

Glosario básico de econometría

Publicación colaborativa

Autores/as: Alumnado Econometría II
Coordinadora: Xulia Guntín Araujo



Glosario básico de econometría

PRÓLOGO

Este glosario é o resultado do traballo colaborativo dun grupo de alumnas e alumnos da materia de Econometría II, realizado baixo a miña titorización, durante o curso 2019-2020.

Forma parte do proxecto de innovación educativa “Aprendizaxe creando e compartindo contidos para a análise cuantitativa no ámbito socioeconómico, a través do Campus Virtual” do grupo de innovación docente “Mellora da aprendizaxe da análise de datos para a Economía e a Empresa”, formado por docentes do Departamento de Economía Cuantitativa da USC. Dito proxecto foi seleccionado na *II Convocatoria de proxectos de innovación educativa en MOOC e de apoio das TIC a docencia* da USC.

O proxecto contemplaba a realización dun “glosario” da materia que debería ter a calidade suficiente para converterse nun recurso docente que puidera ser posto a disposición do alumnado nos cursos seguintes. Non é realmente un glosario ao uso, dado que alumnado tiña liberdade para decidir que incorporaba, coa única restrición de que debían ser termos empregados especificamente na materia de Econometría. Podían definir, explicar, apoiarse nun exemplo, nunha gráfica etc. Buscábase incentivar o traballo rigoroso e de calidade por parte do alumnado e cremos que este obxectivo está conseguido.

O documento foi revisado polas profesoras do Departamento de Economía Cuantitativa, M.^a Carmen López Andión e Ana Iglesias Casal, a quen lle solicitamos a súa lectura para detectar posibles erros que deberan ser subsanados, pero respectando o texto e a selección de entradas feitas polo alumnado. Agradécémolle o seu traballo. Calquera erro é, en todo caso, unicamente responsabilidade miña. Teño como tarefa, implicar a novo alumnado na súa mellora e ampliación.

O desenvolvemento do proxecto viuse afectado pola suspensión da docencia presencial a causa da COVID 19, pero aínda con todas as dificultades conseguimos levar a bo fin este glosario.

Relación de autores/as

Acevedo Iglesias, Sergio	González Castelo, Juan	Marín Lanzarotti, Ana Karina
Castro Saavedra, Jorge	González Hermo,	Mariño Tojo, Juan Ramón
Domínguez Rodríguez, Moisés	María Margarita	Martínez Domínguez, Ariane
Fonte Couce, Rebeca	González Rueda, Manuel	Martínez Mariño, Jessica
Fraga Carlín, Luís	Graña Pérez, Álvaro	Meijón Corujo, Guillermo
Fraga Neira, Alicia	Grela Beres, David	Méndez González, Carmen
García Barreiro, Lucia	Gutiérrez Castiñeiras, Diego	Mosquera Sánchez, Lucía
García Goyanes, Cristina	Iglesias Blanco, Raquel	Muñiz Ayaso, Aitana
García Rodeiro, Carlota	Iglesias Silva, Óscar	Novoa Otero, Laura
García-Ganges Icaza, Soledad	Lago Oliveira, Iván	Otero Carballal, Inés
Gens Noya, Antonio	Lijó Lampón, Diego	Otero Rodríguez, Patricia
Gil López, Diego	Lojo Dieste, Sara	Pena Boado, Carlos
Gómez Campos, Clara	López López, Sandra	Sánchez Cerneira, Adrián
Gómez Pérez, Eva	Maril Carregal, Yago	Valcárcel Lubián, Borja

Xulia Guntín Araujo
Profesora e responsable do Proxecto

- Autocorrelación / 11
Bondade do axuste / 11
Cambios de escala / 11
Ceteris Paribus / 12
Coeficiente de correlación de Pearson / 12
Coeficiente de desigualdade U de Theil / 12
Coeficiente de determinación / 13
Coeficiente de determinación axustado / 13
Coeficientes de reacción / 14
Cointegración / 14
Consistencia / 14
Contraste de hipóteses / 15
Contraste t de significación individual / 15
Correlación contemporánea / 15
Correlograma / 16
Covarianza / 16
Criterio de información de Akaike (AIC) / 16
Criterio de información de Schwarz (SIC) / 17
Criterio de información bayesiano (véxase Criterio de información de Schwarz) / 17
Criterio de información Hannan-Quinn (HQC) / 17
Datos de panel / 17
Diagrama de predición / 17
Distribución normal / 18
Econometría / 19
Econometría Financeira / 19
Ecuación de comportamento / 19
Erro de especificación / 20
Erro de predición / 20
Erro estándar da regresión / 20
Esperanza matemática da perturbación non nula / 21
Esperanza matemática da perturbación nula / 21
Estatístico de proba / 21
Estatístico F / 21
Estimado lineal / 22
Estimador de máxima verosimellanza (EMV) / 22
Estimador inesgado / 22
Estimador óptimo / 23
Estrutura temporal do retardo / 23
Función de autocorrelación (FAC) / 24
Función de autocorrelación parcial (FACP) / 24
Heterocedasticidade / 24
Hipótese alternativa / 25
Hipótese de rango pleno / 25
Hipótese nula / 26
Homocedasticidade / 26
Identificación (Metodoloxía ARIMA) / 26
Incorrelación contemporánea entre X e ε / 27
Intervalo de confianza / 27
Matriz de varianzas-covarianzas / 27
Método da matriz de transformación / 27
Método de mínimos cadrados ordinarios / 28
Método de mínimos cadrados xeralizados / 28
Método de variables instrumentais / 28
Microeconometría / 27
Modelo autorregresivo de medias móbiles (ARMA) / 29
Modelo autorregresivo integrado de media móbil (ARIMA) / 29
Modelo de axuste parcial / 29
Modelo de elección binaria / 30
Modelo de probabilidade lineal / 30
Modelo de regresión / 30
Modelo de regresión lineal clásico / 30
Modelo de regresión lineal xeralizado / 30
Modelo de retardos distribuídos / 31
Modelo de retardos xeométricos / 31
Modelo de Koyck (véxase Modelo de retardos xeométricos) / 32
Modelo econométrico / 32
Modelo econométrico autorregresivo / 32
Modelo econométrico dinámico / 33
Modelo econométrico estático / 33
Modelo econométrico lineal-logarítmico / 33
Modelo econométrico logarítmico-lineal / 33
Modelo econométrico Log-Log / 34
Modelo econométrico multiecuacional / 35
Modelo econométrico non lineal / 35
Modelo econométrico uniecuacional / 35
Multicolinealidade / 35
Multiplicador de impacto / 36
Nesgo dun estimador / 36
Nivel de confianza / 36
Operador de retardos / 36
Operadores nas ecuacións ARIMA / 37
Orde de integración / 37
Ordenada na orixe / 37
P-Valor / 37
Parámetro / 38
Paseo aleatorio / 38
Perturbación aleatoria / 38
Perturbación aleatoria AR(1) / 38
Polinomio de retardo distribuídos de Almon / 39
Predición / 39
Proceso autorregresivo de orde p , AR(p) / 39
Proceso ergódico / 40
Proceso estocástico / 40
Proceso estocástico estacionario / 40
Proceso integrado / 40
Pseudo R^2 de Mc Fadden / 41
Raíz do erro cadrático medio (RECM) / 41
Raíz unitaria / 41

Regresión espuria	/ 42	Test de Dickey-Fuller aumentado (ADF)	/ 46
Regresor significativo	/ 42	Test de Durbin-Watson	/ 47
Regresores estocásticos	/ 43	Test de Goldfeld e Quandt	/ 47
Regresores non estocásticos	/ 43	Test de Jarque-Bera	/ 48
Retardo temporal	/ 43	Test de raíces unitarias	/ 48
Ruído branco	/ 43	Test de White	/ 49
Serie temporal	/ 44	Test Q* de Ljung e Box	/ 49
Serie temporal non estacionaria	/ 44	Trampa das variables ficticias	/ 49
Suma de cadrado dos erros (SCE)	/ 45	Variable aleatoria	/ 50
Tendencia	/ 45	Variable endóxena	/ 50
Teorema de Gauss-Markov	/ 45	Variable esóxena	/ 50
Test da razón de verosimellanza para heterocedasticidade	/ 45	Variable ficticia	/ 51
Test de Breusch-Godfrey	/ 46	Variable retardada	/ 51
Test de Dickey-Fuller (DF)	/ 45	Varianza	/ 51

Glosario

Autocorrelación

Término en inglés: **Autocorrelation**

Nos modelos econométricos, a autocorrelación refírese á existencia de correlación entre as perturbacións correspondentes a distintas observacións.

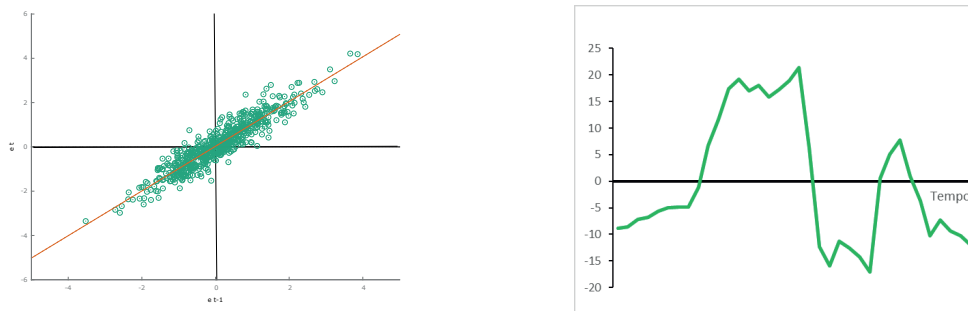
Formalmente:

$$\text{Cov}(\varepsilon_t, \varepsilon_s) \neq 0 \quad \forall t \neq s$$

É dicir, existe autocorrelación cando a covarianza entre dúas perturbacións aleatorias, referidas a distintos momentos do tempo ou distintas unidades económicas é non nula.

Exemplo: Se observamos o consumo dunha familia nun ano é probable que estea correlacionado co consumo da mesma familia no ano anterior, xa que esperaríamos que unha familia non cambiara moito o seu consumo dun ano a outro. Os factores non incluídos explicitamente na ecuación ao ter unha influencia que se mantén durante varios períodos faría que a perturbación fora autocorrelacionada nos distintos períodos.

Gráfico 1. Perturbación autocorrelacionada



Fonte da imaxe: Elaboración propia

Bondade do axuste

Término en inglés: **Goodness-of-fit**

A bondade do axuste nun modelo econométrico intenta medir o ben que se axusta (ou non) o modelo estimado ao conxunto de observacións mostrais. As medidas de bondade en termos xerais resumen as diferenzas entre os valores observados de Y e os estimados co modelo.

