

Introdución aos modelos de datos de panel

Ana Iglesias Casal

Área de Economía Cuantitativa
Departamento Economía Cuantitativa
Facultade de Ciencias Económicas e Empresariais

unidadesdidácticas
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA



Esta obra atópase baixo unha licenza internacional Creative Commons BY-NC-ND 4.0. Calquera forma de reprodución, distribución, comunicación pública ou transformación desta obra non incluída na licenza Creative Commons BY-NC-ND 4.0 só pode ser realizada coa autorización expresa dos titulares, salvo excepción prevista pola lei. Pode acceder Vde. ao texto completo da licenza nesta ligazón: <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.gl>

© Universidade de Santiago de Compostela, 2013

Deseño e maquetación

J. M. Gairí

Edita

Edicións USC

<https://www.usc.gal/publicacions>

DOI

<https://dx.doi.org/10.153049788415876557>

MATERIA: Econometría II

TITULACIÓN: Administración e Dirección de Empresas

PROGRAMA XERAL DO CURSO

Localización da presente unidade didáctica

Bloque Temático I. Incumprimento das hipóteses do Modelo de Regresión Lineal Clásico

Unidade I. Incumprimento das hipóteses do Modelo de Regresión Lineal Clásico

Esperanza matemática non nula

Matriz de varianzas-covarianzas non escalar

Non normalidade

Regresores estocásticos

Non linealidade

Bloque Temático II. O modelo de regresión lineal xeneralizado: hipóteses, propiedades dos estimadores, inferencia e predición

Unidade II. Modelo xeneralizado

Hipóteses

Estimación Mínimo-Cuadrática Xeneralizada. Propiedades dos estimadores

O método da matriz de transformación

Inferencia e Predición

Unidade III. Autocorrelación

Exposición do problema. Causas e consecuencias sobre a estimación MCO

Autocorrelación de primeira orde

Métodos para detectar a autocorrelación

Estimación, contrastación de hipóteses e predición dun modelo con autocorrelación de primeira orde

Unidade IV. Heterocedasticidade

Exposición do problema. Causas e consecuencias sobre a estimación MCO

Métodos para detectar a heterocedasticidade

Estimación, contrastación de hipóteses e predición dun modelo con heterocedasticidade

Bloque Temático III. Modelos dinámicos uniecuacionais. Modelos de series de tempo. Introducción á cointegración

Unidade V. Modelos dinámicos

Clases de modelos con retardos nas variables

Modelos con retardos nas variables esóxenas

Modelos autorregresivos

Unidade VI. Modelos ARIMA e Introducción á Cointegración

Modelos ARIMA: definición e conceptos

Modelos ARIMA: Fases de elaboración

Introdución á Cointegración

Bloque Temático IV. Microeconometría e decisión

Unidade VII. Modelos de variable dependente cualitativa

Modelo lineal de probabilidade

Introdución aos modelos logit e probit

Unidade VIII. Introducción aos modelos de datos de panel

Exposición xeral

Modelos estáticos: Estimadores alternativos e contrastes de especificación

Modelos dinámicos

ÍNDICE

Presentación

Os obxectivos

Os principios metodolóxicos

Os contidos básicos

1. Exposición xeral
 - 1.1. Vantaxes e limitacións derivadas do uso de paneis de datos
 - 1.2. Tipos de variables explicativas
 - 1.3. Datos de panel e efectos inobservables (individuais e temporais)
2. Modelos estáticos: estimadores alternativos e contraste de especificación
 - 2.1. Efectos fixos
 - 2.1.1. Modelo de efectos individuais fixos.
 - 2.1.2. Modelo con efectos individuais e temporais fixos
 - 2.2. Efectos aleatorios
 - 2.2.1. Modelo de efectos individuais aleatorios
 - 2.2.2. Modelo de dobre-vía aleatorio
 - 2.3. Contraste de especificación
3. Modelos dinámicos

Actividades propostas

Avaliación da unidade didáctica

Bibliografía

PRESENTACIÓN

A materia na que se enmarca esta unidade didáctica corresponde ao 3º curso do grao de Administración e Dirección de Empresas. Impártese no segundo semestre e necesita dos coñecementos previos adquiridos na materia introdutoria do primeiro semestre Econometría I.

A materia de Econometría II é a continuidade da materia de Econometría I, de conceptos introdutorios, e proporciona ao estudante técnicas econométricas de dificultade media que permiten a análise cuantitativa da realidade económica e empresarial. Tamén, esboza algúns temas de gran relevancia na modelización econométrica proporcionando as ideas básicas e os conceptos necesarios para poder abordalos nun futuro con profundidade. Así, a materia comeza cunha primeira unidade de incumprimento das hipóteses do MRLNC, sinalando as consecuencias do incumprimento sobre a estimación MCO e as posibles solucións.

A unidade II engloba o modelo xeneralizado, a súa estimación, contrastación de hipótese e predición. Este modelo xorde cando se incumpren as hipóteses de incorrelación e/ou de homocedasticidade do MRLNC.

Nas unidades III e IV preséntanse por separado a autocorrelación e a heterocedaticidade, destacando as causas, as consecuencias sobre a estimación MCO e os métodos para detectar estas situacións así como os métodos de estimación, os contrastes de hipóteses e a predición.

