Econometria
Tema 3: MRLM: Variables exògenes qualitatives

Ramon Alemany

Grau Estadística UB-UPC

Curs 2017-18

Presentació

Bibliografia

- GREENE, W. (1999)
  Análisis econométrico. 3a Ed.
  Capítol 8

- WOOLDRIDGE, J. (2009)
  Introducción a la Econometría. Un enfoque moderno. 4a Ed.
  Capítol 7

- STOCK, J. & WATSON, M. (2012)
  Introducción a la Econometría. 3a Ed.
  Capítol 5

Definició i tipologia de variables qualitatives

1. Definició i tipologia de variables qualitatives

Fins ara hem considerat que en el MRLM tant la variables endògena com les variables explicatives eren de tipus quantitatiu. Però, aquest supòsit és molt restrictiu ja que, en ocasions, pot ser interessant emprar factors qualitatius.

Alguns exemples de factors qualitatius poden ser:

- sexe, estat civil, nivell d'estudis, nacionalitat,
- sector econòmic d'empreses
- marca de diferents productes

Els factors qualitatius es caracteritzen perquè no tenen valors numèrics sinó atributs, categories o modalitats.

Definició i tipologia de variables qualitatives

Els factors qualitatius poden estar presents en un MRLM com a:

- **variables endògenes**: per exemple, viure en un pis llogat o de propietat pot ser explicat per l'edat, la renda, etc. En aquest, cas la variable endògena és de tipus qualitatiu.

- **variables explicatives**: per exemple, volem explicar consum de llibres en funció, entre d'altres, del nivell d'estudis dels individus, i aquests poden ser primaris, mitjans o superiors. En aquest cas, la variable explicativa és de tipus qualitatiu.

Definició i tipologia de variables qualitatives

Les variables qualitatives es poden classificar com:

- **Variables dicotòmiques**: són aquelles que presenten únicament dues categories, nomenades també modalitats o atributs. Exemple: el gènere de les persones, que pot ser masculí o femení.
- **Variables politòmiques**: són aquelles que presenten més de dues categories. Exemple: l'estat civil de les persones, que pot ser solter, casat, divorciat i vidu.

Però les variables qualitatives no es poden introduir en el MRLM amb les seves modalitats atès que no prenen valors numèrics.

Especificació de variables qualitatives en el MRLM

Variables dicotòmiques

2. Especificació de variables qualitatives

La solució és la codificació, és a dir, l'assignació d'un valor numèric a cadascuna de les categories que presenta la variable qualitativa.

Variables dicotòmiques

S'utilitzen dos valors, 0 i 1 per a cadascuna de les modalitats de la variable qualitativa.

Per exemple, si es desitja codificar la variable **"Gènere de les persones"** es poden crear dues variables codificades:

$$\text{GÈNERE1} = \begin{cases} 1 & \text{dona} \\ 0 & \text{home} \end{cases} \quad \text{GÈNERE2} = \begin{cases} 0 & \text{dona} \\ 1 & \text{home} \end{cases}$$

Especificació de variables qualitatives en el MRLM

Variables dicotòmiques

I aquestes variables ja codificades reben el nom de **variables fictícies o dummies**.

Però ens adonem que no és necessari crear dues variables doncs amb una sola ja tenim caracteritzades les dues modalitats.

Per tant, en aquestes circumstàncies, es crea únicament una variable fictícia, sent indiferent l'assignació del valor 1 o 0 a cada atribut.

Especificació de variables qualitatives en el MRLM

Variables politòmiques

Variables politòmiques

Quan la variable a codificar és politòmica la codificació és més complexa.

Per exemple, si volguéssim codificar la variable **"Estat civil"** podríem pensar en operar d'una manera semblant al cas de les dicotòmiques i definir una variable com la següent:

$$\mathrm{ESTCIV} = \begin{cases} 0 & \text{solter} \\ 1 & \text{casat} \\ 2 & \text{divorciat} \\ 3 & \text{vidu} \end{cases}$$

Especificació de variables qualitatives en el MRLM

Variables politòmiques

Però amb això estaríem introduint un ordre que no existeix i tindríem problemes importants per interpretar els resultats obtinguts en l'estimació del model.