Unha das medidas máis utilizadas é o coeficiente de determinación (R^2).

Cambios de escala

Término en inglés: **Data scaling**

Nos modelo econométricos, un cambio de escala refírese a un cambio nas unidades de medida das variables que aparecen no modelo (por exemplo, cambiar de € a \$). Este cambio efectúase multiplicando os datos por unha determinada constante.

- Esta técnica utilízase se efectuamos a estimación dun modelo tomando determinadas unidades de medida pero queremos expresar os resultados da estimación nunhas unidades de medida diferentes.
- O cambio de escala afecta de forma diferente aos distintos resultados da estimación dun modelo, dependendo de se o cambio de escala se produce na variable dependente ou nas variables independentes.

Ceteris Paribus

Término en inglés: **Ceteris Paribus**

Nos modelos econométricos nos que se inclúen múltiples regresores, a interpretación dos coeficientes estimados realízase como “se todos os demais factores relevantes permaneceran constantes”. A isto refírese a expresión que provén do latín “ceteris paribus”.

Coefficiente de correlación de Pearson

Término en inglés: **Pearson Correlation Coefficient**

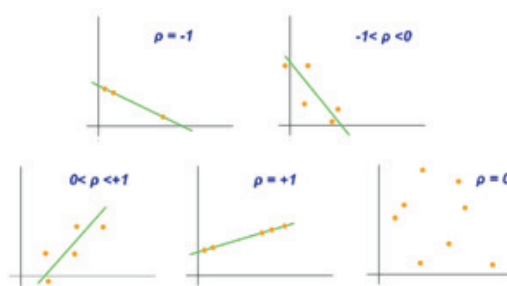
É una medida do grado de asociación lineal entre dúas variables cuantitativas. É un número real comprendido entre -1 e 1. O valor do coeficiente mide a forza da relación, se o valor é próximo a 0 diremos que a relación é débil, pero si o coeficiente toma un valor próximo a +1 ou -1 diremos que a relación é forte. O signo do coeficiente indícanos a relación, si é positivo a relación será positiva e si é negativo dita relación será negativa.

Formalmente:

$$\rho_{x,y} = \frac{\text{Cov}(x,y)}{\sigma_x \sigma_y}$$

É dicir, a covarianza de X e Y entre o produto das desviacións estándar das variables X e Y.

Gráfico 2 Coeficientes de correlación de Pearson



Fonte da imaxe: Wikipedia

Coefficiente de desigualdade U de Theil

Término en inglés: **Theil Inequality Coefficient**

Este coeficiente utilízase en econometría como medida avaliadora da capacidade predictiva dun modelo econométrico. Defínese como:

$$U_{\text{Theil}} = \frac{\sqrt{\sum_{t=1}^n e_t^2}}{\sqrt{\sum_{t=1}^n \hat{y}_t + \sum_{t=1}^n y_t}}$$

Con $0 \leq U_{\text{Theil}} \leq 1$

Se o U de Theil é próximo a cero indicaríanos que é unha boa predición. En cambio, se é próximo a 1, indica que o modelo non serve para predicir. O coeficiente de U de Theil pódese descompoñer en

- U_n : Que indicaría a proporción do erro debido ao nesgo, mediría o erro medio ao facer a predición e indicaría a presenza dalgún error sistemático que debería evitarse. Cando $U_n = 0$ non hai erro de predición sistemático.
- U_v : Proporción do erro debida as diferenzas das varianzas entre os valores observados e os valores de predición. Cando é igual a 0, indica igual variabilidade nos valores de predición e observados.
- U_a : Proporción do erro aleatoria. Indicaría a variación aleatoria no erro. O desexable sería que o erro fose debido a este factor.

Entón, entendemos que se as proporcións do nesgo e varianza son pequenas, a predición é boa.

Coefficiente de determinación

Termo en inglés: **Coefficient of Determination**

O coeficiente de determinación é unha medida da bondade do axuste, da capacidade das variables independentes de explicar o comportamento da variable dependente. O R^2 indícanos a proporción de variación mostral de Y que se explica polos regresores (X) do modelo.

No modelo de regresión lineal pódese calcular a través da seguinte fórmula:

$$R^2 = \frac{SCR}{SCT} = 1 - \frac{SCE}{SCT}$$

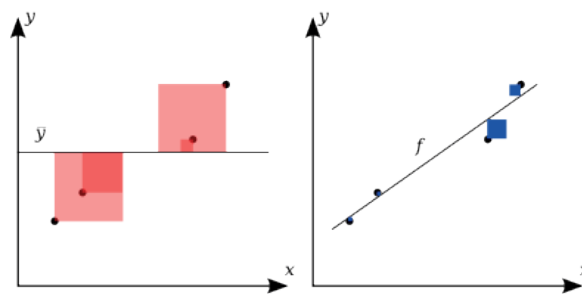
sendo:

SCE=Suma de cadrados dos erros; SCT=Suma de cadrados totais; e SCR=Suma de cadrados da regresión.

Dado que se trata de sumas de cadrados, e cumpríndose que $SCT=SCR+SCE^1$ os seus límites serán cero e un: $0 \leq R \leq 1$. Canto máis próximo estea o resultado á unidade, mellor será o axuste, e canto máis próximo a cero, peor será o axuste.

* Require que o modelo teña ordenada na orixe e que o método de estimación sexa MCO.

Gráfico 3. Coeficiente de determinación



Fonte da imaxe: Wikipedia

Coefficiente de determinación axustado

Termo en inglés: **Adjusted Coefficient of Determination**

O coeficiente de determinación axustado ou corrixido mide a porcentaxe da variación da variable dependente explicada polo modelo (igual que o fai o coeficiente de determinación) tendo en conta ao número de variables explicativas incluídas no modelo. Intenta corrixir o problema do coeficiente de determinación como medida da bondade do axuste dun modelo, problema derivado de que R^2 sempre crece cando se incrementa o número de regresores e non resulta útil para a comparación de modelos cun distinto número de regresores.

Formalmente:

$$\text{Adj}R^2 = \bar{R}^2 = \frac{SCR}{SCT} = 1 - \frac{SCE/T-k-1}{SCT/T-1}$$

ou alternativamente, en función da súa relación con R^2 :

Sendo:

$$\text{Adj}R^2 = \bar{R}^2 = 1 - \frac{T-1}{T-k-1} (1-R^2)$$

T: número de observacións, tamaño mostral; k: número de variables independentes.

Coefficientes de reacción

Termo en inglés: **Reaction Coefficients**

Nun modelo econométrico dinámico con retardos distribuídos, os coeficientes da variable esóxena e os seus retardos, β_p , denomínanse coeficientes de reacción. Na teoría dinámica reciben o nome de multiplicadores (relación entre a causa e o efecto), e amosan a resposta da variábel dependente Y ante un cambio unitario na variábel esóxena X , “ i ” períodos máis tarde.

Exemplo:

No seguinte modelo:

$$y_t = \alpha + \sum_{i=0}^n \beta_i x_{t-i} + \varepsilon_t$$

β_2 mediría o efecto sobre a endóxena, Y , dun cambio unitario da variábel esóxena, X , dous períodos de tempo máis tarde.

Cointegración

Termo en inglés: **Cointegration**

Se existe unha combinación lineal estacionaria de variables non estacionarias, as variables combinadas dise que están cointegradas. A cointegración prodúcese cando as series temporais comparten unha tendencia estocástica común. Nese caso poderíase soste a existencia dunha relación a longo prazo entre as variables.

A existencia dunha relación de cointegración entre as variables dun modelo de regresión permite modelizar a relación a longo prazo entre variables non estacionarias sen risco de ter unha regresión espuria.

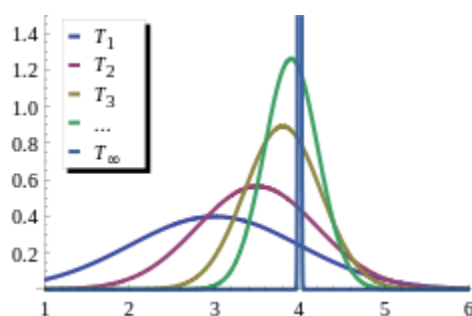
Consistencia

Termo en inglés: **Consistency**

A consistencia é unha propiedade estatística dos estimadores, das mostras grandes, que consiste en que, conforme se vai aumentando o tamaño da mostra, a distribución do estimador tende a concentrarse entono ao valor poboacional verdadeiro do parámetro.

Esta propiedade tamén pode ser definida como: o límite probabilístico dun estimador b calquera cando o tamaño da mostra tende a infinito é igual ao parámetro β , $\text{Plim } b = \beta$

Gráfico 4. Propiedade da consistencia dun estimador



Fonte da imaxe: Wikipedia

Contraste de hipóteses

Término en inglés: **Hypothesis Testing**

Consiste en determinar se é aceptable, partindo de datos mostrais, que a característica ou o parámetro poboacional estudado tome un determinado valor ou estea dentro duns determinados valores. O contraste de hipótese estatístico basearase na información proporcionada pola mostra, de modo que, se **rexecemos a hipótese**, queremos indicar que os datos da mostra ofrecen certa evidencia sobre a súa falsidade; e se a **aceptamos** simplemente queremos significar que non se rexeita.

Un contraste de hipótese consiste, polo tanto, en estudar dúas hipóteses: **H_0 (hipótese nula)** e **H_1 (hipótese alternativa)**, de maneira que se divide os resultados mostrais en dúas zonas; unha zona de rexeitamento da hipótese nula e outra de aceptación de dita hipótese.

Cando se realiza un contraste pódense cometer dous tipos de erros: O **erro tipo I** é o erro que se comete na decisión do contraste cando se rexeita a hipótese nula (H_0), sendo correcta (certa); o **erro tipo II** é o erro que se comete na decisión do contraste cando se acepta a hipótese nula (H_0) cando realmente é falsa. Os contrastes de hipóteses constrúense de forma que a probabilidade de cometer o erro tipo I sexa bastante pequena. Dita probabilidade denótase por α e denomínase nivel de significación do contraste.

Contraste t de significación individual

Término en inglés: **t-test of Individual Significance**

Nos modelos econométricos un contraste de significación individual dun parámetro determinado verifica se, coa información dispoñible na mostra, podemos rexeitar ou non que o parámetro β_i é estatisticamente distinto de cero.