A unidade V aborda modelos econométricos uniecuacionais que inclúen relacións non contemporáneas entre as variables que interveñen nos mesmos. A seguinte unidade incorpora a análise de series temporais univariantes, estudando as súas características evolutivas e a dependencia entre as súas observacións. Ademais, introduce o concepto de proceso estocástico, a definición de estacionariedade, e desenvolve a tipoloxía de modelos univariantes, e as distintas fases de elaboración dun modelo ARIMA. Por último, trátase de modo introdutorio o fenómeno da regresión espuria e a interpretación da cointegración entre dúas series temporais coa mesma orde de integración, así como algúns contrastes para detectala.

A unidade VII introduce sucintamente algúns modelos de variable dependente cualitativa.

Na unidade VIII, que trataremos a continuación, desenvólvese o tratamento da combinación de series temporais e datos de corte transversais. Nalgúns casos específicos, a escaseza de datos temporais invalida os procedementos clásicos de contrastación estatística polo que a incorporación da variabilidade transversal permite incorporar robustez os resultados. Esta metodoloxía é importante para o futuro profesional dos estudantes, xa que tanto no ámbito empresarial como económico é frecuente encontrarse ante este tipo de mostras. Nestes casos, esta metodoloxía permite incrementar a potencia dos contrastes e obter estatísticos con mellores propiedades asintóticas. Esta unidade tamén é susceptible de tratamento nunha materia denominada Econometría que se imparte no Máster de Economía: Organización Industrial e Mercados Financeiros.

A unidade didáctica ten asignada dúas sesións expositivas, nas que se desenvolverán os contidos teóricos, e dúas sesións interactivas onde se levarán a cabo prácticas de ordenador para a resolucións de aplicacións de datos de panel a realidade empresarial. A totalidade de horas empregadas é de 7, dúas corresponden ás sesións expositivas e 5 ás interactivas.

OS OBJECTIVOS

Os obxectivos que a presente unidade didáctica pretende que desenvolva o alumnado son:

- **obxectivo 1.** Reflexionar de forma crítica sobre as vantaxes e limitacións da aplicación de modelos econométricos a datos de panel;
- **obxectivo 2.** Comparar o modelo de efectos fixos e o modelo de efectos aleatorios, e explicar que nos conduce a considerar que as diferenzas individuais sexan aleatorias;
- **obxectivo 3.** Contrastar a existencia de efectos fixos ou efectos aleatorios e utilizar o test de Hausman para valorar se o estimador de efectos aleatorios é inconsistente;
- **obxectivo 4.** Adquirir destreza na aplicación dos métodos econométricos ao estudo científico da realidade empresarial mediante o manexo de paquetes informáticos específicos;
- **obxectivo 5.** Interpretar traballos econométricos de dificultade media;
- **obxectivo 6.** Adquirir os coñecementos necesarios para afrontar temas econométricos máis avanzados que o axuden na toma de decisións de carácter económico.

OS PRINCIPIOS METODOLÓXICOS

A metodoloxía utilizada está relacionada co carácter aplicado da materia á problemática económica e empresarial.

A presentación dos contidos farase nas sesións expositivas mediante presentacións visuais, ademais proporcionarase ao estudante os recursos adicionais no curso virtual para o correcto proceso de aprendizaxe autónoma. O curso virtual emprégase non só como repositorio de materiais senón coma unha ferramenta de interacción entre alumnado e profesorado a través de foros, mensaxería.

As clases interactivas desenvolveranse na aula de informática en grupos de alumnos máis reducidos para que o alumnado consiga unha maior comprensión dos

contidos básicos, se familiarice coas novas ferramentas informáticas e coa resolución de problemas concretos no ámbito da empresa.

Estas sesións teórico-prácticas poderán complementarse con actividades concretas que o alumnado deberá realizar de forma autónoma, nas que se desenvolveran habilidades para elaborar e presentar estudos referentes á realidade empresarial.

OS CONTIDOS BÁSICOS

1. Exposición xeral

Os paneis de datos son mostras formadas por observacións recollidas a n axentes ou unidades económicas ao longo de t instantes de tempo. Os axentes económicos dos que proveñen os datos nun panel poden ser persoas, fogares, estados, países ou empresas, segundo cada enquisa particular. Tamén nos referiremos a eles como unidades de sección cruzada ou individuos.

Podemos distinguir diferentes tipos de datos de panel dependendo da súa amplitude transversal e profundidade temporal. Así, os paneis nos que o número de axentes é moito maior có número de períodos denomínanse Paneis Micro, mentres que os paneis centrados nunha ampla dimensión temporal denomínanse Paneis Macro. No caso de contar cun panel cunha ampla dimensión tanto temporal como transversal falaríamos dun Campo Aleatorio.

É importante subliñar que, nun sentido estrito, non son datos de panel os paneis rotatorios ou a mera agregación de cortes transversais independentes —este conxunto de datos denomínanselle pseudo paneis—. Para construír un elemento verdadeiramente útil para a inferencia é necesario que a variabilidade temporal e transversal corresponda a unha mesma mostra de individuos para todas as observacións. Neste senso, no caso da análise empresarial, a gran heteroxeneidade dificulta a construción de verdadeiros paneis.