La solució consisteix en crear diverses variables fictícies. Podríem crear tantes variables fictícies com categories tingui la variable qualitativa:

$$\text{ESTCIV1} = \begin{cases} 1 & \text{solter} \\ 0 & \text{altre cas} \end{cases} \quad \text{ESTCIV2} = \begin{cases} 1 & \text{casat} \\ 0 & \text{altre cas} \end{cases}$$

$$\text{ESTCIV3} = \begin{cases} 1 & \text{divorciat} \\ 0 & \text{altre cas} \end{cases} \quad \text{ESTCIV4} = \begin{cases} 1 & \text{vidu} \\ 0 & \text{altre cas} \end{cases}$$

Especificació de variables qualitatives en el MRLM

Trampa de les fictícies

Trampa de les fictícies:

Per a no tenir un problema de multicol·linealitat perfecta, si existeix terme independent en l'equació cal introduir tantes variables fictícies com categories tingui la variable qualitativa menys una.

$$\boxed{\textbf{nº fictícies} = \textbf{nº categories - 1}}$$

Especificació de variables qualitatives en el MRLM

Trampa de les fictícies

Per exemple, en el cas de la variable **"Gènere"**, si introduïm en el model les dues variables fictícies:

$$\text{GÈNERE1} = \begin{cases} 1 & \text{dona} \\ 0 & \text{home} \end{cases} \quad \text{GÈNERE2} = \begin{cases} 0 & \text{dona} \\ 1 & \text{home} \end{cases}$$

$$\mathbf{y_i = \boldsymbol{\beta}_1 + \boldsymbol{\beta}_2 x_{2i} + \boldsymbol{\beta}_3 \text{GENERE1}_i + \boldsymbol{\beta}_4 \text{GENERE2}_i + u_i}$$

Aleshores es verificarà que:

$$\mathbf{\text{GENERE1}_i + \text{GENERE2}_i = 1 = x_{1i}}$$

de forma que $\rho(\mathrm{X}) < \mathrm{k} + 2$ i no hi haurà solució única al sistema d'equacions normals MQO.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

3. Esquemes additiu, multiplicatiu i interaccions

Suposem que per un conjunt d'individus especifiquem un model de Consum a partir de la Renda i volem recollir els efectes diferencials entre homes i dones:

$$\text{Consum}_i = \beta_1 + \beta_2 \text{Renda}_i + \beta_3 \text{Genere}_i + u_i \qquad i = 1, 2, \ldots, N$$

i on la variable fictícia: $\text{GENERE} = \begin{cases} 1 & \text{home} \\ 0 & \text{dona} \end{cases}$

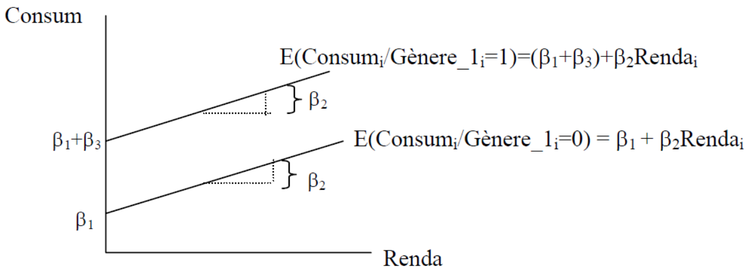
Aleshores:

$$\mathbf{E}(\text{Consum}_i \mid \text{Renda}_i \text{Genere}_i = 1) = \beta_1 + \beta_2 \text{Renda}_i + \beta_3$$
$$\mathbf{E}(\text{Consum}_i \mid \text{Renda}_i \text{Genere}_i = 0) = \beta_1 + \beta_2 \text{Renda}_i$$

i β_3 és l'**efecte diferencial** en el consum entre homes i dones.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu



En aquest model el contrast de significació individual del paràmetre $\boldsymbol{\beta_3}$ suposa contrastar si el consum esperat d'una dona és igual al d'un home suposant que té la mateixa renda disponible.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

Si rebutgéssim la hipòtesis nul·la del contrast de significació individual de β_3, la conclusió serà que el consum de les dones és diferent al dels homes, a igualtat de renda disponible.

Si no rebutgéssim la hipòtesis nul·la, la conclusió serà que el consum de les dones no és diferent al dels homes, a igualtat de renda disponible.

En el cas de que haguéssim rebutjat la hipòtesi nul·la, l'estimació del model quan no s'inclou la variable fictícia donaria lloc a un error d'omissió de variable rellevant.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

També podem introduir una variable que presenti més de dues categories (politòmica). En aquest cas cal tenir una especial precaució ja que una codificació incorrecta podria portar a interpretacions errònies.

Seguint amb el model de Consum, pensem que pot haver diferències en funció de l'estat civil de les persones. En aquest cas, tindríem una variable qualitativa amb quatre modalitats: solter, casat, divorciat i vidu.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

Per analitzar aquesta possibilitat hauríem de crear les variables fictícies necessàries:

$$\text{SOL} = \begin{cases} 1 & \text{solter} \\ 0 & \text{altre cas} \end{cases} \quad \text{CAS} = \begin{cases} 1 & \text{casat} \\ 0 & \text{altre cas} \end{cases}$$

$$\text{DIV} = \begin{cases} 1 & \text{divorciat} \\ 0 & \text{altre cas} \end{cases} \quad \text{VID} = \begin{cases} 1 & \text{vidu} \\ 0 & \text{altre cas} \end{cases}$$

Un cop definides les variables fictícies, cal introduir-les en el MRLM.