As hipóteses que se formulan son:

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0$$

Un estatístico que se pode utilizar para realizar os contrastes individuais é un estatístico t que se define como o valor estimado do parámetro β_i dividido pola súa desviación típica. Este estatístico t, se a hipótese nula é certa, distribúese como unha t-student con $T-(k+1)$ sendo T o tamaño da mostra e $k+1$ o número de regresores do modelo (incluído o regresor ficticio).

Se no contraste de significación individual se rexeita a hipótese nula ou sexa que $\beta_i = 0$ podemos dicir que a variable explicativa á cal acompaña o parámetro ten un efecto estatisticamente significativo sobre o regresando.

Correlación contemporánea

Término en inglés: **Contemporaneous Correlation**

Na modelización econométrica cando se asume como hipótese que os regresores do modelo son estocásticos, o termo de correlación contemporánea emprégase para referirse a correlación entre os regresores e o termo do erro ou perturbación aleatoria no mesmo momento de tempo, cando se trata de datos temporais, ou para a mesma observación mostral cando se trata de datos atemporais.

Formalmente:

$$\text{Cov}(x_{it}, \varepsilon_t) \neq 0 \quad \forall t=t'$$

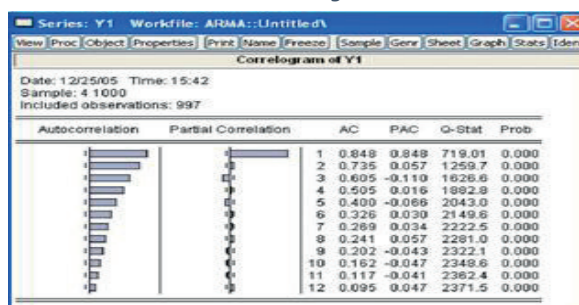
Correlograma

Termo en inglés: **Correlogram**

O correlograma é unha representación gráfica dos coeficientes de autocorrelación (e coeficientes de autocorrelación parcial) estimados para distintos retardos dunha serie de tempo. Emprégase para analizar o comportamento dunha serie temporal en relación ao seu carácter estacionario (ou non) e o tipo de proceso estocástico que se pode intuír xerou a serie en estudo.

No correlograma recóllese, nun diagrama de barras, os coeficientes de autocorrelación simple (función de autocorrelación) e os coeficientes de autocorrelación parcial (función de autocorrelación parcial) da serie temporal. Utilízase na elaboración dun modelo ARIMA para unha serie temporal, no enfoque de Box-Jenkins, na etapa de identificación do modelo. A análise do correlograma permitíranos analizar se a serie é ou non estacionaria e identificar a orde do polinomio autorregresivo (p) e o de medias móbiles (q), de selo caso.

Gráfico 5. Correlograma



Fonte da imaxe: Elaboración propia

Covarianza

Termo en inglés: **Covariance**

A covarianza é un parámetro estatístico que nos indica o grado de relación que existe entre dúas variables aleatorias respecto ás súas medias. É dicir, sabemos se hai asociación no comportamento das variables. A fórmula da covarianza entre as variables X e Y exprésase da seguinte maneira:

$$\text{Cov}(X,Y)=E[(X-E(X))(Y-E(Y))]$$

Interpretación dos resultados:

- Se $\text{Cov}(X, Y) > 0$, a relación entre as variables é directa (positiva), cando aumenta X, Y tamén aumenta.
- Se $\text{Cov}(X, Y) < 0$, a relación entre as variables é inversa (negativa), cando X aumenta, Y diminúe.
- Se $\text{Cov}(X, Y) = 0$, non existe relación entre as variables.

Criterio de información de Akaike (AIC)

Termo en inglés: **Akaike's Information Criterion**

Proporciona unha medida para a selección dun modelo entre varios modelos alternativos, e o seu obxectivo é calcular unha medida que indique como de próximos están os modelos alternativos ao verdadeiro modelo xerador dos datos.

O seu cálculo realízase a partir da seguinte expresión:

$$\text{AIC}(k)=2k-2\text{Ln}L$$

onde k é o número de parámetros estimados no modelo econométrico (β, σ^2), e L o valor máximo da función de verosimilitud do modelo estimado. Unha vez calculado o AIC para cada modelo pódense clasificar de mellor a peor, sendo o “mellor” modelo o que teña o mínimo valor do AIC.

Criterio de información de Schwarz (ou criterio de información bayesiano)

Termo en inglés: **Schwarz Information Criterion (or Bayesian Information Criterion)**

O Criterio de Información de Schwarz (SIC) tamén coñecido como criterio de información Bayesiano (BIC), é unha medida que axuda na selección do modelo que mellor se axuste aos datos proporcionados, entre modelos alternativos. Elaborouse a partir do criterio de información de Akaike pero é máis preciso pois ten en conta o número de observacións (T). Pódese definir coa seguinte fórmula:

$$\text{SIC}(k) = k(\ln T) - 2L_n L$$

onde $L_n L$ é o máximo do logaritmo da función de verosimellanza (indica o valor estimado do parámetro que é máis compatible cos datos observados), k é o número de parámetros estimados no modelo de regresión (β, σ^2) e T é o tamaño da mostra.

Na selección entre modelos, escolleríamos aquel modelo que presente un menor BIC.

Criterio de información Hannan-Quinn

Termo en inglés: **Hannan-Quinn Information Criterion**

O criterio de Información de Hannan-Quinn (HQC) emprégase en econometría como criterio para seleccionar entre un conxunto finito de modelos. Pode ser considerado unha variante do BIC (Criterio de Información Bayesiano) cunha pequena penalización da magnitude do tamaño da mostra:

$$\text{HQC}(k) = 2kL_n(\ln T) - 2L_n L$$

Onde $L_n L$ indica o máximo do logaritmo da función de verosimellanza obtida cos valores estimados dos parámetros do modelo econométrico (β, σ^2); k o número de parámetros estimados no modelo econométrico, e T é o tamaño da mostra.

Cando se nos presentan modelos alternativos estimados, o que teña un menor valor HQC será o considerado óptimo, xa que implicaría un menor número de variables explicativas e un mellor axuste.

Datos de panel

Termo en inglés: **Panel data**

En Econometría definimos os datos de panel como aquel conxunto de datos que ten tanto dimensión temporal como atemporal. É dicir, teremos datos de panel cando temos observacións das mesmas unidades (N) en distintos momentos de tempo (T). A nosa mostra será de tamaño $N \cdot T$.

A característica que define este tipo de datos é que se obteñen datos da mesma unidade individual aleatoria ao longo do tempo (tamén se lles denomina datos lonxitudinais).

Diagrama de predición

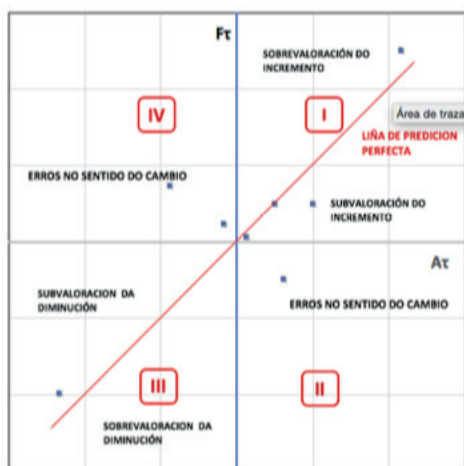
Termo en inglés: **Prediction Diagram**

Trátase dunha maneira de valorar a capacidade predictiva do modelo econométrico a través dun gráfico das variacións reais e previstas. Para facelo, calcúlase a porcentaxe da variación real (A) e porcentaxe da variación prevista (F) do regresando en cada período de predición:

E represéntase no plano os pares de valores (A_τ, F_τ) obtidos, xunto coa bisectriz do primeiro ao terceiro cuadrante, que é a recta de predicións perfectas.

$$A_\tau = \frac{y_\tau - y_{\tau-1}}{y_{\tau-1}} * 100; F_\tau = \frac{\hat{y}_\tau - y_{\tau-1}}{y_{\tau-1}} * 100$$

Gráfico 6. Gráfico de variacións reais e previstas



Fonte da imaxe: Elaboración propia

Para que o modelo proporcione boas predicións é necesario, como mínimo, que permita anticipar correctamente o sentido do cambio no regresando, e dicir, que se no período τ o regresando realmente aumenta/diminúe, o modelo prediga un aumento/diminución. Isto sucede cando os porcentaxes de variación real e previsto son do mesmo signo, no gráfico, sobre os cuadrante I y III, en cambio, se $A\tau$ e $F\tau$ teñen signos distintos, nin sequera se anticipa correctamente o sentido no que cambia a variable, no gráfico, sobre os cuadrante II y IV.

Distribución normal

Termo en inglés: **Normal Distribution**

Unha variable aleatoria normal é unha variable aleatoria continua que pode tomar calquera valor q no que a súa función de densidade de probabilidade presenta forma de campá simétrica. A función de densidade da distribución normal (tamén coñecida como distribución gaussiana) ven dada por:

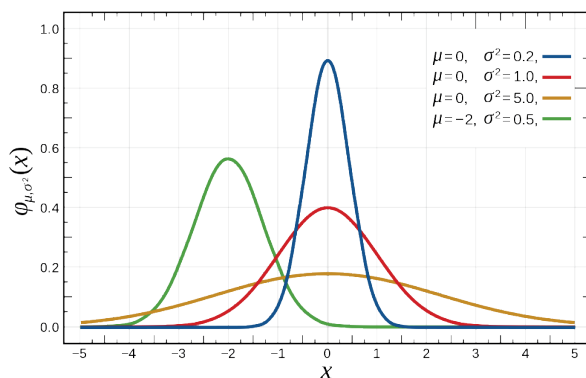
$$\phi_{\mu,\sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$

μ : media da distribución.

σ : desviación estándar da distribución.

Dicimos que X ten unha distribución normal de media μ e desviación estándar σ denótase como $N(\mu,\sigma)$.

Gráfico 7. Función densidade distribución normal



A liña vermella correspóndese á distribución normal estándar coa función de densidade dunha normal.

Fonte da imaxe: Wikipedia

Nos modelos econométricos o suposto de normalidade na perturbación aleatoria é base para realizar contrastes de hipóteses e estimación por intervalo. Ademais emprégase este suposto na estimación por máxima verosimellanza dos parámetros do Modelo Clásico Normalizado.

Econometría

Termo en inglés: **Econometrics**

A econometría é unha disciplina científica que combina á teoría económica, a inferencia estatística e as matemáticas. A asociación destes tres campos do coñecemento científico, converxen nesta nova ciencia social con características propias.

Ten por obxecto a explicación e a predición de fenómenos económicos, mediante o uso de modelos expresados en forma matemática e a utilización de métodos estatísticos de estimación e contrastes.