Os tamaños relativos das dúas dimensións inflúen sobre a natureza das preguntas que o/a investigador/ra pretende analizar e tamén sobre o tratamento dado o panel de datos. De tal forma que cando a dimensión de sección cruzada é superior á temporal a análise céntrase fundamentalmente na heteroxeneidade entre as unidades de sección cruzada.

Exemplos de paneis en España son:

- Central de Balances do Banco de España que recolle os balances anuais dun gran número de empresas.
- Encuesta Permanente de Consumo e Encuesta de Presupuestos Familiares, ambas elaboradas polo Instituto Nacional de Estadística (INE), que recollen información sobre a evolución dos gastos de consumo dun elevado número de familias.
- Encuesta Industrial de Empresas, tamén elaborada polo INE, proporciona unha análise da estrutura da actividade industrial en España.

1.1. Vantaxes e limitacións derivadas do uso de paneis de datos

Baseándose en Hsiao (2003) e Baltagi (2005) enumeraremos algunhas vantaxes e limitacións da utilización de datos de panel. As vantaxes dos datos de panel pódense resumir en:

1. Control da heteroxeneidade individual: os datos transversais e temporais non son capaces, por si mesmos, de controlar a heteroxeneidade inherente no comportamento dos individuos, empresas, rexións ou países, corréndose o risco de obter estimacións innegadas cando se utilizan datos dun tipo ou doutro. Referímonos á omisión de variables que conduce a nesgo nas estimacións. Porén, a través do uso dos datos de panel poden controlarse estes efectos específicos, transversais ou temporais, sexan observables ou non -xeralmente non o serán-.
2. Proporcionan datos con maior cantidade de información, con maior grao de variabilidade e con menor nivel de colinealidade entre os regresores. Tamén aumenta o número de graos de liberdade e, polo tanto, dá lugar a unha maior eficiencia nas estimacións.
3. Son un medio adecuado para estudar procesos dinámicos de axuste xa que a partir deles, se son o suficientemente longos, pódese entrever a velocidade dos axustes promovidos por cambios nas políticas económicas.
4. Axudan a identificar e medir efectos que non son detectables con datos puros de corte transversal ou de series temporais. Supoñamos que temos datos de corte transversal correspondentes a un determinado número de mulleres nos que unha variable é a participación media anual no mercado de traballo. Se unha muller participou unha media do 50%, esta cifra puido ser xerada por dúas situacións diferentes. Ou ben traballou o 50% dos anos considerados, ou ben traballou a tempo parcial durante todos os anos. As implicacións de ambas situacións son distintas e só a utilización de datos de panel permitiría discriminar ambas situacións.
5. Permiten construír e contrastar modelos máis complicados que en contextos exclusivamente de sección cruzada ou temporais. Ademais, é necesario introducir menos restricións a hora de estimar modelos de retardos distribuídos en relación cos que son necesarios cando se traballa con series temporais.
6. Evítase o nesgo que resulta cando se traballa coas variables agregadas posto que as unidades transversais dun panel de datos normalmente refírense a individuos, familias ou empresas.

Por outro lado, entre as limitacións cabe destacar as seguintes:

1. Problemas de deseño da mostra e de recollida de datos relacionados con inadecuadas taxas de cobertura, falta de resposta, frecuencia e lapso temporal, período de referencia, etc.

2. Distorsións provocadas por erros de medida que poden aparecer por respostas ambiguas, erros de memoria, respostas incorrectas deliberadas, ou por nesgo que pode introducir o propio encuestador.
3. Problemas de selección da mostra tales como non aleatoriedade, auto-selección, falta de resposta inicial ou abandono.
4. En xeral a escasa dimensión temporal que invalida algúns argumentos asintóticos, fai que a maior parte dos mesmos teña que recaer no tamaño de corte transversal.

1.2. Tipos de variables explicativas

O conxunto de variables explicativas que se poden incluír nun modelo de datos de panel son:

- Variables que varían entre as unidades transversais e no tempo (x_{it}).
- Variables que non varían no tempo senón só no corte transversal (z_i).
- Variables que só varían na dimensión temporal pero non na transversal (ω_t).

1.3. Datos de panel e efectos inobservables (individuais e temporais)

Un modelo de regresión básico para datos de corte transversal vén dado pola expresión

$$y_i = x_i' \beta + \varepsilon_i \quad (1)$$

onde $i=1,2,\dots,N$ e N é o número de unidades de sección cruzada.

Cando cada unidade transversal se observa durante T_i períodos de tempo, o modelo resultante será

$$y_{it} = x_{it}' \beta + \varepsilon_{it} \quad (2)$$

onde $i=1,\dots,N$, e $t=1,\dots,T_i$. Este é o prototipo dun modelo *estático* con datos de panel.

Un modelo de regresión dinámico simple para unha unidade económica podería expresarse como:

$$y_t = \lambda y_{t-1} + x_t' \beta + \varepsilon_t \quad (3)$$

se se observan repetidamente varias unidades, teríase un modelo

$$y_{it} = \lambda y_{i,t-1} + x_{it}' \beta + \varepsilon_{it} \quad (4)$$

que é o prototipo dun modelo de datos de panel *dinámico*.