Atès el problema de la trampa de les fictícies, tenim dues opcions per introduir aquestes variables fictícies:

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

1. La primera opció passa per **definir una modalitat de la variable qualitativa politòmica com categoria base**, incloent en el MRLM tantes variables fictícies com modalitats hi hagi menys una (la de la categoria base). Aquesta manera de procedir permet mantenir el terme constant.

Així, si prenem com categoria base el ser SOL, hauríem d'introduir les variables fictícies CAS, DIV i VID en el model de regressió inicial obtenint la següent expressió:

$$\text{Consum}_{\mathbf{i}} = \boldsymbol{\beta}_{\mathbf{1}} + \boldsymbol{\beta}_{\mathbf{2}}\text{CAS}_{\mathbf{i}} + \boldsymbol{\beta}_{\mathbf{3}}\text{DIV}_{\mathbf{i}} + \boldsymbol{\beta}_{\mathbf{4}}\text{VID}_{\mathbf{i}} + \boldsymbol{\beta}_{\mathbf{5}}\text{Renda}_{\mathbf{i}} + \mathbf{u}_{\mathbf{i}}$$

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

I prenent esperances condicionades tindríem:

$$\mathbf{E}(\text{Consum}_\mathbf{i} \mid \text{solter}) = \boldsymbol{\beta}_\mathbf{1} + \boldsymbol{\beta}_\mathbf{5}\text{Renda}_\mathbf{i}$$

$$\mathbf{E}(\text{Consum}_\mathbf{i} \mid \text{casat}) = \boldsymbol{\beta}_\mathbf{1} + \boldsymbol{\beta}_\mathbf{2} + \boldsymbol{\beta}_\mathbf{5}\text{Renda}_\mathbf{i}$$

$$\mathbf{E}(\text{Consum}_\mathbf{i} \mid \text{divorciat}) = \boldsymbol{\beta}_\mathbf{1} + \boldsymbol{\beta}_\mathbf{3} + \boldsymbol{\beta}_\mathbf{5}\text{Renda}_\mathbf{i}$$

$$\mathbf{E}(\text{Consum}_\mathbf{i} \mid \text{vidu}) = \boldsymbol{\beta}_\mathbf{1} + \boldsymbol{\beta}_\mathbf{4} + \boldsymbol{\beta}_\mathbf{5}\text{Renda}_\mathbf{i}$$

i la interpretació dels paràmetres de les variables fictícies es realitza en relació a la categoria base.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

Aleshores:

- $\boldsymbol{\beta_1}$ reflecteix el consum esperat d'una persona soltera quan no té renda disponible.
- $\boldsymbol{\beta_2}$ és la diferència entre el consum esperat d'una persona casada i una soltera, amb independència de la seva renda.
- $\boldsymbol{\beta_3}$ és la diferència entre el consum esperat d'una persona divorciada i una soltera, amb independència de la seva renda.
- $\boldsymbol{\beta_4}$ és la diferència entre el consum esperat d'una persona vídua i una soltera, amb independència de la seva renta.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

Tanmateix, també és possible veure la relació que existeix entre cadascun dels grups de la variable qualitativa. Així, tindrem:

- $(\boldsymbol{\beta_3 - \beta_2})$ és la diferència entre el consum esperat d'una persona divorciada i una casada, amb independència de la seva renda.
- $(\boldsymbol{\beta_4 - \beta_2})$ és la diferència entre el consum esperat d'una persona vídua i una casada, amb independència de la seva renda.
- $(\boldsymbol{\beta_4 - \beta_3})$ és la diferència entre el consum esperat d'una persona vídua i una divorciada, amb independència de la seva renda.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

En aquest model podríem contrastar si efectivament existeixen diferències estadísticament significatives en el consum dels individus en funció del seu estat civil.

El procediment seria exactament igual al del cas de variables dicotòmiques. Així, s'hauria de contrastar la significació individual dels paràmetres $\boldsymbol{\beta_1}$,$\boldsymbol{\beta_2}$, $\boldsymbol{\beta_3}$,$\boldsymbol{\beta_4}$.

Si es rebutja la hipòtesi nul·la això voldrà dir que, per a la mateixa quantitat de renda, les persones realitzen un consum diferent segons sigui el seu estat civil.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

A més, es podrien fer contrastos sobre els efectes diferencials entre estats civils sobre el consum esperat.

Per exemple, es podria plantejar el contrast de si el consum esperat d'una persona divorciada és igual al d'una persona vídua, amb independència de la seva renda.

En aquest cas, les hipòtesis nul·la i alternativa serien:

$$\begin{cases} H_0 : \beta_3 = \beta_4 \\ H_A : \beta_3 \neq \beta_4 \end{cases}$$

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

2. La segona opció possible per evitar la trampa de les fictícies passa per incloure en el MRLM tantes variables fictícies com modalitats, però **excloent el terme independent**.