Tradicionalmente, a econometría estivo centrada nas relacións económicas agregadas, por exemplo o comportamento do PIB, tipos de interese, taxas de inflación etc. Pero dende 1970, os métodos econométricos empréganse cada vez máis en modelos microeconómicos que describen o comportamento individual, familiar ou empresarial, como por exemplo o gasto dunha familia en función da súa renda.

Actualmente, a econometría desempeña un papel importante no traballo empírico en tódolos campos da economía.

Econometría financeira

Termo en inglés: **Financial econometrics**

A econometría financeira afronta o estudo cuantitativo dos problemas que se formulan desde as finanzas. Ao pertencer á Econometría, o estudo realízase empregando técnicas matemáticas e estatísticas e a teoría económica.

Así, na econometría financeira constrúense modelos econométricos financeiros que poden afrontar cuestións como a estimación da volatilidade dos mercados, a xestión do risco, a análise dos beneficios e perdas dun mercado financeiro, a xestión da carteira de investimentos etc.

Por exemplo, podemos estar interesados facer unha valoración completa dun activo en concreto, coñecer que efecto poden ter os tipos de interese ou a taxa de inflación, ou outros factores, ademais do risco que ven implícito nunha valoración deste tipo.

Ecuación de comportamento

Termo en inglés: **Behavioural Equation**

Nun modelo econométrico un dos elementos que o constitúen son as ecuacións. Segundo o carácter empírico das mesmas, poderemos clasificalas en: Ecuacións de comportamento, institucionais ou legais, tecnolóxicas, de definición ou de identidade e de equilibrio móbil.

As ecuacións de comportamento explican o xeito de actuar dos axentes económicos, é dicir, son aquelas que reflexan o comportamento destes axentes (consumidores, produtores, investidores). Este tipo de ecuacións son as máis destacadas dentro dos modelos macroeconométricos, é dicir, en aqueles modelos nos que o obxectivo principal é o de explicar o comportamento empírico do sistema económico: a demanda, a oferta, o investimento, o consumo, o aforro etc.

Un exemplo deste tipo de ecuacións de comportamento sería a función de consumo na que se establece que o gasto en consumo, ou o comportamento de consumo, é sensibles aos ingresos ou á función de ingresos:

$$C_t = \beta_0 + \beta_1 Y_d_t + \epsilon_t$$

Sendo C_t , consumo, Y_d_t ingreso dispoñible, e sendo ϵ_t o termo da perturbación aleatoria e β_0 e β_1 os parámetros estruturais.

Erro de especificación

Término en inglés: **Specification Error**

Na modelización econométrica, un erro de especificación é o incumprimento de calquera dos supostos básicos do modelo lineal xeral. Nun sentido máis concreto, os erros de especificación céntranse habitualmente na omisión de regresores relevantes, na inclusión de regresores irrelevantes, ou nunha forma funcional incorrecta (a cal en moitos casos, pode verse como unha mala especificación da matriz de regresores X).

A cuestión de interese é como afecta, aos estimadores MCO/MCX, que se cometa algún tipo de erro de especificación.

Erro de predición

Término en inglés: **Forecast Error**

É a diferenza entre o valor observado da variable Y no período de predición e o valor que o modelo predí para Y nese período:

$$e_{\tau} = y_{\tau} - \hat{y}_{\tau}$$

sendo

- e_{τ} = erro de predición, no período de predición τ
- y_{τ} = valor observado de Y no período de predición τ
- \hat{y}_{τ} = valor que o modelo predí para Y no período de predición τ

Este erro tamén se refire a diferenza entre o valor observado e o valor que o modelo predí para unha observación que non pertence á mostra coa que se realizou a estimación do modelo.

Erro estándar da regresión

Término en inglés: **Standard error of the regression**

Para analizar a bondade do axuste nun modelo econométrico de regresión lineal unha das medidas empregadas é o Erro Estándar da regresión que, expresada en porcentaxe, indícanos o erro estándar que cometemos ao estimar o modelo con respecto á media do regresando.

Calcúlase como:

$$\%ES = \frac{\sqrt{SCE / T - k - 1}}{\bar{Y}} * 100$$

Sendo:

- SCE: Suma de cadrados dos erros.
- T: Número de observacións.
- k+1: Número de regresores do modelo (incluíndo o regresor ficticio).

A %ES dun modelo con axuste perfecto sería 0, e a medida que nos afastamos de dito valor, o axuste realizado sería menos bo.

Esperanza matemática da perturbación non nula

Término en inglés: **Non-zero expected value of random error (or random disturbance)**

Nun modelo econométrico, se a esperanza matemática da perturbación aleatoria non é nula afectará as propiedades dos estimadores MCO, concretamente deixarían de ser inesgados. Formulamos dúas posibles situacións:

— $E\varepsilon_t = \mu$ para todo t

Non nula pero constante. Neste caso podemos obter ELIO de todos os parámetros do modelo excepto da ordenada na orixe, cuxa estimación sería nesgada.

— $E\varepsilon_t = \mu_t$ para todo t

Non nula e variando en t . Os EMCO de β serán nesgados.

Esta circunstancia pode producirse por unha incorrecta especificación do modelo, como a omisión dalgún regresor relevante.

Esperanza matemática da perturbación nula

Término en inglés: **Null expected value of random error (or random disturbance)**

É unha das hipóteses relativas á perturbación aleatoria. Significa que, en promedio, o efecto conxunto sobre o comportamento da variable dependente de todos os factores ou variables recollidas na perturbación aleatoria é nulo.

Formalmente:

$E\varepsilon_t = 0$ para todo t

A consecuencia disto, se a esperanza da perturbación é cero, o valor esperado da variable dependente non viría influído pola perturbación, cumpríndose que:

$$E(Y) = X\beta$$

É dicir, o valor esperado (promedio) de y só depende de x pero non da perturbación aleatoria.

Estatístico de proba

Término en inglés: **Test Statistic**

Un estatístico de proba é unha variable aleatoria que se calcula a partir de datos da mostra e que se utiliza para tomar unha decisión nun contraste de hipóteses. Permite probar se a diferenza observada entre os datos e a hipótese nula formulada no contraste de hipóteses é real ou se debe a unha variación aleatoria. Dito estatístico é unha variable aleatoria cunha distribución mostral determinada, preferentemente con probabilidades tabuladas. En econometría son moi frecuentes estatísticos que teñen distribucións como a normal, t , χ^2 ou F .

Estatístico F

Término en inglés: **F-Statistic**

O estatístico F utilízase para o contraste de hipóteses conxuntas relativas a máis dun coeficiente da regresión, sendo un dos seus usos máis comúns avaliar se un conxunto ou subconxunto de variables independentes teñen efecto conxunto sobre o regresando.

Para o caso concreto da análise de nulidade conxunta para todos os parámetros (significatividade conxunta de todas as variables explicativas), pero sen ter en conta o regresor ficticio, o contraste formularíase do seguinte modo:

Hipótese que se contrasta:

$$H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$$

H_1 : Algún ou todos os β_i distintos de 0

Estatístico de proba:

$$F = \frac{SCR/k}{SCE/T-k-1}$$

que baixo o suposto de H_0 certa seguirá unha distribución F de Snedecor con k graos de liberdade no numerados e T-k-1 graos de liberdade no denominador.

Sendo:

- SCR: Suma de cadrados da regresión.
- SCE: Suma de cadrados dos erros.
- T: Número de observacións.
- k+1: Número de regresores do modelo (incluído o regresor ficticio).

Estimador lineal

Térmo en inglés: **Lineal Estimator**

Un estimador lineal é aquel que é unha función lineal dos elementos da mostra. No caso do estimador “b” obtido segundo o método MCO:

$$b = (X'X)^{-1}X'Y$$

Considerando X unha matriz non estocástica, o produto $(X'X)^{-1}X'$ será tamén non estocástico, e b será un estimador lineal por ser unha función lineal dos valores mostrais de Y.

Estimador de máxima verosimellanza (EMV)

Térmo en inglés: **Maximum Likelihood Estimator (MLE)**

O método de estimación de máxima verosimellanza formula que se elixan como estimadores dos parámetros β e σ^2 do MRLNC aqueles valores dos parámetros que maximicen a función de verosimellanza.

Baixo a hipótese do modelo de regresión lineal normal clásico, o vector Y é un vector de variables aleatorias independentes con distribución normal. A función de densidade conxunta, cando se considera como función dos parámetros (β e σ^2) coñécese como función de verosimellanza, denotada como $L(\beta, \sigma^2)$. Os estimadores de máxima verosimellanza son os valores de estes parámetros que xerarían con maior probabilidade a mostra observada.

Estes estimadores teñen as propiedades asintóticas: asintoticamente inesgados, asintoticamente eficientes e consistentes; e demóstrase que nun MRLNC os EMV coinciden cos EMCO.

Estimador inesgado

Térmo en inglés: **Unbiased Estimator**

Que un estimador sexa inesgado significa que a súa esperanza matemática coincide co valor do parámetro que se quere estimar. A inesgadez é unha das propiedades que se lle esixe aos estimadores dos parámetros.

No caso dos modelos econométricos, é unha propiedade que tamén se busca ao facer as estimación dos parámetros do modelo. Nun MRLC, a non nesgadez é unha propiedade dos estimadores MCO. É dicir, pódemos demostrar que nun modelo clásico $Eb=\beta$, sendo b o estimador MCO.

Cando a esperanza matemática do estimador non coincide co verdadeiro valor do parámetro, pódese dicir que o estimador ten un nesgo.

Estimador óptimo

Termo en inglés: **Best Estimator (Best Lineal Unbiased Estimator, BLUE)**

Defínese como óptimo o estimador con menor varianza entre a clase de estimadores lineais e non nesgados. Baixo os supostos do modelo clásico, mediante o teorema de Gauss-Markov demóstrase que os estimadores MCO son ELIO dos parámetros do modelo.

O estimador lineal, inesgado e óptimo (ELIO) de β será:

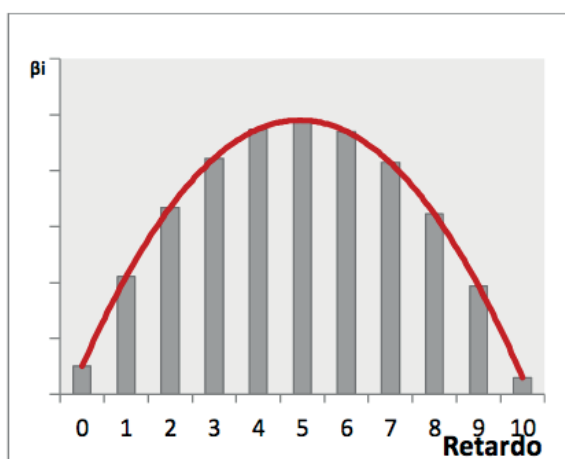
- Unha función lineal de Y .
- Non nesgado.
- De varianza mínima entre os estimadores lineais e non nesgados.