Se $T_i = T$, é dicir, se se observa o mesmo número de períodos a todas as unidades transversais dirase que o panel de datos está completo, e falarase de paneis de datos

equilibrados (*balanced*) cando o número de observacións transversais é o mesmo para cada período de tempo ($N_j=N$); noutro caso dirase que o panel é incompleto ou non equilibrado (*unbalanced*).

Para controlar a presenza de efectos inobservables individuais suponse que $\varepsilon_{it} = \alpha_i + v_{it}$, onde α_i recolle a heteroxeneidade transversal persistente non observada e v_{it} representa o termo da perturbación. Segundo se asuma que o efecto α_i é un parámetro fixo ou unha variable aleatoria terase o *modelo de efectos fixos* ou o *modelo de efectos aleatorios*.

Cando tamén existen efectos temporais persistentes e non observados, considérase a descomposición $\varepsilon_{it} = \alpha_i + \delta_t + v_{it}$, onde agora δ_t representa os efectos temporais inobservables, específicos de cada período, e non incluídos explicitamente entre os regresores.

Un concepto importante no contexto de datos de panel é o que se denomina heteroxeneidade e, asociada a ela, nesgo de heteroxeneidade. O principal obxectivo perseguido coa utilización de datos de panel é analizar como varía o comportamento dos axentes ou unidades económicas individuais en función de determinadas características socioeconómicas incluídas no modelo especificado como variables causais. Porén, non todos os axentes ou unidades comparten a mesma forma de actuación fronte a unhas mesmas variables causais. É dicir, o ignorar efectos específicos de individuos e de tempo que existen entre as unidades de tempo ou de corte transversal, e que non se capturan coas variables incluídas no modelo, pode conducirnos á presenza de heteroxeneidade dos parámetros do modelo.

Por esta razón, faise necesario considerar a existencia de efectos latentes non observables específicos de cada axente ou unidade, que son xeralmente constantes no tempo, e que inciden na súa toma de decisións. Se estes efectos latentes existen e non se recollen explicitamente no modelo producirase un problema de variables omitidas.

2. Modelos estáticos: estimadores alternativos e contraste de especificación

Un modelo de datos de panel podería formularse en termos plenamente xenéricos como:

$$y_{it} = \mu_{it} + \beta_{1it}x_{1it} + \beta_{2it}x_{2it} + \beta_{3it}x_{3it} + \dots + \beta_{kit}x_{kit} + v_{it} \quad (5)$$

onde $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$; permitindo a presenza de múltiples parámetros individuais e temporais e unha definición sen restricións sobre a composición e propiedades do vector de perturbacións aleatorias.

Porén, tal representación non resulta viable polo que, xeralmente recórrase a formulacións máis restritivas tanto en termos paramétricos como con relación a os supostos sobre o vector de perturbacións. Seguindo a Johnston (1989) podemos ordenar 7 tipos de especificacións nunha taxonomía sobre os modelos de datos de panel:

Táboa 1: Taxonomía de modelos de datos de panel

Modelo	Ordenada na orixe	Coefficientes da pendente	Vector de perturbacións
I	Común para todo «i» e «t»	Común para todo «i» e «t»	Matriz de varianzas-covarianzas escalar
II	Común para todo «i» e «t»	Común para todo «i» e «t»	Matriz de varianzas-covarianzas non escalar
III	Variando sobre «i»	Común para todo «i» e «t»	Efectos Fixos
IV	Variando sobre «i»	Común para todo «i» e «t»	Efectos Aleatorios
V	Variando sobre «i» e «t»	Común para todo «i» e «t»	Efectos Fixos
VI	Variando sobre «i» e «t»	Común para todo «i» e «t»	Efectos Aleatorios
VII	Variando sobre «i»	Variando sobre «t»	Matriz de varianzas-covarianzas escalar ou non

De todos os modelos presentados na táboa 1, dous deles son os máis utilizados: III Modelo de Efectos Fixos e IV Modelo de Efectos Aleatorios.

A especificación xeral dun modelo estático de datos de panel é:

$$y_{it} = x'_{it}\beta + \alpha_i + \lambda_t + v_{it} \quad (6)$$

onde $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$

Nestes modelos, o comportamento do axente i no momento t (y_{it}) é explicado por unha serie de variables explicativas (x_{it}). Ademais, existen efectos non observables específicos de cada axente e invariantes o longo do tempo (α_i) que pódense considerar sempre aleatorios sen perda de xeneralidade, e efectos temporais non observables invariantes entre individuos que non están incluídos na regresión como variables explicativas (λ_t). A estes efectos denomínanse *variables latentes*, e a eles non se lles asocia ningún parámetro, dado que ao non ser observables non permiten que sexan identificados por separado parámetro e variable.

Aínda que non aparecen na ecuación anterior por simplicidade, tamén poden incluírse como regresores: variables esóxenas que afectan de maneira específica a cada axente e que permanecen invariables ao longo do tempo (z_i), e variables que evolucionando no tempo afectan de igual maneira a todos os individuos do panel (w_t).

