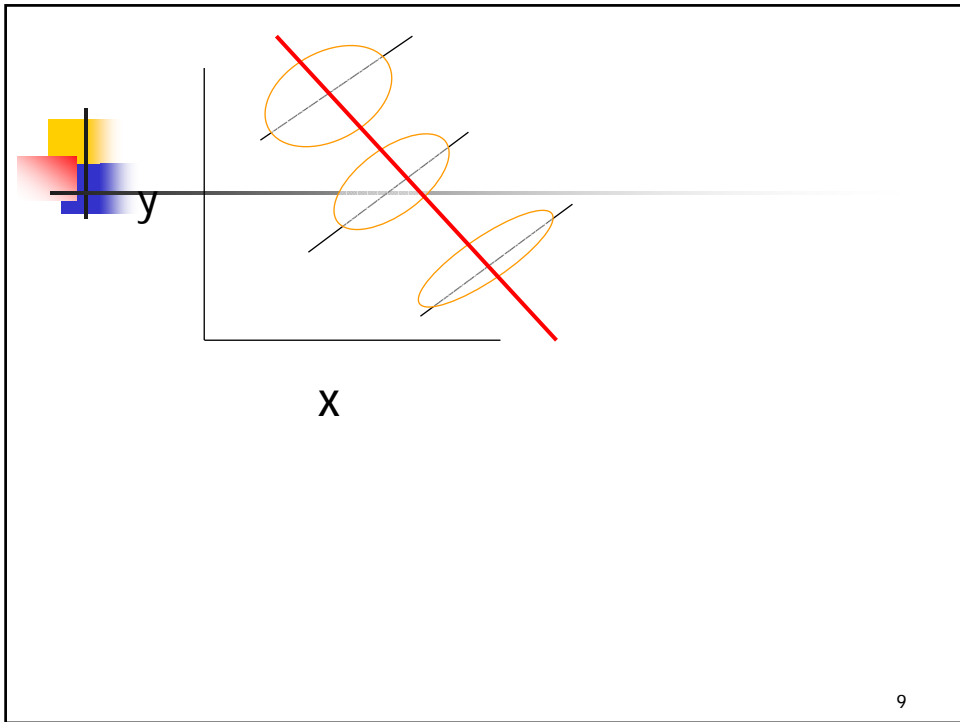
$$y_{it} = \alpha^* + \beta x_{it} + u_{it}$$

