

Modelos de datos de panel

Román Salmerón Gómez

Grado en Economía

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos aleatorios

Selección de modelos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos aleatorios

Selección de modelos

Contenidos

Introducción

Introducción

El modelo

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos
aleatorios

Selección de modelos

Introducción

Hasta el momento se han aplicado técnicas de regresión para datos de sección cruzada o de serie temporal, por lo que faltaría abordar el caso en el que se dispone de la combinación de ambos: los datos de panel. Esto es, datos de distintos individuos (hogares, empresas, regiones, etc.) en distintos momentos del tiempo (días, semanas, meses, trimestres, años, etc).

Los paneles pueden ser:

- Balanceados cuando el número de períodos es igual para todos los individuos y no balanceados cuando es diferente.
- Corto o micro cuando el número de sujetos, N , es mayor que el número de períodos, T y largo o macro cuando T es mayor que N .

Ventajas:

- Permite trabajar con un mayor número de observaciones.
- Permite capturar la heterogeneidad entre los individuos o en el tiempo.
- Se reduce la colinealidad entre las variables explicativas.

Un ejemplo de datos de panel es aquel en el que se dispone de la siguiente información:

Comunidad Autónoma	Año	PIB
Andalucía	2010	139397
Andalucía	2011	141348
Andalucía	2012	140757
...
Aragón	2010	35023
Aragón	2011	32232
Aragón	2012	33234
...

En la tabla anterior se recoge el PIB de cada una de las comunidades autónomas españolas en tres años consecutivos.

Consideremos el modelo

$$y_{it} = \alpha_{it} + x'_{it}\beta + u_{it}, \quad u_{it} \sim N(0, \sigma_u^2), \quad (1)$$

donde x'_{it} es un vector que contiene k variables predeterminadas, β es un vector de k parámetros, i representa a los individuos ($i = 1, \dots, N$), t representa el tiempo ($t = 1, \dots, T$) y α_{it} recoge la heterogeneidad provocada por los efectos de los individuos y/o tiempo provocada por variables no observables.

Dependiendo de la consideración que se le dé al término independiente se distinguen tres enfoques:

Modelo agrupado: es constante para todos los individuos y en los períodos (es decir, $\alpha_{it} = \alpha$).

Efectos fijos: el término independiente puede ser distinto para cada individuo (es decir, $\alpha_{it} = \alpha_i$), cada período (es decir, $\alpha_{it} = \alpha_t$) o ambos.

Efectos variables o aleatorios: el término independiente, α_{it} , es una variable aleatoria.

Contenidos

Introducción

Efectos de grupo

Modelo agrupado

Modelo de efectos fijos

Modelo de efectos
aleatorios

Selección de modelos

Efectos de grupo

Cuando $\alpha_{it} = \alpha$ tenemos el modelo agrupado (o “pooled”):

$$y_{it} = \alpha + x'_{it}\beta + u_{it}, \quad u_{it} \sim N(0, \sigma_u^2),$$

que será estimado por MCO.

El inconveniente de este modelo es que ignora la estructura de panel de los datos y puede incumplir la hipótesis de no autocorrelación entre las perturbaciones.

Además, podríamos estar interesados en determinar si estimamos un modelo agregado o T modelos de series temporales. Si las perturbaciones cumplen las hipótesis básicas, podemos aplicar un test de Chow:

$$F = \frac{(\vec{e}_r' \vec{e}_r - \vec{e}' \vec{e}) / (N - 1)}{\vec{e}' \vec{e} / (N \cdot T - k)} \sim F_{N-1, N \cdot T - k},$$

donde \vec{e}_r son los residuos del modelo agrupado y \vec{e} son los residuos de estimar los modelos por separado.

Este test es similar a realizar un contraste mediante una F usando variables ficticias para la constante y la pendiente.

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos fijos

Modelo de efectos fijos

Estimación

Modelo de efectos
aleatorios

Selección de modelos

Modelo de efectos fijos

Modelo de efectos fijos individuales: Sea el modelo

$$y_{it} = \alpha_i + x'_{it}\beta + u_{it}, \quad u_{it} \sim N(0, \sigma_u^2).$$

En este modelo tenemos N términos independientes que recogen las diferencias entre los distintos individuos y que se conocen como efectos fijos individuales. La variación de los efectos fijos individuales proviene de las variables omitidas que varían entre los distintos individuos pero no en el tiempo.

Modelo de efectos fijos temporales: Un modelo con efectos fijos temporales es aquel que permite tener en cuenta las variables que son constantes entre individuos, pero evolucionan en el tiempo:

$$y_{it} = \alpha_t + x'_{it}\beta + u_{it}, \quad u_{it} \sim N(0, \sigma_u^2).$$

En este caso tenemos T interceptos que recogen los efectos fijos temporales que provienen de las variables omitidas que varían en el tiempo, pero no entre los individuos.