A distinción crucial entre os enfoques utilizados para datos de panel reside nos supostos que se asumen acerca deses factores específicos de cada axente ou individuo. Eses supostos fan referencia a se os efectos específicos están correlacionados ou non coas variables explicativas (x_{it}). De tal forma, que se α_i ou λ_t está correlacionado coas variables explicativas pode ser conveniente facer inferencia condicional sobre as realización dos α_i ou λ_t na mostra —efectos fixos— (a distribución da variable endóxena condiciónase ao valor de ditos parámetros), mentres que se os α_i ou

λ_t non están correlacionados coas variables explicativas é natural facer inferencia incondicional —efectos aleatorios—.

O enfoque de efectos fixos é razoable cando se confía en que as diferenzas entre unidades poden ser recollidas a través de diferentes ordenadas no orixe. Ademais, é a especificación adecuada cando nos centramos nun conxunto específico de unidades económicas e a nosa inferencia está restrinxida ó comportamento desas unidades económicas. Polo tanto, pode ser aplicado no caso de que as conclusións se refiran unicamente ás unidades incluídas na mostra e non se queira xeneralizar esas conclusións a unidades fora da mostra, é dicir cando a mostra inclúe a todas as unidades de decisión do problema analizado.

Noutros casos pode ser máis apropiado considerar os termos específicos como aleatoriamente distribuídos entre as unidades económicas. Isto sería o máis adecuado se consideramos que as unidades económicas da mostra foron tomadas dunha poboación grande, en cuxo caso sería aplicable o enfoque de efectos aleatorios.

No modelo de efectos aleatorios as variables latentes α_i ou λ_t , ao non ser observables, pasan a formar parte dun termo de perturbación composto do modelo (ε_{it}), e esta é a razón pola que se denominan tamén modelos de erros compostos. É dicir:

$$y_{it} = x'_{it}\beta + (\alpha_i + \lambda_t + \nu_{it}) = x'_{it}\beta + \varepsilon_{it} \quad (7)$$

En notación matricial

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \beta + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

onde cada submatriz ou subvector contén T observacións temporais e representa á correspondente unidade económica das N que forman parte da mostra.

2.1. Efectos fixos

2.1.1. Modelo de efectos individuais fixos.

Imos examinar, en primeiro lugar, o caso máis verosímil no que existen correlacións entre as variables latentes α_i , aleatorias e non observables, e as demais variables explicativas. Posto que o problema da inconsistencia está xerado pola presenza dos efectos individuais, cabe preguntarse se existe algunha transformación do modelo que elimine tales efectos. En realidade, existen dúas transformacións diferentes con esta propiedade mediante as cales podemos obter estimadores consistentes, estes son os estimadores intragrupos (*within*) e en primeiras diferenzas que se expoñen a continuación.

O estimador intragrupos (*within*) baséase en que se se promedia a ecuación

$$y_{it} = x'_{it}\beta + \alpha_i + v_{it} \quad (8)$$

para cada individuo, obtense a ecuación

$$\bar{y}_i = \bar{x}'_i\beta + \alpha_i + \bar{v}_i \quad (9)$$

e restando á ecuación (8) a (9) prodúcese a expresión

$$y_{it} - \bar{y}_i = (x'_{it} - \bar{x}'_i)\beta + (v_{it} - \bar{v}_i) \quad (10)$$

Aplicando MCO (Mínimos cadrados ordinarios) a esta ecuación obtéñense estimacións do vector β , denominándose estimadores de efectos fixos ($\hat{\beta}_{EF}$) ou intragrupos ($\hat{\beta}_{IG}$).

Se as variables x_{it} son estritamente esóxenas —modelo estático—, o estimador MCO será un estimador consistente de β . O requisito de que as variables x_{it} sexan esóxenas, e non só predeterminadas, para garantir a consistencia do estimador débese á presenza da media mostral das perturbacións do modelo orixinal ($\bar{v}_i$) no termo de perturbación do modelo transformado (ecuación 10).

Como se desprende das anteriores expresións, o estimador intragrupos utiliza unicamente a variación que se produce entre as observacións procedentes de cada individuo, pero non a través de todo o panel de datos. Por non utilizar toda a información mostral, o estimador intragrupos non é, en xeral, eficiente; con todo si é eficiente cando eses efectos específicos son deterministas.

Un procedemento alternativo para obter os estimadores intragrupos é considerar os α_i constantes e aplicar o método mínimo cuadrático con variables ficticias á ecuación 8. Isto consiste en estimar por MCO introducindo N variables ficticias, resultando a seguinte ecuación

$$y_{it} = i\alpha_i + x'_{it}\beta + v_{it} \quad (11)$$

En notación matricial

$$\begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} i & 0 & \dots & 0 \\ 0 & i & \dots & 0 \\ . & . & \dots & . \\ 0 & 0 & \dots & i \end{bmatrix} \alpha_i + \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix} \beta + \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{bmatrix}$$

sendo i un vector duns de orden $T \times 1$.