D'aquesta manera el model de regressió seria:

$$\mathrm{C_i} = \boldsymbol{\beta_1^*}\mathrm{SOL_i} + \boldsymbol{\beta_2^*}\mathrm{CAS_i} + \boldsymbol{\beta_3^*}\mathrm{DIV_i} + \boldsymbol{\beta_4^*}\mathrm{VID_i} + \boldsymbol{\beta_5}\mathrm{Renda_i} + \mathbf{u_i}$$

a partir del qual es pot veure que, en termes de valor esperat, tindríem:

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

$$\mathbf{E}(\text{Consum}_\mathbf{i} \mid \text{solter}) = \boldsymbol{\beta}_\mathbf{1}^* + \boldsymbol{\beta}_\mathbf{5}\text{Renda}_\mathbf{i}$$

$$\mathbf{E}(\text{Consum}_\mathbf{i} \mid \text{casat}) = \boldsymbol{\beta}_\mathbf{2}^* + \boldsymbol{\beta}_\mathbf{5}\text{Renda}_\mathbf{i}$$

$$\mathbf{E}(\text{Consum}_\mathbf{i} \mid \text{divorciat}) = \boldsymbol{\beta}_\mathbf{3}^* + \boldsymbol{\beta}_\mathbf{5}\text{Renda}_\mathbf{i}$$

$$\mathbf{E}(\text{Consum}_\mathbf{i} \mid \text{vidu}) = \boldsymbol{\beta}_\mathbf{4}^* + \boldsymbol{\beta}_\mathbf{5}\text{Renda}_\mathbf{i}$$

I es pot concloure fàcilment que cadascun dels coeficients $\boldsymbol{\beta}_\mathbf{1}^*$, $\boldsymbol{\beta}_\mathbf{2}^*$, $\boldsymbol{\beta}_\mathbf{3}^*$ i $\boldsymbol{\beta}_\mathbf{4}^*$ recull el consum esperat de les persones solteres, casades, divorciades i vídues respectivament quan la variable explicativa Renda és igual a 0.

Esquemes additiu, multiplicatiu i interaccions

Esquema additiu

Fins ara, s'ha analitzat si el consum pot dependre del gènere dels individus o del seu estat civil, amb independència de la seva renda disponible. Així, s'ha vist com introduir en el model una o varies variables fictícies.

Aquestes variables s'introdueïen de la mateixa manera que les variables quantitatives vistes fins ara. Aquesta manera d'introduir variables fictícies suposa que únicament varia el terme independent, i fa que aquestes variables tinguin una influència additiva.

Per tant, les denominem **variables fictícies additives** o variables fictícies introduïdes en el MRLM segons un esquema additiu.

Esquemes additiu, multiplicatiu i interaccions

Esquema multiplicatiu

La introducció de variables fictícies mitjançant un esquema additiu no és l'única possibilitat d'incloure aquest tipus de variables en el model.

Seguint amb l'exemple on analitzàvem si el consum depenia de la renda disponible i del gènere dels individus, podríem qüestionar-nos **si la renda disponible afecta de manera diferent al consum dels homes i de les dones**.

Per analitzar aquesta possibilitat és necessari introduir la variable qualitativa que considera el gènere dels individus en el model de manera que interactuï amb la variable renda disponible.

Esquemes additiu, multiplicatiu i interaccions

Esquema multiplicatiu

Això s'aconsegueix incloent variables que siguin el producte de les variables qualitatives i quantitatives.

Aquest procediment d'inclusió de variables qualitatives es denomina esquema multiplicatiu, i les denominem **variables fictícies multiplicatives**.

Introduïm com variables explicatives tots els productes de les variables fictícies amb les quantitatives però tenint en compte no caure en multicol·linalitat perfecte.

Novament tenim dues formes d'especificar el model:

Esquemes additiu, multiplicatiu i interaccions

Esquema multiplicatiu

1.Introduir de manera multiplicativa **tantes variables fictícies com número de categories menys 1**.

En aquest cas el model quedaria:

$$\text{Consum}_i = \beta_1 + \beta_2 \text{Renda}_i + \beta_3 \text{Genere*Renda}_i + u_i$$

A partir d'aquest model, tindrem que l'esperança condicionada del consum quan es tracta d'un home o d'una dona seran:

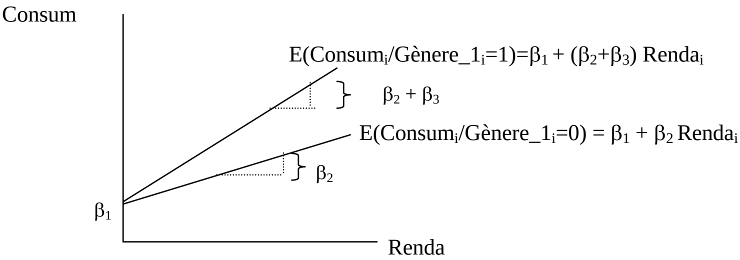
$$E(\text{Consum}_i \mid \text{Renda}_i, \text{Genere}_i = 1) = \beta_1 + (\beta_2 + \beta_3)\text{Renda}_i$$

$$E(\text{Consum}_i \mid \text{Renda}_i, \text{Genere}_i = 0) = \beta_1 + \beta_2 \text{Renda}_i$$