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos fijos

Modelo de efectos fijos

Estimación

Modelo de efectos
aleatorios

Selección de modelos

Modelo de efectos fijos individuales y temporales: Si las variables omitidas son constantes en el tiempo pero varían entre los individuos, mientras que otras son constantes para los individuos pero varían en el tiempo, entonces se habla de efectos fijos individuales y temporales:

$$y_{it} = \alpha_{it} + x'_{it}\beta + u_{it}, \quad u_{it} \sim N(0, \sigma_u^2).$$

Estimación intra grupos o within con dummies

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos fijos

Modelo de efectos fijos

Estimación

Modelo de efectos
aleatorios

Selección de modelos

Consideremos el caso del modelo de efectos fijos individuales. En éste, además del término independiente, podemos recoger mediante variables ficticias la variación entre los individuos. En tal caso se estimaría el siguiente modelo:

$$y_{it} = \alpha_1 + \sum_{j=2}^N \alpha_j d_j + x'_{it} \beta + u_{it},$$

donde α_1 es el efecto del primer individuo (categoria base) y:

- d_j toma el valor 1 si el dato corresponde al individuo j y cero en otro caso.
- α_j son los coeficientes de las variables ficticias y representan el grado en que los valores de los interceptos del resto de individuos difieren respecto del intercepto base α_1 (por ejemplo, $\alpha_1 + \alpha_2$ representa el efecto individual del individuo 2).

En el caso de efectos fijos temporales la variable binaria d_j tomará el valor 1 si el dato corresponde al tiempo j y cero en otro caso. Mientras que bajo ambos supuestos se tendría: $y_{it} = \alpha_1 + \sum_{j=2}^N \alpha_j d_j + \delta_1 + \sum_{l=2}^T \delta_l d'_l + x'_{it} \beta + u_{it}$.

Estimación intra grupos o within con dummies

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos fijos

Modelo de efectos fijos

Estimación

Modelo de efectos
aleatorios

Selección de modelos

En el planteamiento anterior pueden surgir ciertos inconvenientes:

- Pérdida de grados de libertad.
- Aumento del problema de la multicolinealidad.
- Las perturbaciones pueden presentar heteroscedasticidad entre las unidades.
- Las perturbaciones pueden estar autocorrelacionadas en el tiempo.
- Si N (y/o T) es muy grande, habrá un número elevado de regresores: por ejemplo, 100 observaciones implica 99 dummies. Lo cual puede incluso llegar a hacer inviable la estimación.

Estimación intra grupos o within corregida por la media

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos fijos

Modelo de efectos fijos

Estimación

Modelo de efectos
aleatorios

Selección de modelos

Una alternativa consiste en la utilización de una regresión sobre las variables obtenidas como desviaciones respecto a las medias individuales. De esta forma, se evitan los efectos individuales.

Partiendo del modelo (1) escrito con respecto a las medias (de cada individuo):

$$\bar{y}_{.t} = \alpha_t + \bar{x}'_{.t}\beta + \bar{u}_{.t},$$

donde $\bar{y}_{.t} = \frac{1}{N} \sum_{i=1}^N y_{it}$, y restándoselo al modelo (1), se obtiene el modelo en desviaciones respecto de la media:

$$y_{it} - \bar{y}_{.t} = (x_{it} - \bar{x}_{.t})'\beta + (u_{it} - \bar{u}_{.t}). \quad (2)$$

Este último modelo será el modelo a estimar por MCO (también se denomina intragrupos con efectos fijos o *fixed effect FE*).

Para estimar los efectos de los individuos podemos usar $\hat{\alpha}_t = \bar{y}_{.t} - \bar{x}'_{.t}\hat{\beta}$.

Estimación entre grupos o between

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos fijos

Modelo de efectos fijos

Estimación

Modelo de efectos
aleatorios

Selección de modelos

Los modelos entre grupos o *between* analiza la variabilidad entre unidades de sección cruzada, por tanto, usan las medias de los datos temporales de cada individuo. Por tanto, se llegaría a un modelo análogo al (2) en el cual las medias se obtienen para cada periodo temporal (en lugar de para cada individuo). En la práctica, se utilizan poco porque los modelos con efectos aleatorios son superiores (el estimador *between* ignora la información temporal existente dentro de los individuos).

Las estimaciones obtenidas intra y entre grupos en desviaciones a las medias son iguales y coinciden también con la estimación realizada mediante variables binarias. Además, si las perturbaciones son normales se pueden utilizar los métodos usuales para realizar la inferencia.