O estimador en primeiras diferenzas baséase no feito de que aplicando o operador de primeiras diferenzas na expresión (8) obtense a ecuación

$$\Delta y_{it} = (\Delta x_{it})' \beta + \Delta v_{it} \quad (12)$$

onde se eliminou a variable latente. Esta transformación non se pode aplicar se as variables x_{it} son constantes ao longo do tempo, xa que ditas variables desaparecerían

da formulación en primeiras diferenzas, non podendo estimarse os seus coeficientes. Neste caso, a nova perturbación aleatoria do modelo transformado non estará incorrelacionada, salvo cando a perturbación orixinal fose un paseo aleatorio. A pesar diso, o estimador MCO continúa sendo un estimador consistente dos elementos do vector b . Cando a perturbación no modelo transformado xa non é un ruído branco, entón é recomendable a utilización do estimador MCX (Mínimo cuadrático Xeneralizado), que coincide numericamente co estimador intragrupos.

2.1.2. Modelo con efectos individuais e temporais fixos

O modelo anterior pódese estender a un modelo de efectos fixos de «dobre-vía» (*two-way*) no que aparecen tamén efectos inobservables temporais.

$$y_{it} = x'_{it}\beta + \alpha_i + \lambda_t + v_{it} \quad (13)$$

Isto equivalería a introducir dos conxuntos de variables ficticias, unhas individuais e outras temporais, e en principio o estimador MCO tería as mesmas propiedades que no modelo anterior. Para obter o estimador de b habería que realizar unha transformación intragrupos (*within*) similar ó caso anterior, efectuando a regresión de $y_{it} - \bar{y}_{i..} - \bar{y}_{.t} + \bar{y}_{...}$ sobre $x_{it} - \bar{x}_{i..} - \bar{x}_{.t} + \bar{x}_{...}$. Sendo $\bar{y}_{i..}$ a media temporal, $\bar{y}_{.t}$ a media por individuos e $\bar{y}_{...}$ a media total.

2.2. Efectos aleatorios

2.2.1. Modelo de efectos individuais aleatorios

Cando non existe correlación entre as variables latentes e as demais variables explicativas a estimación por MCO en niveis proporciona estimadores consistentes. Con todo, mesmo no caso máis restritivo de que v_{it} sexa homocedástica e non presente autocorrelación, a perturbación aleatoria do modelo de erros compostos ε_{it} si presenta autocorrelación. Isto suxire que se pode obter un estimador máis eficiente que o MCO utilizando MCX (Mínimos cadrados Xeneralizados), e a ese estimador MCX coñécese co nome de estimador de Balestra-Nervole $\hat{\beta}_{BN}$.

Sexa o modelo:

$$y_{it} = x'_{it}\beta + \varepsilon_{it} \quad (14)$$

onde $\varepsilon_{it} = \alpha_i + v_{it}$ con $i = 1, 2, \dots, N$ e $t = 1, 2, \dots, T$, no que se verifican as seguintes hipóteses:

1. $E v_{it} = E \alpha_i = E(v_{it}, \alpha_i) = 0$ para todo i, t .
2. $E(x'_{it} v_{it}) = E(x'_{it} \alpha_i) = 0_k$ para todo i, t .
3. $E \alpha_i \alpha_j = \sigma_\alpha^2$ si $i = j$, e igual a cero en caso contrario.
4. $E v_{it} v_{jt} = \sigma_v^2$ si $i = j$, e igual a cero en caso contrario.

É útil neste caso ver a formulación do modelo en bloques de T observacións para cada unha das unidades atemporais i . Para esas T observacións teremos que V_i será: $V_i = E(\epsilon_i' \epsilon_i) = \sigma_v^2 I_T + \sigma_\alpha^2 i i'$ sendo i' un vector fila duns de orde $1 \times T$.

En notación matricial

$$V_i = \begin{bmatrix} \sigma_v^2 + \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_v^2 + \sigma_\alpha^2 & \cdots & \sigma_\alpha^2 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \cdots & \sigma_v^2 + \sigma_\alpha^2 \end{bmatrix}$$

Pódese observar como se detecta a presenza de autocorrelación entre as perturbacións aleatorias de cada individuo (ϵ_i) debida á correlación entre as variables non observables que son incorporadas na perturbación aleatoria.

A matriz V_i é a mesma para cada individuo da mostra, polo que a matriz de varianzas covarianzas V pódese expresar como $V = I_N \otimes V_i$, xa que as perturbacións entre unidades transversais non están correlacionadas.

En notación matricial

$$V = \begin{bmatrix} V_1 & 0 & \cdots & 0 \\ 0 & V_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & V_N \end{bmatrix}$$

Os estimadores MCX son lineais, innesgados e óptimos (ELIO) dos parámetros do modelo, e calcúlanse como:

$$\hat{\beta} = \left[\sum_{i=1}^N x_i' V_i^{-1} x_i \right]^{-1} \sum_{i=1}^N x_i' V_i^{-1} y_i$$

sendo a matriz de covarianzas do estimador igual a:

$$\left[\sum_{i=1}^N x_i' V_i^{-1} x_i \right]^{-1}$$

Estimar por MCX este modelo é equivalente a efectuar a estimación das variables transformadas da forma seguinte:

$$y_{it} - \theta \bar{y}_i \text{ e igual coas variables explicativas, sendo } \theta = 1 - \sqrt{\frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\alpha^2}}$$

Un procedemento alternativo para obter os estimadores MCX é a través dos estimadores intragrupos e entregrupos, xa que os estimadores MCX son unha media ponderada matricial deses dous estimadores.