Estrutura temporal do retardo

Termo en inglés: **Lag Structure**

Nun modelo dinámico de retardos distribuídos, os parámetros β_i que na ecuación acompañan as variábeis retardadas, forman o que se chama a estrutura temporal do retardo. Dita estrutura pódese representar graficamente, representando os parámetros en función do retardo. Asíumese como hipótese natural que suma dos β_i é converxente, que non tende a infinito.

Exemplo de estrutura temporal do retardo: O valor dos β_i vai crescendo ata un determinado retardo no que comezan a diminuír. É dicir, neste exemplo, o efecto dun cambio na variable vai medrando ata un determinado retardo a partir do cal o efecto da variable comeza a diminuír.



Fonte da imaxe: Guntín, X. (2020). *Modelos dinámicos*. Recuperado de cv.usc.es.

Función de autocorrelación (FAC)

Término en inglés: **Autocorrelation Function**

A función de autocorrelación (FAC) representa os coeficientes de autocorrelación en función da distancia temporal ou do retardo entre as variables que forman unha serie temporal.

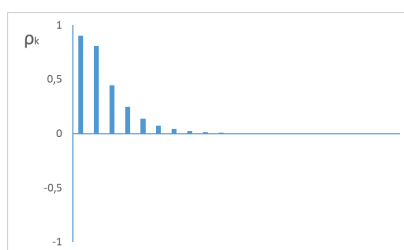
Os coeficientes de autocorrelación de orde k dun proceso estocástico (Y_t), miden o grao de asociación lineal existente entre dúas variables aleatorias do proceso separadas k períodos (Y_t e Y_{t-k}). Calcúlanse como a ratio das covarianzas de Y_t e Y_{t-k} e as súas respectivas varianzas, dando lugar ao conxunto dos coeficientes de autocorrelación $\rho_{k(t)}$.

$$\rho_k = \frac{E[Y_t - \mu][Y_{t-k} - \mu]}{\sqrt{E[Y_t - \mu]^2 E[Y_{t-k} - \mu]^2}} = \frac{\text{Cov}(Y_t, Y_{t-k})}{\sqrt{\text{Var}(Y_t) \text{Var}(Y_{t-k})}}$$

Dado que cada ρ_k é un coeficiente de autocorrelación, soe dicirse que a función de autocorrelación representa a intensidade e a duración da memoria do proceso (Y_t).

Un exemplo da función de autocorrelación para un proceso estocástico estacionario AR(1) con parámetro positivo:

Gráfico 8. Función de autocorrelación



Fonte da imaxe: Elaboración propia

Función de autocorrelación parcial (FACP)

Término en inglés: **Partial Autocorrelation Function.**

A función de autocorrelación parcial ten en conta os valores de correlación entre dúas variables aleatorias do proceso separadas entre si “ k ” períodos e en función dos valores intermedios entre elas. É dicir, recolle a correlación lineal existente entre as variables do proceso estocástico asociadas a distintos momentos de tempo, descontado o efecto que teñen os retardos intermedios sobre elas.

$$\phi_{kk} = \text{corr}(Y_t, Y_{t-k} / Y_{t-1}, Y_{t-2}, \dots, Y_{t-k+1})$$

Xunto á FAC sérvenos para identificar a orde dos procesos nun modelo ARMA e ARIMA.

Heterocedasticidade

Término en inglés: **Heterocedasticity.**

Nun modelo econométrico, cando a varianza dos perturbacións aleatorias (ou erros aleatorios) non é constante ao longo da mostra, é dicir, nas distintas observacións que realizamos, dise que existe heterocedasticidade. Implica o incumprimento dunha das hipóteses básicas do modelo de regresión lineal que denominamos clásico. As causas son diversas, nalgúns casos prodúcese heterocedasticidade porque, pola súa propia natureza, algunhas variables son heterocedásticas.

Formalmente exprésase como:

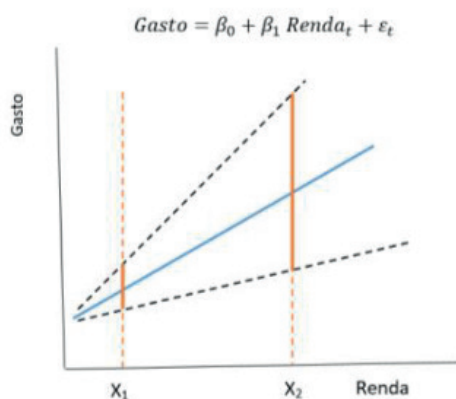
$$V(\epsilon_t) = E(\epsilon_t - E\epsilon_t)^2 = E\epsilon_t^2 = \sigma_t^2$$

para calquera t (dado que no modelo clásico $E\epsilon_t = 0$).

Exemplo:

No exemplo que se mostra a continuación vemos como se relaciona o gasto das familias coa súa renda. Dispoñemos dun grupo de familias con rendas baixas, para este grupo haberá moi pouca diferenza entre as familias que deciden aforrar todo e as que deciden gastar todo debido a que os reducidos ingresos non permiten moita variación; por outro lado temos outro grupo de familias con rendas elevadas para as cales pode existir unha diferenza maior entre familias cunha alta propensión ao aforro e familias cunha baixa propensión ao aforro. Entón, como podemos observar na gráfica, aparecería a heterocedasticidade.

Gráfico 9. Heterocedasticidade



Fonte da imaxe: Elaboración propia

Hipótese alternativa

Termo en inglés: **Alternative hypothesis**

Nun contraste de hipóteses, para comprender a hipótese alternativa, débese partir do concepto da hipótese nula. A hipótese nula indica que un parámetro da poboación, por exemplo, a media da poboación, é igual a un valor hipotético específico. Normalmente obtense a partir dun estudo especializado a partir de análises anteriores. Chámase H_0 e é a hipótese que se quere probar porque é a que se formula en primeiro lugar e a que se manterá a menos que os datos indiquen que é falsa.

A hipótese alternativa, que se denota como H_1 , indica unha afirmación contraria á hipótese nula. É a negación da hipótese nula e da hipótese que hai que verificar. Representa polo tanto o que cada un trata de probar, a conclusión que quere demostrar co seu estudo, negando a hipótese nula.

Hipótese de rango pleno

Termo en inglés: **Full Rank Hypothesis**

Refírese a unha das hipóteses que se establece na especificación dun modelo econométrico. Concretamente, supón que a matriz de observacións dos regresores X (de tamaño $T \times (k+1)$) ten rango igual a $k+1$, é dicir, o rango desta matriz é igual ao número de columnas da matriz, o que supón que todas as súas columnas son linealmente independentes. Nótese que tamén se require que $T > k+1$.

Formalmente:

$$\text{rg}(X) = k+1 < T$$

Hipótese nula

Término en inglés: **Null Hypothesis**

Para explicar este termo é preciso aclarar primeiro que é unha hipótese estatística.

Unha hipótese estatística é unha afirmación en relación a unha característica da poboación, un enunciado sobre os valores dalgún parámetro da poboación da cal se toma a mostra. Para saber se a nosa afirmación é correcta necesitaremos contrastar a hipótese. O contraste da hipótese é un procedemento que responde á pregunta de se a diferenza entre o valor da mostra e o valor da poboación se debe a unha diferenza real ou é unha variación aleatoria.

Neste contexto, a hipótese nula, denotada habitualmente como H_0 , é a hipótese que se proba. Especifica un só valor do parámetro se a hipótese é simple ou un conxunto de valores se é composta.

Homocedasticidade

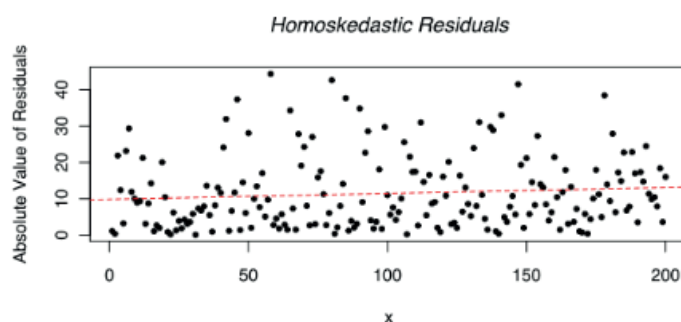
Término en inglés : **Homoscedasticity**

É unha hipótese relativa á perturbación aleatoria, propia do modelo de regresión lineal clásico. A homocedasticidade existe cando as perturbacións aleatorias teñen a mesma dispersión ou variabilidade ao longo da mostra. Formalmente a expresamos como:

$$V(\epsilon_t) = E(\epsilon_t - E\epsilon_t)^2 = E\epsilon_t^2 = \sigma^2 \quad \forall t$$

(dado que no modelo clásico $E\epsilon_t = 0$).

Gráfico 10. Homocedasticidade



Fonte da imaxe: Wikipedia

Identificación (metodoloxía ARIMA)

Término en inglés: **ARIMA: Model Identification**

No metodoloxía de Box-Jenkins, para a elaboración dun modelo ARIMA, hai unha serie de pasos básicos que se deben dar. Podemos falar de catro fases, sendo a primeira delas a de identificación.

Identificación: Debemos decidir que transformacións aplicar para converter a serie en estacionaria (en caso de non selo) é determinar as ordes de p , d , e q (P, D , e Q no caso de que existise parte estacional), é dicir, identificar un modelo tentativo.

- Para saber se a serie é estacionaria podemos empregar o gráfico de evolución temporal ou a función de autocorrelación temporal (correlograma). Hai que ter en conta que para que unha serie sexa estacionaria, a media e a varianza teñen que oscilar arredor dun valor aproximado. A través do gráfico de evolución temporal comprobaremos se a serie é estacionaria en media e varianza. A aplicación de diferenzas pode converter unha serie non estacionaria en estacionaria. Determinaríamos así o valor de “ d ”. Outra transformación posible é aplicar logaritmos.
- Por outra banda, debemos determinar os valores de p e q , para o que empregaremos a función de autocorrelación (FAC) e a función de autocorrelación parcial (FACP) estimadas.

Incorrelación contemporánea entre X e ε

Termo en inglés: **Contemporaneous Uncorrelation**

Nun modelo econométrico no que se consideran que os regresores son estocásticos, definimos a incorrelación contemporánea como aquela situación na que os regresores (estocásticos) e a perturbación aleatoria do modelo, non sendo independentes, non presentan correlación no mesmo momento tempo ou para a mesma observación mostral.