Esquemes additiu, multiplicatiu i interaccions

Esquema multiplicatiu

Gràficament i sota el supòsit hipotètic que $\boldsymbol{\beta_3 > 0}$:



A partir d'aquest gràfic es pot comprovar que la inclusió de la variable fictícia de manera multiplicativa té efectes sobre el pendent del model però no sobre el terme independent.

Esquemes additiu, multiplicatiu i interaccions

Esquema multiplicatiu

A més, es pot contrastar si estadísticament aquesta variable és significativa. Això suposaria contrastar si existeixen diferències en l'efecte d'un increment en una unitat de la variable Renda sobre el consum d'un home i d'una dona.

Les hipòtesis nul·la i alternativa serien:

$$\begin{cases} H_0 : \beta_3 = 0 \\ H_A : \beta_3 \neq 0 \end{cases}$$

Esquemes additiu, multiplicatiu i interaccions

Esquema multiplicatiu

2. La segona possibilitat, per resoldre el problema de la multicol·linealitat perfecta, és introduir de manera multiplicativa tantes variables fictícies com modalitats tingui la variable qualitativa però **no incloure de manera aïllada la variable quantitativa que interactua amb les fictícies**. El model quedaria:

$$\mathrm{C_i} = \beta_1^* + \beta_2^* \mathrm{Genere1{*}Renda_i} + \beta_3 \mathrm{Genere2{*}Renda_i} + u_i$$

I l'esperança condicionada a que el consum el faci un home o una dona serà:

$$E(\mathrm{C_i} \mid \mathrm{Renda_i}, \mathrm{Genere1_i} = 1, \mathrm{Genere2_i} = 0) = \beta_1^* + \beta_2^* \mathrm{Renda_i}$$

$$E(\mathrm{C_i} \mid \mathrm{Renda_i}, \mathrm{Genere1_i} = 0, \mathrm{Genere2_i} = 1) = \beta_1^* + \beta_3^* \mathrm{Renda_i}$$

Esquemes additiu, multiplicatiu i interaccions

Esquema multiplicatiu

En aquest cas, per contrastar si existeix un efecte diferencial en el pendent per ambdós gèneres s'haurà de contrastar la igualtat entre els coeficients dels efectes d'interacció del model, és a dir:

$$\begin{cases} H_0 : \beta_2^* = \beta_3^* \\ \\ H_A : \beta_2^* \neq \beta_3^* \end{cases}$$

En el cas de que la variable qualitativa que volguéssim incloure fos politòmica, el procediment seria exactament el mateix.

Esquemes additiu, multiplicatiu i interaccions

Esquema mixt

Addicionalment, ens podríem plantejar si ser home o dona influeix simultàniament tant en el consum esperat per a qualsevol renda disponible com en el consum quan hi ha una variació unitària de la variable explicativa Renda.

Per reflectir aquesta possibilitat s'hauria d'introduir en el model la variable fictícia tant de manera additiva com multiplicativa, és a dir, mitjançant el que es nomena un **esquema mixt**.

Esquemes additiu, multiplicatiu i interaccions

Esquema mixt

D'aquesta manera, l'especificació seria:

$$\mathrm{C}_{\mathbf{i}} = \boldsymbol{\beta}_{\mathbf{1}} + \boldsymbol{\beta}_{\mathbf{2}}\mathrm{Renda}_{\mathbf{i}} + \boldsymbol{\beta}_{\mathbf{3}}\mathrm{Genere}_{\mathbf{i}} + \boldsymbol{\beta}_{\mathbf{4}}\mathrm{Genere}^{*}\mathrm{Renda}_{\mathbf{i}} + \mathbf{u}_{\mathbf{i}}$$

i partir d'aquest model, tindrem que l'esperança condicionada del consum quan es tracta d'un home o d'una dona és:

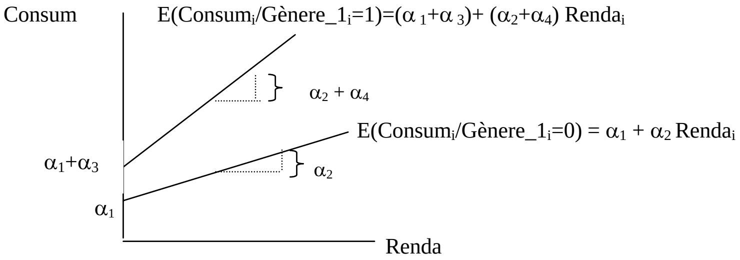
$$\mathbf{E}(\mathrm{C}_{\mathbf{i}} \mid \mathrm{Renda}_{\mathbf{i}}, \mathrm{Genere}_{\mathbf{i}} = \mathbf{1}) = \boldsymbol{\beta}_{\mathbf{1}} + \boldsymbol{\beta}_{\mathbf{3}} + (\boldsymbol{\beta}_{\mathbf{2}} + \boldsymbol{\beta}_{\mathbf{4}})\mathrm{Renda}_{\mathbf{i}}$$

$$\mathbf{E}(\mathrm{C}_{\mathbf{i}} \mid \mathrm{Renda}_{\mathbf{i}}, \mathrm{Genere}_{\mathbf{i}} = \mathbf{0}) = \boldsymbol{\beta}_{\mathbf{1}} + \boldsymbol{\beta}_{\mathbf{2}}\mathrm{Renda}_{\mathbf{i}}$$

Esquemes additiu, multiplicatiu i interaccions

Esquema mixt

Gràficament:



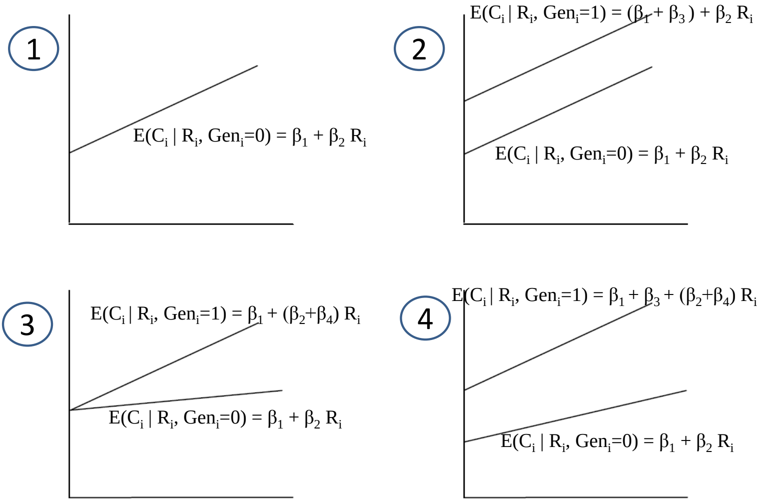
És a dir, la introducció de la variable fictícia segons un esquema mixt portaria a que tant el terme independent com el pendent del model de regressió fossin diferents entre homes i dones.

Esquemes additiu, multiplicatiu i interaccions

Arribats a aquest punt, seria lògic plantejar-nos si hi ha una diferència significativa entre homes i dones i, en cas d'existir, si aquesta afecta al terme independent i/o al pendent del model de regressió especificat.

En realitat, respondre a aquesta pregunta implica seleccionar un dels quatre models següents:

Esquemes additiu, multiplicatiu i interaccions



Esquemes additiu, multiplicatiu i interaccions

En conseqüència podrem contrastar una especificació en front d'una altra.

Model 1 vs Model 4 (Test de Chow):

$$\begin{cases} H_0 : \beta_3 = \beta_4 = 0 \\ H_A : \text{Rebuig} H_0 \end{cases}$$

Model 2 vs Model 4:

$$\begin{cases} H_0 : \beta_4 = 0 \\ H_A : \text{Rebuig} H_0 \end{cases}$$

Model 3 vs Model 4:

$$\begin{cases} H_0 : \beta_3 = 0 \\ H_A : \text{Rebuig} H_0 \end{cases}$$

Esquemes additiu, multiplicatiu i interaccions

Termes d'Interacció

En el cas que tinguem en el model més d'una variable qualitativa ens podríem plantejar si són significatius els efectes creuats entre les categories o atributs de les qualitatives.

Per reflectir aquesta possibilitat s'haurien d'introduir en el model els **termes d'interacció** com a producte entre les distintes variables fictícies que hem especificat.

Suposem que tenim:

Sexe: $S_i = \begin{cases} 1 & \text{Home} \\ 0 & \text{Dona} \end{cases}$ Nivell d'estudis: $E_i = \begin{cases} 1 & \text{Superiors} \\ 0 & \text{altre cas} \end{cases}$

Esquemes additiu, multiplicatiu i interaccions

Termes d'Interacció

I especifiquem el model:

$$Y_i = \beta_1 + \beta_2 X_i + \beta_3 S_i + \beta_4 E_i + u_i$$

Les esperances condicionades seran:

Homes:

$$E(Y_i \mid X_i, S_i = 1, E_i = 1) = \beta_1 + \beta_2 X_i + \beta_3 + \beta_4$$

$$E(Y_i \mid X_i, S_i = 1, E_i = 0) = \beta_1 + \beta_2 X_i + \beta_3$$

Dones:

$$E(Y_i \mid X_i, S_i = 0, E_i = 1) = \beta_1 + \beta_2 X_i + \beta_4$$

$$E(Y_i \mid X_i, S_i = 0, E_i = 0) = \beta_1 + \beta_2 X_i$$

Esquemes additiu, multiplicatiu i interaccions

Termes d'Interacció

En forma de taula de doble entrada:

	$E_i = 1$	$E_i = 0$	Efecte Dif.
$S_i = 1$	$\boldsymbol{\beta_1 + \beta_2 X_i + \beta_3 + \beta_4}$	$\boldsymbol{\beta_1 + \beta_2 X_i + \beta_3}$	$\boldsymbol{\beta_4}$
$S_i = 0$	$\boldsymbol{\beta_1 + \beta_2 X_i + \beta_4}$	$\boldsymbol{\beta_1 + \beta_2 X_i}$	$\boldsymbol{\beta_4}$
Efecte Dif.	$\boldsymbol{\beta_3}$	$\boldsymbol{\beta_3}$	

Amb aquesta especificació només captem les diferències entre Sexes, sigui quin sigui el Nivell d'Estudis, $\boldsymbol{\beta_3}$; i les diferències entre Estudis sigui quin sigui el Sexe, $\boldsymbol{\beta_4}$.

Amb aquesta especificació no recollim les interrelacions entre les categories de les variables qualitatives.

Esquemes additiu, multiplicatiu i interaccions

Termes d'Interacció

Si ara especifiquem el model:

$$\mathsf{Y_i} = \boldsymbol{\beta_1} + \boldsymbol{\beta_2}\mathsf{X_i} + \boldsymbol{\beta_3}\mathsf{S_i} + \boldsymbol{\beta_4}\mathsf{E_i} + \boldsymbol{\beta_5}\mathsf{S^*E_i} + \mathbf{u_i}$$

Ara la taula anterior quedarà:

	$E_i = 1$	$E_i = 0$	Efecte Dif.
$S_i = 1$	$\boldsymbol{\beta_1} + \boldsymbol{\beta_2}X_i + \boldsymbol{\beta_3} + \boldsymbol{\beta_4} + \boldsymbol{\beta_5}$	$\boldsymbol{\beta_1} + \boldsymbol{\beta_2}X_i + \boldsymbol{\beta_3}$	$\boldsymbol{\beta_4} + \boldsymbol{\beta_5}$
$S_i = 0$	$\boldsymbol{\beta_1} + \boldsymbol{\beta_2}X_i + \boldsymbol{\beta_4}$	$\boldsymbol{\beta_1} + \boldsymbol{\beta_2}X_i$	$\boldsymbol{\beta_4}$
Efecte Dif.	$\boldsymbol{\beta_3} + \boldsymbol{\beta_5}$	$\boldsymbol{\beta_3}$	

i les diferències entre Sexes dependran del Nivell d'Estudis, i les diferències entre Nivells d'Estudis dependran del Sexe.

Aplicacions de les variables fictícies

Observacions típiques

4. Aplicacions de les variables fictícies

Observacions típiques

Les observacions atípiques són un problema pel model de regressió en la mesura que poden distorsionar els resultats de l'estimació. Una opció possible quan hi han observacions atípiques és eliminar-les de la mostra, tot i que això suposa reduir la mida mostral i no considerar determinada informació.

En aquesta situació, la utilització de variables fictícies és una opció que permet mantenir l'observació atípica en la mostra però eliminant la seva influència sobre les estimacions.

Aplicacions de les variables fictícies

Observacions típiques

Suposem que en l'estimació d'un model de regressió identifiquem l'observació j com atípica. En aquest cas, podríem generar una variable fictícia $\mathrm{D_i}$ de la següent manera:

$$\mathrm{D_i} = \begin{cases} \mathbf{0} & \text{si l'observació és la número } \boldsymbol{j} \\ \mathbf{1} & \text{si l'observació no és la número } \boldsymbol{j} \end{cases}$$

Per tant, es tracta d'una variable fictícia on tots els seus valors són 0 excepte el de l'observació atípica.

Si s'introdueix aquesta variable fictícia com una variable explicativa més (esquema additiu) eliminarem la influència d'aquesta observació.

Aplicacions de les variables fictícies

Observacions típiques

L'avantatge d'introduir aquesta variable fictícia és que el residu de l'estimació per aquesta observació atípica serà zero. D'aquesta manera, el model ajustarà exactament aquesta observació, impedint que tingui influència sobre l'estimació de la resta de paràmetres i dels seus estadístics corresponents.

L'inconvenient associat a la inclusió d'aquesta variable fictícia és que els resultats obtinguts per la resta de coeficients seran els mateixos que els obtinguts en el cas d'eliminar aquesta observació de la mostra, tenint en aquest cas el model una variable explicativa més.