Por iso imos definir os estimadores entregupos (between), que se obteñen a partir da aplicación de MCO ao modelo que relaciona a media aritmética das observacións de cada grupo da mostra. É dicir, consiste en estimar por MCO a seguinte relación:

$$\bar{y}_i = \bar{x}_i' \beta + (\alpha_i + \bar{v}_i) \quad (15)$$

Esta estimación só será consistente cando os efectos individuais non observables estean incorrelacionados coas variables explicativas.

O estimador MCX obtense como

$$\hat{\beta} = \hat{F}^I b^I + \hat{F}^E b^E$$

onde

$$\hat{F}^E = I - \hat{F}^I$$

$$\hat{F}^I = [S_{XX}^I + \lambda \quad S_{XX}^E]^{-1} S_{XX}^I \quad \lambda = \frac{\sigma_v^2}{\sigma_v^2 + T\sigma_\alpha^2} = (1 - \theta)^2$$

sendo $\hat{F}^E$ a matriz do estimador *entregupos*, $\hat{F}^I$ a matriz do estimador *intragrupos*, b^E o estimador *entregupos*, b^I o estimador *intragrupos*, S_{XX}^I e S_{XX}^E as matrices de momentos das sumas de cadrados dentro de cada grupo e entre grupos respectivamente.

$$S_{XX}^I = \sum_i \sum_t (x_{it} - \bar{x}_i)(x_{it} - \bar{x}_i)' \quad S_{XY}^I = \sum_i \sum_t (x_{it} - \bar{x}_i)(y_{it} - \bar{y}_i)$$

$$S_{XX}^E = \sum_i T(\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})' \quad S_{XY}^E = \sum_i T(\bar{x}_i - \bar{x})(\bar{y}_i - \bar{y})$$

$$b^I = (S_{XX}^I)^{-1} S_{XY}^I \quad b^E = (S_{XX}^E)^{-1} S_{XY}^E$$

Na práctica, tense que traballar co estimador MCXF (Mínimos cuadrático Xeneralizado Factible). Xeralmente utilízase como estimación da varianza σ_v^2 a varianza residual obtida do estimador within multiplicada polo factor $(NT-k)/(NT-N-k)$,

e como estimación da varianza σ_α^2 o $\hat{\sigma}_v^2 - \frac{\hat{\sigma}_b^2}{T}$ valor onde $\hat{\sigma}_b^2$ é a varianza residual

estimada do estimador *between*. O estimador MCXF resultante será asintoticamente o mellor estimador lineal e consistente, e distribuirase normalmente.

As estimacións en niveis (MCO ou MCX) e as estimacións en desviacións adoitan proporcionar resultados moi distintos nas aplicacións prácticas. Isto é, en xeral, unha indicación de que hai diferenzas inobservables entre individuos que nesgan as estimacións en niveis. Por outra banda, en paneis nos que a dimensión de corte transversal é grande e a dimensión temporal é pequena as estimacións intragrupos son moito máis imprecisas que as estimacións en niveis.

2.2.2. Modelo de dobre-vía aleatorio

Nun modelo de dobre vía

$$y_{it} = x'_{it}\beta + (\alpha_i + \lambda_t + \nu_{it}) = x'_{it}\beta + \varepsilon_{it} \quad (16)$$

pódese considerar os efectos λ_t como aleatorios e estimar o devandito modelo aplicando MCXF. Porén, é moi usual, sobre todo en modelos para datos de panel non lineais, considerar os efectos individuais como aleatorios e os temporais como efectos fixos, estimándose estes últimos mediante a introdución das correspondentes variables ficticias.

En resumo, podemos distinguir dous enfoques:

- Enfoque de efectos fixos: o efecto individual está correlacionado coas variables explicativas (x_{it}), neste caso os estimadores MCO son inconsistentes;
- Enfoque de efectos aleatorios: como non existe tal correlación non se produce o problema de inconsistencia o estimar por MCO, pero pódese tratar de gañar eficiencia na estimación dos parámetros a través da consideración da estrutura da matriz de varianzas-covarianzas da nova perturbación aleatoria (ε_{it}), é dicir aplicando MCX no caso de que exista heterocedasticidade ou autocorrelación.

2.3. Contraste de especificación

Unha vez abordados os diferentes tipos de modelos resulta pertinente preguntarse que modelo debe utilizarse. A elección debe basearse sempre na xustificación dos enfoques previamente mencionada. Só cando o/a investigador/ra non ten clara a relación entre α_i e x_{it} pódese utilizar o contraste de Hausman para decidir entre o modelo de efectos fixos ou o de efectos aleatorios.

No contraste de especificación de Hausman a hipótese nula supón a incorrelación entre as variables latentes e as demais variables incluídas no modelo ($H_0: E(\alpha_i / x_{i1}, \dots, x_{iT}) = 0$). Este test compara directamente os dous estimadores, o de efectos fixos ($\hat{\beta}_{IG}$) e o de efectos aleatorios ($\hat{\beta}_{BN}$), e o estatístico proposto presenta a seguinte forma,

$$m = \hat{q}' [\text{Var}(\hat{q})]^{-1} \hat{q}$$

onde, $\hat{q} = \hat{\beta}_{IG} - \hat{\beta}_{BN}$ e $\text{Var}(\hat{q}) = \text{Var}(\hat{\beta}_{IG}) - \text{Var}(\hat{\beta}_{BN})$.