Formalmente:

$$\begin{aligned} \text{Cov}(x_{it}, \varepsilon_t) &= 0 \quad \forall t=t' \\ \text{Cov}(x_{it}, \varepsilon_{t'}) &\neq 0 \quad \forall t \neq t' \end{aligned}$$

Neste suposto, os estimadores mínimo cadráticos ordinarios (EMCO) dun modelo que cumpre os supostos do modelo clásico agás o relativo ao carácter non estocástico dos regresores, serán nesgados e non lineais, e unicamente manteñen a propiedade da consistencia.

Intervalo de confianza

Termo en inglés: **Confidence Interval**

É unha técnica de estimación empregada en inferencia estatística que nos permite atopar dous valores ($a, b, a < b$) entre os cales se estima que estará certo valor descoñecido cunha determinada probabilidade (nivel de confianza), é dicir, estes valores van acoutar un intervalo probabilístico dentro do cal vaise atopar o parámetro poboacional.

Formalmente:

$$P(\theta \in [a, b]) = 1 - \alpha \quad \text{ou} \quad P(a \leq \theta \leq b) = 1 - \alpha$$

Matriz de varianzas-covarianzas

Termo en inglés: **Variance-covariance matrix**

Dado un vector de variables aleatorias Y , de tamaño $T \times 1$, chamamos matriz de varianzas-covarianzas á matriz cadrada $T \times T$ que dispoña na súa diagonal principal as varianzas de cada elemento de Y e nos elementos non diagonais as correspondentes covarianzas entre cada par de elementos do vector aleatorio Y .

Chamando V a esta matriz:

$$V = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1T} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{T1} & \sigma_{T2} & \cdots & \sigma_T^2 \end{bmatrix}$$

sendo $\sigma_i^2 = \text{Var}(y_i)$ e $\sigma_{ij} = \text{Cov}(y_i, y_j)$.

Dado que $\text{Cov}(y_i, y_j) = \text{Cov}(y_j, y_i)$ é inmediato comprobar que unha matriz de varianzas-covarianzas é simétrica.

Método da matriz de transformación

Termo en inglés: **Transformation matrix**

O chamado método da matriz de transformación consiste en identificar unha matriz tal que ao premultiplicar o modelo de regresión lineal xeralizado con dita matriz, o modelo resultante sexa un modelo de regresión lineal clásico.

Pódese demostrar que, derivado das propiedades das matrices definidas positivas, e tendo en conta que a matriz de varianzas-covarianzas da perturbación aleatoria $V(\varepsilon)$ nun MRLX é definida positiva, dita matriz existe e debe cumprir a condición de que $T'T = \Omega^{-1}$, sendo, T a matriz de transformación é a Ω a matriz que se deriva de V , $V = \sigma^2 \Omega$, con σ^2 un factor de proporcionalidade descoñecido.

Metodo de mínimos cadrados ordinarios

Termo en inglés: **Ordinary Least Squares (OLS)**

O método de Mínimos Cadrados Ordinarios (MCO) é un método de estimación dos parámetros dun modelo econométrico de regresión lineal que se basea no criterio de minimizar a suma de cadrados dos erros (diferenza entre o valor observado de regresando e o seu valor estimado). É dicir, o vector de estimadores mínimo cadrados ordinarios (EMCO) de β é o vector b que minimiza a SCE.

A súa ecuación é: $b = (X'X)^{-1} X'Y$

Método de mínimos cadrados xeralizados

Termo en inglés: **Generalized Least Squares (GLS)**

O método de Mínimos Cadrados Xeralizados (MCX) é un método de estimación dos parámetros descoñecidos nun modelo de regresión lineal. Este método permite obter estimadores con boas propiedades estatísticas nos modelos de regresión lineais xeralizados, nos que a matriz de varianzas-covarianzas da perturbación non se supón que é unha matriz escalar, polo que a os supostos de homocedasticidade e/ou incorrelación non se requiren. É dicir, incorpora o suposto de heterocedasticidade e/ou un certo grao de autocorrelación entre as perturbacións das distintas observacións. Os estimadores obtidos serán lineais, non sesgados e óptimos ademais de consistentes.

A fórmula do estimador MCX é: $b_{MCX} = (X'\Omega^{-1}X)^{-1} X'\Omega^{-1}Y$

Método de variables instrumentais

Termo en inglés: **Instrumental Variable Method**

O método de Variables Instrumentais é un método de estimación dos modelos econométricos que permite obter estimadores consistentes cando as variables explicativas están correlacionadas coa perturbación.

Un caso no que pode ser empregado é nos modelos autorregresivos nos que existe un regresor endóxeno retardado (e, polo tanto, estocástico). Se a perturbación do modelo non é Ruído Branco hai un problema de correlación contemporánea entre o regresor estocástico e a perturbación (ϵ). Nestes modelos a estimación por MCO non é válida pois os estimadores obtidos deixan de ter boas propiedades estatísticas. Neste caso a estimación polo método de Variables Instrumentais permite obter como resultado estimadores consistentes.

O método de Variables Instrumentais baséase en substituír parcialmente o regresor estocástico correlacionado coa perturbación aleatoria por outra variable con información similar á contida no regresor endóxeno pero non correlacionada coa perturbación. Este tipo de variables son denominadas variables instrumentais ou instrumentos. No caso dos modelos autorregresivos adóitase utilizar como instrumento o retardo da variable esóxena como variable instrumental da endóxena retardada.

Microeconomía

Termo en inglés: **Microeconomics**

Segundo Cameron e Trivedi (2005)*, a análise microeconómica refírese á “análise de datos a nivel individual sobre o comportamento económico de individuos e empresas”.

Trátase dun campo de estudo dentro da economía que aplica os métodos econométricos para contrastar diferentes hipóteses e situacións relacionadas coas decisións que toman os axentes microeconómicos tales como individuos e empresas. Combinando as ferramentas propias da econometría e o coñecemento microeconómico, a microeconomía axuda a estimar posibles relacións entre variables habituais no día a día dos axentes económicos e as súas posibles interrelacións.

* Cameron A.C. e Trivedi, P.K. (2005): *Microeconometrics: Methods and Applications*. Cambridge: Cambridge University Press.

Modelo autorregresivo de medias móviles (ARMA)

Termo en inglés: **Autoregressive Moving Average Model**

O modelo autorregresivo de media móbil, ARMA(p,q) é un proceso estocástico estacionario que ten unha compoñente autorregresiva, AR(p) e outra compoñente de media móbil MA(q). Ao ser estacionario o polinomio autorregresivo debe cumprir as condicións de estacionariedade, mentres que a compoñente de media móbil debe cumprir a condición de invertibilidade.

Formalmente, podemos expresar o modelo ARMA (p,q) como:

$$\begin{aligned} Y_t &= \delta + \phi_1 Y_{t-1} + u_t - \theta_1 u_{t-1} \\ (1 - \phi_1 L) Y_t &= \delta + (1 - \theta_1 L) u_t \\ \Phi(L) Y_t &= \delta + \Theta(L) u_t \end{aligned}$$

sendo u_t unha variable ruído branco.

Modelo autorregresivo integrado de media móbil (ARIMA)

Termo en inglés: **Autoregressive Integrated Moving Average Model (ARIMA)**

O modelo autorregresivo integrado de media móbil, ARIMA, é un modelo que trata de recoller o comportamento dunha serie temporal e predicir os seus valores futuros. Ten en conta a dependencia existente entre os datos, é dicir, cada observación nun momento dado é modelizado en función de valores anteriores.

Os modelos ARIMA están formados por unha parte autorregresiva (AR) outra parte de media móbil (MA) e, a diferenza dos modelos ARMA, son aplicados ás series temporais que amosan evidencia de falta de estacionariedade pero nos que a toma de diferenzas pode eliminar a non estacionariedade. A este tipo de procesos se lles denomina procesos integrados.

Modelo de axuste parcial

Termo en inglés: **Partial Adjustment Model**

Os modelos econométricos dinámicos poden xerarse por hipóteses de comportamento dos axentes económicos. Neste contexto, o modelo de axuste parcial, introducido por Nerlove (1958)*, supón que non existen incertezas acerca do futuro pero si rixidez no comportamento.

Supoñendo que hai un valor de equilibrio óptimo a longo prazo Y_t^* , dito valor pode expresarse en función das condicións esóxenas. No modelo de Nerlove, supónse que o nivel de capital desexado (Y^*) é unha función lineal da produción corrente (X).

$$Y_t^* = \beta_0 + \beta_1 X_t + \epsilon_t \quad (1)$$

Dado que Y^* é inobservable, Nerlove postula que calquera cambio observado na existencia de capital en calquera momento do tempo t é unha fracción do cambio desexado nese período. É dicir:

$$Y_t - Y_{t-1} = \delta (Y_t^* - Y_{t-1}) \quad (2)$$

Sendo δ o coeficiente de axuste ($0 < \delta < 1$).

Substituíndo (1) en (2) obteríamos o modelo autorregresivo que finalmente se estima.

* Nerlove, M. (1958): "Distributed lags and Estimation of Long-Run Supply and Demand Elasticities: Theoretical Considerations". *American Journal of Agricultural Economics*, vol. 40 (2), 301-310.

Modelo de elección binaria

Término en inglés: **Binary Choice Model**

Os modelos de elección binaria tratan problemas asociados á toma de decisións cando os axentes económicos afronta un proceso de elección binaria. Esta elección depende da probabilidade asociada a cada unha das alternativas posibles que pode ter un individuo.

Neste modelos asúmese que a elección depende de que a utilidade que obtén un individuo supere á que lle proporciona a alternativa. Esta utilidade depende dos valores que toman as características do axente económico e da opción a elixir (son os regresores do modelo), representadas pola combinación lineal $X_i\beta=Z_i$.

Exemplo: a modelización dun banco da decisión de conceder ou non un préstamo seguindo como criterio para concedelos ou non a probabilidade de devolución do cliente.

Modelo de probabilidade lineal

Término en inglés: **Lineal Probability Model**

O modelo de probabilidade lineal denota un modelo de regresión no que a variable dependente é binaria ou dicotómica, é dicir, toma unicamente os valores 1 ou 0 para indicar a ocorrencia ou non dun determinado evento.

Dado que Y é unha variable binaria

$$E(Y|X_1, \dots, X_k) = \Pr(Y=1|X_1, \dots, X_k),$$

e o Modelo de probabilidade lineal sería:

$$\text{Prob}(y=1|x_1, x_2, \dots, x_k) = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \epsilon_t$$

Nestes modelos os coeficientes da regresión, β_p , indican o cambio na probabilidade de que $Y=1$ asociado a un cambio unitario no regresor X_p , mantendo constantes os outros regresores.