Caso 1. Constantes heterogéneas $\alpha^*_i \neq \alpha^*$

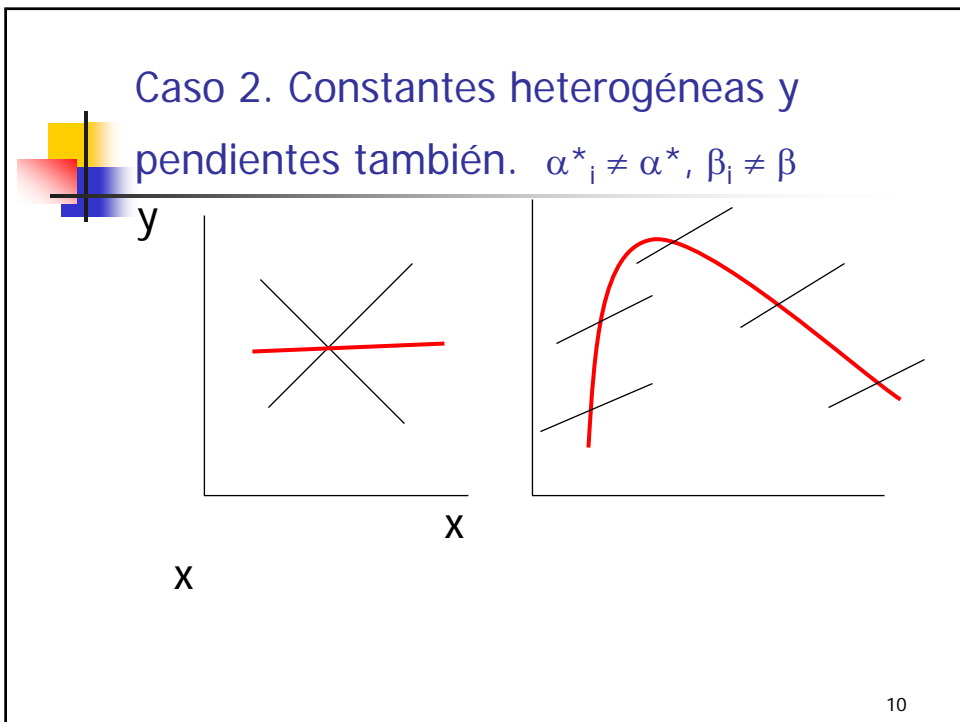


x

8



9



10



- La homogeneidad de constantes y/o pendientes se comprueba mediante las hipótesis nula y alternativa correspondientes en cada caso.
- Usando los contrastes F habituales.
- Lo veremos en el Tema 2.

11



Inconvenientes

- Datos más complejos.
- Métodos econométricos algo más específicos y complejos. Menos familiares e intuitivos.
- Heterogeneidad no tratada.
- Paneles incompletos.

12

3. Dos modelos habituales

- Tratamiento de la heterogeneidad.

- 1. Constantes y pendientes iguales

$$y_{it} = \alpha^* + \beta x_{it} + u_{it}$$

- 2. Constantes distintas y pendientes iguales

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

13

Modelo de efectos fijos: o intragrupos

- Constantes distintas, pendientes iguales

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

- Usamos una variable ficticia para cada individuo (o empresa).
- Se puede estimar por MCO, mínimos cuadrados ordinarios.
- Estimadores de β insesgados y consistentes con N ó T ó ambos $\rightarrow \infty$.

14

Modelo de efectos variables o componentes del error



$$y_{it} = \mu + \beta x_{it} + v_{it} \quad i = 1, \dots, N, \\ t = 1, \dots, T.$$

- $v_{it} = \alpha_i + u_{it} \quad (+ \lambda_t)$
- Estimación por MCG, mínimos cuadrados generalizados.
- Más eficiente, que MCO, pero inconsistente si hay correlación entre efectos individuales y regresores.

15

Elegir entre ambos modelos: efectos fijos o aleatorios



- Criterios elección entre **MODELOS**:
- 1. Mundlak.
- **Modelo de Efectos fijos.** Si la muestra es la población, son pocos individuos, se representan a sí mismos, su identidad es importante.

16




- **Modelo de efectos aleatorios.** Si la muestra representa a una población más amplia, la identidad de cada individuo no es importante, si cada individuo representa a otros, si la población es grande.

17



- ¿Y cómo se estimaría en caso de aleatorios?
- Elección del **ESTIMADOR** en caso de efectos ALEATORIOS.
- En el caso de que haya correlación entre efectos individuales y regresores, MCG sería inconsistente y habría que estimar por MCO con variables ficticias.

18


- 
- ¿Cómo sabemos si hay correlación entre efectos individuales y regresores?

- **Contraste de Hausman.**

- H_0 : No hay correlación entre α_i y las x_i
- H_1 : Si hay correlación entre α_i y las x_i

- Procedimiento práctico: mide la distancia entre las dos estimaciones: por MCO con variables ficticias y por MCG.

19

- 
- Si la distancia es grande, habría correlación entre efectos y regresores, MCG sería inconsistente, y habría que estimar por MCO con variables ficticias.

- Si la distancia es pequeña, no habría correlación entre efectos individuales y regresores, MCG sería consistente y podremos estimar por MCG que es más eficiente que MCO con variables ficticias.

20



Enfoque generalizable a:

- errores de medida en las variables,
- modelos dinámicos,
- ecuaciones simultáneas,
- máxima verosimilitud,
- variable dependiente limitada,
- heteroscedasticidad,
- autocorrelación,
- paneles incompletos,

21



4. Conclusiones.

- Panel: tipo de datos cada vez más frecuente.
- Ordenadores y programas disponibles.
- Metodología accesible y rentable a medio plazo, si se va a utilizar.
- Metodología de datos panel es útil para trabajos de investigación aplicada.

22