O estatístico m distribúese como unha χ^2 con $k-1$ graos de liberdade no caso de que a H_0 sexa certa. Sendo $\hat{\beta}_{BN}$ o estimador de Balestra-Nerlove (efectos aleatorios), que é consistente e ten varianza mínima cando H_0 é certa, e $\hat{\beta}_G$ o estimador intragrupos (efectos fixos), o cal é consistente con independencia de que se verifique ou non a H_0 .

Por tanto, a regra de decisión derívase do feito de que ambos os dous estimadores tenderán a coincidir cando se verifique a H_0 e a diferir se se incumpre.

Xa que se a hipótese nula é certa o estimador de Balestra e Nerlove $\hat{\beta}_{BN}$ é asintoticamente máis eficiente que o estimador intragrupos $\hat{\beta}_{IG}$. Porén, se a hipótese nula non é certa o estimador intragrupos $\hat{\beta}_{IG}$ manterá a consistencia, pero o estimador de Balestra e Nerlove $\hat{\beta}_{BN}$ será nesgado e inconsistente.

3. Modelos dinámicos

O modelo de datos de panel dinámico básico pódese expresar

$$y_{it} = \lambda y_{i,t-1} + x'_{it}\beta + \varepsilon_{it} \quad (17)$$

onde $\varepsilon_{it} = \alpha_i + v_{it}$

Todos os estimadores vistos ata agora para o caso de modelos estáticos son inconsistentes e polo tanto tense que recorrer á obtención de novos estimadores aplicando métodos de variables instrumentais.

ACTIVIDADES PROPOSTAS

Actividades do alumnado fóra da aula

- Localización dun artigo ou traballo de investigación que utilice esta metodoloxía, para proceder a lectura comprensiva do mesmo. Entrega posterior nunha cara dun folio das seguintes cuestións: título do artigo, especificación do modelo, variable dependente do modelo, tipo de mostra utilizada (unidades atemporais e período temporal), tipos de variables explicativas, método de estimación e principais conclusións obtidas.

Actividades do alumnado na aula

- Realización dunha práctica sobre datos de panel efectuando tanto a estimación como os contrastes de especificación mediante o manexo do programa econométrico Eviews. Os datos utilizados pódense obter a partir de manuais recomendados na bibliografía.

AVALIACIÓN DA UNIDADE DIDÁCTICA

A avaliación desta unidade didáctica constará:

- Participación activa nas clases e resolución da actividade proposta. Constitúe o 5% da nota da materia na que se valorará por unha parte as intervencións realizadas na aula e por outra a contribución na resolución da actividade levaba a cabo na presente unidade. Para valorar esta contribución, será fundamental a calidade da presentación, a idoneidade das argumentacións e da resolución.
- Resolución dunha proba final da materia na que a parte correspondente o bloque temático IV, que inclúe esta unidade didáctica, supoña o 5% da nota total.

BIBLIOGRAFÍA

- ARELLANO, Manuel e Olympia BOVER(1990): «La econometría de los datos de panel», *Investigaciones Económicas*, Fundación SEPI, vol. 14(1), 3-45.
- ARELLANO, Manuel. (2003). *Panel Data Econometrics*. Oxford University Press, Oxford.
- BALTAGI, Badi. H. (2005). *Econometric Analysis of Panel Data*, 3rd. Ed. John Wiley & Sons Ltd. West Sussex, England.
- GREEN, William. H. (1999): *Análisis Económico*, 3^a. Ed. Prentice Hall Iberia, Madrid.
- GUISÁN, M. Carmen.(1997): *Econometría*, Ed. McGraw-Hill, Madrid.
- HILL, Carter.R ; William .E.,GRIFFITHS e Guay. C.,LIM (2008). *Principles of Econometrics*, 3rd. Ed. John Wiley & Sons, Hoboken.
- HSIAO, Cheng. (2003). *Analysis of Panel Data*, 2nd. Ed. Cambridge University Press, Cambridge.
- JOHNSTON, J. (1989). *Métodos de Econometría*. Ed. Vicens Vivens, Barcelona.
- WOOLDRIDGE, Jeffrey. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. The MIT Press Cambridge, Massachusetts.

Recursos en internet*Instituto Nacional de Estadística*

- Encuesta de Presupuestos Familiares:
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=%2Ft25%2Fp458&file=inebase&L=0> [citado 17 xullo 2013]
- Encuesta Industrial de Empresas:
<http://www.ine.es/jaxi/menu.do?type=pcaxis&path=/t05/p048/&file=inebase> [citado 17 xullo 2013]

Banco de España

- Central de Balances:
<http://www.bde.es/bde/es/areas/cenbal/> [citado 17 xullo 2013]



Unha colección orientada a editar materiais docentes de calidade e pensada para apoiar o traballo do profesorado e do alumnado de todas as materias e titulacións da universidade

unidadesdidácticas
UNIVERSIDADE DE SANTIAGO DE COMPOSTELA