Modelo de regresión

Término en inglés: **Regression Model**

Un modelo de regresión é un modelo que permite describir como inflúe unha variable X sobre outra variable Y.

X: Chámasele variable independente, variable explicativa ou regresor.

Y: Chámasele variable dependente, variable explicada ou regresando.

O obxectivo con este modelo é obter estimación razoables de Y para os distintos valores de X a partir de unha mostra de T pares de valores $(x_1, y_1), \dots, (x_T, y_T)$.

Na econometría, esta relación suponse estocástica, non determinística, e podemos expresala como: $y_t = f(x_t) + \epsilon_t$ sendo ϵ a perturbación ou erro aleatorio descoñecido e $f(x)$ a función que relaciona as variables dependente e independente.

Modelo de regresión lineal clásico

Término en inglés: **Classical Linear Regression Model**

Trátase dun modelo econométrico que intenta explicar a variable dependente y en función de un conxunto de variables independentes x_i ($i=1,2,\dots,k$), e de un termo aleatorio, a perturbación aleatoria ϵ , segunda a seguinte ecuación:

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + \epsilon_t$$

asumindo as seguintes hipóteses:

En relación a ε , esta compórtase como unha variable ruído branco, é dicir:

- $E(\varepsilon_t) = 0$ para todo t
- $V(\varepsilon_t) = \sigma$ para todo t
- $Cov(\varepsilon_t, \varepsilon_s) = 0$ para todo t, s con $t \neq s$

En notación matricial, estas dúas últimas hipóteses supoñen que a matriz de varianzas-covarianzas da perturbación aleatoria é unha matriz escalar.

En relación aos regresores:

- Os regresores son non estocásticos
- A matriz de regresores ten rango pleno

Ademais requirimos que o número de observación sexa maior que o número de parámetros a estimar.

Supoñemos que non existen erros de especificación.

Modelo de regresión lineal xeralizado

Termo en inglés: **Generalized Linear Regression Model**

É un modelo econométrico que podemos especificar como

$$y_t = \beta_0 + \beta_1 x_{1t} + \dots + \beta_k x_{kt} + \varepsilon_t$$

E no que se cumpre as seguintes hipóteses:

- Esperanza matemática da perturbación nula $E(\varepsilon_t) = 0$ para todo t
- Matriz de varianzas-covarianzas da perturbación non escalar, $V(\varepsilon) = \sigma \Omega$, onde Ω é unha matriz cadrada, definida positiva. Nun modelo xeralizado asúmese para a perturbación a hipóteses de heterocedasticidade e/ou autocorrelación.
- Matriz de regresores non estocástica (substituíble por matriz estocástica e independente da perturbación).
- Hipótese de rango pleno $Rg(X) = k+1$

O número total de observacións será maior que o número total de parámetros $T > k+1$ Supoñemos que non existen erros de especificación.

Modelo de retardos distribuídos

Termo en inglés: **Distributed Lag Models**

Son modelos de regresión dinámicos que se caracterizan por incluír como regresores os valores das variables explicativas actuais e os seus valores retardados un ou varios períodos de tempo. Polo tanto os efectos dos cambios en x prodúcense ao longo de varios períodos de tempo. Dentro deste tipo de modelo podemos distinguir dous subtipos:

1. Modelo de retardos distribuídos finitos: Neste tipo de modelo existe un número de retardos finito.

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_n x_{t-n} + \varepsilon_t$$

2. Modelo de retardos distribuídos infinitos: Neste tipo de modelo, non existe un número finito de retardos a introducir.

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \varepsilon_t$$

Modelo de retardos xeométricos / Modelo de Koyck

Termo en inglés: **Geometric Lag Model/Koyck Model**

Nun modelo con retardos distribuídos infinitos,

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \epsilon_t$$

a hipótese de Koyck formula que o valor dos parámetros das variables esóxenas retardadas, β_i , decrecen xeometricamente segundo a seguinte expresión:

$$\beta_i = \beta_0 \lambda^i, \quad 0 < \lambda < 1$$

Supón que a medida que nos referimos a períodos máis distantes no tempo, menor é a influencia de X sobre Y. A λ chámase coeficiente de diminución e a velocidade a que diminúen os retardos é $1-\lambda$.

Baixo este suposto, o modelo de retardos distribuídos infinitos pódese transformar no modelo autorregresivo de primeira orde:

$$y_t = \beta_0 x_t + \lambda y_{t-1} + \epsilon_t^*$$

O método de estimación que debe ser empregado dependerá da relación existente entre a endóxena retardada e a perturbación aleatoria do modelo, que neste modelo se deduce é un MA(1).

Modelo econométrico

Termo en inglés: **Econometric Model**

Un modelo econométrico é unha representación formal do coñecemento relativo a un fenómeno económico construída para cuantificar e contrastar as relacións entre variables, postuladas polos modelos económicos, a partir da evidencia empírica. Porén, tamén é posible establecer modelos econométricos non vinculados con ningunha teoría económica específica.

Esta formado por: variables endóxenas (regresando, explicadas polo modelo), variables explicativas (regresores, inflúen no comportamento das endóxenas), perturbación aleatoria (parte non sistemática do modelo, expresa a diferenza entre o verdadeiro valor da variable explicada e o seu valor esperado de acordo co modelo) e parámetros (factores de ponderación correspondentes a cada variable explicativa, que miden os efectos das fluctuacións destas sobre a variable endóxena). O modelo require igualmente definir as ecuacións e a súa forma funcional.

Modelo econométrico autorregresivo

Termo en inglés: **Autoregressive Econometric Model**

Modelo econométrico no que o valor actual da variable endóxena depende non só dos valores das esóxenas, senón tamén de observacións pasadas da variable de interese estudada, é dicir, a endóxena retardada (Y_{t-1}). A inclusión da endóxena retardada supón que o cambio nas variables esóxenas producen efectos infinitos sobre a endóxena actual.

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 y_{t-1} + \epsilon_t$$

Se queremos estudar as vendas dunha empresa no momento t, entenderemos que dependen da variable esóxena “prezo do produto” no momento t pero tamén do prezo do produto nos períodos anteriores, que determinarán as vendas nos períodos anteriores. En definitiva no modelo econométrico introduciríamos como regresores o prezo do produto no momento t e o valor das vendas retardado un período.

Modelo econométrico dinámico

Termo inglés: **Dynamic Econometric Model**

Estudan a relación entre a variable endóxena (Y) e as variables esóxenas (X) pero en distintos momentos do tempo. É dicir, a diferenza dos modelos estáticos nos que a cada conxunto de variables esóxenas lle correspondería un conxunto único de valores actuais das endóxenas, nos modelos dinámicos, estes valores actuais das endóxenas non só dependen dos valores das esóxenas, senón do “estado” anterior da economía, representado nos modelos econométricos coa incorporación de retardos nas variables esóxenas.

Así pois, o que caracteriza aos modelos dinámicos é que as reaccións aos estímulos económicos non se producen nin inmediatamente nin de forma instantánea.

Modelo econométrico estático

Termo en inglés: **Static Econometric Model**

Un modelo econométrico dise estático cando todas as súas variables explicativas están referidas ao mesmo período de tempo que o regresando (t).

$$y_t = \beta_0 + \beta_1 x_t + \beta_2 x_{2t} + \epsilon_t$$

Isto significa que unha variación no regresor (x_t) só lle afecta a y nese mesmo período de tempo, o resto dos períodos non se ven afectados por ese cambio (existe só un efecto inmediato, contemporáneo, que non se propaga a outros períodos).

Modelo econométrico lineal-logarítmico

Termo en inglés: **Lineal-logarithmic Model**

O modelo lineal-logarítmico é unha función de regresión na que a variable independente (X) é logarítmica, é dicir:

$$y_t = \beta_0 + \beta_1 \log(x_t) + \epsilon_t$$

O modelo lineal-logarítmico interprétase como: o incremento do 1% no regresor X provocará un cambio en Y de $0,01 \cdot \beta_1$

Exemplo:

Relacionamos as notas académicas obtidas cunha variable de renda da familia do alumno/a. A ecuación estimada sería:

$$\text{Cualificación}_t = 3 + 1,25 \cdot \text{Log}(\text{Renda familiar})$$

En relación a esta ecuación podemos dicir que ante un aumento do 1% da renda familiar dáse un aumento de $0,01 \cdot 1,25$ puntos nas cualificacións, é dicir, 0,0125 ptos.

Modelo econométrico logarítmico-lineal (forma funcional exponencial)

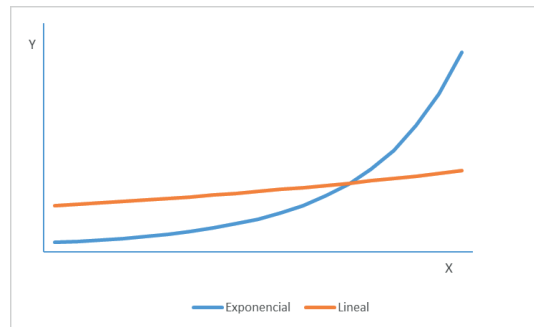
Termo en inglés: **Logarithmic Linear Model (Exponential Form)**

Cando se considera que a variable de interese crecerá exponencialmente a relación coas variables explicativas debería modelizarse cunha forma funcional exponencial:

$$y_t = e^{(\beta_0 + \beta_1 x_t + \epsilon_t)}$$

Graficamente:

Gráfico 11. Relación funcional exponencial



Fonte da imaxe: Elaboración propia

Para estimar mediante MCO este tipo de modelos, a aproximación habitual é linealizalo tomando logaritmos a ambos lados da ecuación, dando lugar ao modelo logarítmico-lineal (log-linear):

$$\log(y_t) = \beta_0 + \beta_1 x_t + \epsilon_t$$

Nesta forma funcional, β_1 indícanos a semielasticidade de Y respecto a X, e indica (multiplicado por 100) a variación porcentual de Y cando X aumenta nunha unidade. O modelo logarítmico lineal é un modelo semilogarítmico.

Modelo econométrico log-log

Término en inglés: **Log-Log Model**

O modelo econométrico dobremente logarítmico (ou log-log) é un modelo lineal cunha forma funcional non lineal no que tanto a variable dependente como as variables independentes están expresadas en logaritmos.