Aplicacions de les variables fictícies

Estacionalitat

Estacionalitat

Quan la variable endògena que es vol analitzar és de sèrie temporal amb una periodicitat inferior l'any pot ser necessària la correcció de factors estacionals, és a dir, l'eliminació del seu component estacional, amb objecte de que l'atenció se centri en els factors explicatius del model.

En aquest context, la consideració de variables fictícies en l'especificació del model permet tractar diferències en el comportament segons les estacions.

Aplicacions de les variables fictícies

Estacionalitat

Considerem, per exemple, un model explicatiu del consum **trimestral** d'un bé en funció de la renda:

$$\mathbf{C_t = \beta_1 + \beta_2 R_t + u_t \quad t = 1, 2, \ldots, T}$$

Però creiem que el consum trimestral del bé pot estar influït pel trimestre en què es realitza aquest consum.

Per analitzar aquesta possibilitat podríem incloure en el model variables fictícies. Hauríem de definir les següents variables:

$$\mathrm{D}_{4t} = \begin{cases} 1 & \text{si 4rt trim.} \\ 0 & \text{altre cas} \end{cases} \quad \mathrm{D}_{3t} = \begin{cases} 1 & \text{si 3r trim.} \\ 0 & \text{altre cas} \end{cases} \quad \mathrm{D}_{2t} = \begin{cases} 1 & \text{si 2n trim.} \\ 0 & \text{altre cas} \end{cases}$$

de manera que el primer trimestre forma la categoria base.

Aplicacions de les variables fictícies

Estacionalitat

A continuació, les introduïm de manera additiva en el model:

$$C_t = \beta_1 + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \beta_5 R_t + u_t$$

On els coeficients que acompanyen a les variables fictícies, β_2, β_3 i β_4 recullen l'**efecte estacional** diferencial en el Consum esperat respecte la categoria base, és a dir, el primer trimestre de l'any.

Alternativament, podríem haver introduït quatre variables fictícies, una per cada trimestre de l'any, i eliminar el terme constant:

$$C_t = \beta_1 D_{1t} + \beta_2 D_{2t} + \beta_3 D_{3t} + \beta_4 D_{4t} + \beta_5 R_t + u_t$$

Aplicacions de les variables fictícies

Canvi Estructural

Canvi Estructural

Les variables qualitatives també permeten fer el contrast de l'existència de canvi estructural. Un canvi estructural suposa un comportament diferent del model per diferents parts de la mostra.

Aquest fet es concreta en què tots o alguns dels paràmetres de la regressió no es mantenen fixes o constants per a tot el període mostral.

Aplicacions de les variables fictícies

Canvi Estructural

Quan les dades són temporals, els canvis estructurals acostumen a estar provocats per canvis institucionals, normatius o fets exògens no recollits en el model i que succeeixen d'una manera puntual.

En el cas de dades de tall transversal, els canvis plantegen diferents comportaments com a conseqüència de l'existència de característiques qualitatives diferents.

En aquest context, la utilització de variables fictícies permet analitzar si existeix canvi estructural.

Aplicacions de les variables fictícies

Canvi Estructural

Per verificar la permanència estructural fem servir el **Contrast de Chow**. Estimem dues regressions diferents per les dues parts de la informació mostral i contrastem la hipòtesi nul·la de que cap dels paràmetres del model varia d'una submostra a l'altra, front la hipòtesi alternativa de que tots els paràmetres varien.

Aquesta matisació és de notable interès ja que implica que del rebuig de la hipòtesi nul·la del contrast de Chow deduïm que no és cert que la constant i el pendent del model en ambdues submostres coincideixin. Però, no podem determinar si el canvi estructural afecta tan sols a la constant, tan sols al pendent o són els dos paràmetres els que canvien.

Aplicacions de les variables fictícies

Canvi Estructural

La solució a aquesta limitació es troba en la utilització de variables fictícies.

En aquest cas, el procediment per contrastar l'existència de ruptura o canvi estructural total, en la constant i en el pendent, que seria equivalent al contrast de Chow, es realitzaria a partir de la introducció de variables fictícies segons un esquema mixt, és a dir, additiu i multiplicatiu.

Però, a més, la utilització de variables fictícies permet també contrastar si el canvi s'ha produït únicament en la constant o en el pendent.

Aplicacions de les variables fictícies

Canvi Estructural

Si es vol analitzar la possibilitat d'un canvi estructural parcial en la constant s'introduirien variables fictícies additives, mentre que si se vol contrastar l'existència d'un canvi estructural parcial en el pendent s'introduirien variables fictícies multiplicatives.

Aquestes variables fictícies prendrien valor 0 per una de les submostres i valor 1 per l'altra.

Econometria
Tema 3: MRLM: Variables exògenes qualitatives

Ramon Alemany

Grau Estadística UB-UPC

Curs 2017-18