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

**Modelo de efectos
aleatorios**

Modelo de efectos
aleatorios

Selección de modelos

Modelo de efectos aleatorios

Modelo de efectos aleatorios

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos aleatorios

Modelo de efectos aleatorios

Selección de modelos

En el modelo de efectos aleatorios se considera que la ordenada en el origen, α_{it} , es una variable aleatoria. Si suponemos que está correlacionada con las variables explicativas, se incumple una de las hipótesis básicas del modelo lineal general (las variables explicativas no pueden estar correlacionadas con el término de perturbación), por lo que se hace necesario una alternativa para poder estimar. Si se supone que α_{it} se puede descomponer en una parte constante, α , y otra aleatoria, ϵ_i , la cual suponemos que depende del individuo i -ésimo pero que es constante en el tiempo, esto es, $\alpha_{it} = \alpha + \epsilon_i$, el modelo (1) quedaría:

$$y_{it} = \alpha_{it} + x'_{it}\beta + u_{it} = \alpha + \epsilon_i + x'_{it}\beta + u_{it} = \alpha + x'_{it}\beta + w_{it},$$

donde se han agrupado los dos términos no observables: $w_{it} = \epsilon_i + u_{it}$. Además, las perturbaciones aleatorias del modelo suponemos que cumplen las hipótesis básicas:

$$\epsilon_i \sim N(0, \sigma_\epsilon^2), \quad u_{it} \sim N(0, \sigma_u^2),$$

donde ϵ_i representa las perturbaciones para los datos de corte transversal y u_{it} son las perturbaciones de la combinación de los datos temporales y de corte transversal. En tal caso, las perturbaciones son homoscedásticas ya que $E(w_{it}^2) = \sigma_\epsilon^2 + \sigma_u^2$.

Sin embargo, se puede demostrar que los términos de perturbación de un individuo en dos momentos diferentes del tiempo están correlacionados:

$$\text{corr}(w_{it}, w_{is}) = \frac{\sigma_{\epsilon}^2}{\sigma_{\epsilon}^2 + \sigma_u^2} \neq 0, \quad \forall t \neq s,$$

por lo que la estimación por MCO del modelo sería ineficiente por presentar las perturbaciones autocorrelacionadas. Para corregir este problema: MCG.

Las estimaciones por MCG se obtienen aplicando MCO al modelo agrupado transformado (MAT):

$$\begin{aligned} y_{it} - \lambda \bar{y}_i &= \beta_1(1 - \lambda) + \beta_2(x_{2it} - \lambda \bar{x}_{2i}) + \dots + \beta_k(x_{kit} - \lambda \bar{x}_{ki}) \\ &+ (v_{it} - \lambda \bar{v}_i), \end{aligned}$$

donde $\lambda = 1 - \sqrt{\frac{1}{1 + T \cdot \frac{\sigma_{\epsilon}^2}{\sigma_u^2}}}$.

Para estimar el modelo agrupado transformado es necesario disponer de una estimación de λ . Cuando $\hat{\lambda} = 0$, el MAT coincide con el modelo agrupado y cuando $\hat{\lambda} = 1$ coincide con el modelo de efectos fijos. Así, $\hat{\lambda}$ proporciona una medida de cuánto nos acercamos a un caso u otro.

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos
aleatorios

Selección de modelos

Prueba de
Breusch-Pagan y test
de Hausman

Selección de modelos

Prueba de Breusch-Pagan y test de Hausman

Contenidos

Introducción

Efectos de grupo

Modelo de efectos fijos

Modelo de efectos aleatorios

Selección de modelos

Prueba de Breusch-Pagan y test de Hausman

Para contrastar la hipótesis de que no hay efectos aleatorios se puede usar la prueba de Breusch–Pagan, que es un test de multiplicadores de Lagrange. La hipótesis nula es que se trata de un modelo agrupado, frente a la alternativa de efectos aleatorios. Bajo la hipótesis de normalidad, el estadístico se define como:

$$LM = \frac{NT}{2(T-1)} \left(\frac{\sum_{i=1}^N (\sum_{t=1}^T e_{it})^2}{\sum_{i=1}^N \sum_{t=1}^T e_{it}^2} - 1 \right)^2 \sim \chi_1^2.$$

El test de Hausman se utiliza para determinar si el modelo que se debe elegir es de efectos fijos o variables. El estadístico de contraste es:

$$(\hat{\beta}_{EF} - \hat{\beta}_{EA})' \left[Var(\hat{\beta}_{EF}) - Var(\hat{\beta}_{EA}) \right]^{-1} (\hat{\beta}_{EF} - \hat{\beta}_{EA}) \sim \chi_{k-1}^2.$$

donde k es el número de parámetros del modelo de efectos fijos. Si el valor del estadístico es superior al de las tablas se rechaza la hipótesis nula de efectos aleatorios y, por tanto, elegiríamos el modelo de efectos fijos (ya que es probable que las perturbaciones estén correlacionadas con alguna variable explicativa).