Modelo de regresión log-log

$$\log(y_t) = \beta_0 + \beta_1 \log(x_t) + \epsilon_t$$

Variable dep. (Y)	Variable Indep. (X)	Interpretación do parámetro
Log(y)	Log(x)	$\% \Delta Y = \beta_1 \% \Delta X$

Nestes modelos, a pendente β_1 mide a elasticidade de Y con respecto a X, isto é a variación porcentual de Y ante un cambio porcentual de X.

Un exemplo de grande interese para a rama económica é a elasticidade-prezo da demanda que viría dada por β_1 se Y representa a cantidade demandada e X o prezo.

Modelo econométrico multiecuacional

Término en inglés: **Multiple-equation econometric model**

Os modelos multiecuacionais son modelos cuxa especificación implica a máis dunha variable endóxena, tendo en consecuencia máis dunha ecuación. Este tipo de modelos poden ser de dous tipos:

Con interdependencia: Cando as variables endóxenas son a súa vez variables explicativas noutra ou noutras ecuacións do modelo

$$\begin{aligned} y_{1t} &= \beta_0 + \beta_1 x_{1t} + \beta_2 y_{2t} + \epsilon_{1t} \\ y_{2t} &= \delta_0 + \delta_1 y_{1t} + \delta_2 x_{2t} + \epsilon_{2t} \end{aligned}$$

Sen interdependencia: Son modelos que non presentan interdependencia no mesmo momento de tempo:

$$\begin{aligned} y_{1t} &= \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_{1t} \\ y_{2t} &= \delta_0 + \delta_1 y_{1t} + \delta_2 x_{2t} + \epsilon_{2t} \end{aligned}$$

Modelo econométrico non lineal

Término en inglés: **Non linear models**

Un modelo defínese como non lineal se o efecto dun cambio unitario na variable explicativa sobre a variable dependente non é constante. Dentro dos modelos non lineais, os modelos que son non lineais nas variables, pero seguen sendo lineais nos parámetros non requiren procedementos especiais de estimación. Poden converterse en modelos lineais mediante transformacións matemáticas e se lles denomina modelos non lineais linealizables. Seguen tendo, desde esta perspectiva, a consideración de modelos lineais.

Porén, se o modelo non é lineal en parámetros si precisarían métodos específicos de estimación.

Un exemplo de este tipo de modelo, é a función de produción de elasticidade de substitución constante (CES):

$$Y = A [\delta K^{-\beta} + (1 - \delta) L^{-\beta}]^{-\nu} / \beta$$

Y : Produción

K : Capital

L : Recurso de traballo

A : Eficiencia do estado da tecnoloxía (positivo, si $A > 0$)

δ : Participación de cada factor de produción (relativas, cando $0 < \delta < 1$)

β : Substitución

ν : Rendementos a escala (constantes, se se cumpre $\nu = 1$)

Modelo econométrico uniecuacional

Término en inglés: **Single-equation econometric model**

Na clasificación dos modelos econométricos referímonos aos modelos uniecuacionais como aqueles nos que só hai unha variable endóxena corrente, que é a que figura no lado esquerdo da ecuación, polo tanto o modelo está conformado unicamente por unha ecuación. Os modelos uniecuacionais tamén poden clasificarse en :

Simple: Teñen unicamente unha variable explicativa.

$$y_t = \beta_0 + \beta_1 x_{1t} + \epsilon_t$$

Múltiple: Teñen máis dunha variable explicativa.

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \epsilon_t$$

Multicolinealidad

Término ao inglés: **Multicollinearity**

Consiste na existencia de relacións lineais entre dúas ou máis variables explicativas do modelo econométrico.

- Dirase que a multicolinealidad é perfecta cando ditas relacións lineais sexan exactas, incumpríndose a hipótese de rango pleno ($\text{rg}(X) < k+1$) xa que non todos os x_i son linealmente independentes.
- Dirase que existe multicolinealidad elevada se algunha variable está altamente relacionada con outra ou outras. Neste caso, manteríase a hipótese de rango pleno aínda que tería importantes consecuencias negativas sobre os resultados da estimación, pola obtención de estimadores con desviacións típicas estimadas moi elevadas. Este é o problema que se debe afrontar.

Exemplo:

x_1 : Produción agraria de España nun período.

x_2 : Litros de auga por m^2 caída en España nese mesmo período.

x_3 : Auga a disposición para regadíos en varios momentos do ano nese mesmo período.

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \beta_3 x_{3t} + \varepsilon_t$$

Neste exemplo poderíamos ter un problema de multicolinealidade elevada xa que a variable x_2 estaría condicionando a variable x_3 .

Multiplicador de impacto

Termo en inglés: **Impact multiplier or short run multiplier**

Nos modelos con retardos nas variables esóxenas, modelos con retardos distribuídos, os coeficientes interprétanse como multiplicadores. No modelo:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_n x_{t-n} + \varepsilon_t$$

o coeficiente β_0 que acompaña á variable esóxena x_t (sen retardos) medirá o cambio, en promedio, na variable endóxena y_t que provoca un incremento nunha unidade da variable esóxena x_t no mesmo momento de tempo. Será o multiplicador de impacto. Polo tanto, o multiplicador de impacto cuantifica o efecto a curto prazo que se produce na endóxena ante os cambios unitarios da esóxena.

Nesgo dun estimador

Termo en inglés: **Bias of an estimator**

Chámase nesgo dun estimador, á diferenza entre a esperanza matemática (valor esperado) de dito estimador e o verdadeiro valor do parámetro que se estima.

Sexa un estimador T dun parámetro θ , o nesgo exprésase como:

$$\text{nesgo}(T) = E(T) - \theta$$

O nesgo do estimador pode ser nulo e o estimador diríase non nesgado ou centrado, é dicir, T sería un estimador non nesgado cando:

$$\text{nesgo}(T) = E(T) - \theta = 0 \rightarrow E(T) = \theta$$

Nivel de confianza

Termo en inglés: **Confidence level**

Na estimación por intervalo, o nivel de confianza refírese a porcentaxe dos intervalos que podemos esperar que conteñan o verdadeiro valor do parámetro, tomando todas as mostras posibles dunha poboación. Por exemplo, se supoñemos que seleccionamos todas as posibles mostras da mesma poboación, e calculamos o intervalo de confianza para cada mostra, un nivel de confianza do 95% implica que o 95% dos intervalos de confianza incluírían o verdadeiro valor do parámetro. Indícase por $1-\alpha$ e habitualmente dáse en porcentaxe $(1-\alpha)100\%$. Falamos de nivel de confianza xa que unha vez extraída a mostra, o intervalo de confianza calculado conterá ou non o verdadeiro valor do parámetro o no, o que sabemos é que de repetir o proceso con todas as mostras posibles, poderíamos afirmar que o $(1-\alpha)\%$ dos intervalos así construídos conterían o verdadeiro valor del parámetro.

Operador de retardos

Termo en inglés: **Lag operator**

Na análise de series temporais, o operador de retardos L utilízase para retardar unha variable en h períodos. $L^h x_t = x_{t-h}$

Exemplo:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \varepsilon_t$$

E utilizando o operador de retardos L :

$$y_t = \alpha + \beta_0 x_t + \beta_1 Lx_t + \varepsilon_t = \alpha + (\beta_0 + \beta_1 L)x_t + \varepsilon_t$$

Ao operador de retardos tamén se lle denota pola letra B .

Operadores nas ecuacións ARIMA

Termos en inglés: **Lag operator, differences operator**

Para escribir correctamente a ecuación correspondente a un modelo ARIMA, temos que coñecer os operadores que temos que empregar na súa modelización.

Os operadores a empregar son

- Operador de retardos L : este operador retarda a variable á que acompaña “ h ” períodos

$$L^h Y_t = Y_{t-h}$$

- Operador de diferenzas Δ : transforma unha variable tomando diferenzas de orden “ d ”

$$\Delta^d Y_t = (1-L)^d Y_t$$

Orde de integración

Termo en inglés: **Order of integration**

Cando se traballa con series temporais en ocasións as series son non estacionarias. Neste contexto, a orde de integración dunha serie (non estacionaria) refírese ao número de veces que é necesario diferenciar dita serie para convertela en estacionaria, dando lugar as series integradas.

No enfoque de Box-Jenkins, para a elaboración dun modelo ARIMA, se é necesario diferenciar dúas veces o proceso para convertelo en estacionario diríamos que a súa orde de integración é dous e o denotamos como $I(2)$. Este concepto tamén é crucial na análise da cointegración entre series temporais que se relaciona nun modelo econométrico.

Ordenada na orixe

Termo en inglés: **Intercept**

A ordenada na orixe nun modelo de regresión lineal indica o valor esperado da variable dependente cando todas as variables independentes son iguais a cero

Exemplo: [Neste vídeo](#) da Khan Academy explícase a interpretación da ordenada na orixe.

P-valor

Termo en inglés: **P-value**

Tamén denominado p -value (directamente do inglés), defínese como a probabilidade de que un valor estatístico calculado en base a datos mostrais sexa posible dada una hipótese nula certa. Axuda a diferenciar resultados que son produto do chou do mostreo de resultados que son estatisticamente significativos. Se o valor p cumpre coa condición de ser menor que un nivel de significación establecido arbitrariamente, este considérase como un resultado estatisticamente significativo e, polo tanto, permite rexeitar a hipótese nula.

No caso habitual en Econometría no que manexamos un nivel de significación do 5%, un valor p menor que 0,05 é un valor pequeno, o cal indicará unha forte evidencia contra a hipótese nula, polo que vaise rexeitar dita hipótese. Un valor p maior a 0,05 indica unha evidencia mínima contra a hipótese nula, polo que non se poderá rexeitar.

Parámetro

Término en inglés: **Parameter, Coefficient**

Nun modelo econométrico, un parámetro (β_i) é un factor que mide a influencia que teñen as variábeis explicativas (regresores) sobre a explicada (regresando). Os parámetros son valores constantes e descoñecidos que se tratarán de estimar a través de métodos de estimación do modelo. β_i indica canto aumenta ou diminúe y_t cando x_i aumenta nunha unidade, e o resto se mantén constante (*ceteris paribus*).

Paseo aleatorio

Término en inglés: **Random Walk**

Un paseo aleatorio é un proceso estocástico non estacionario que se caracteriza porque a súa primeira diferenza é un proceso estacionario Ruído Branco. É dicir:

$$y_t = y_{t-1} + u_t$$

Sendo u_t proceso Ruído Branco.

Poderíamos consideralo como un AR(1), $y_t = \phi_1 y_{t-1} + u_t$, cun coeficiente $\phi_1 = 1$, o que supón un incumprimento da condición de estacionariedade, polo que o proceso non é estacionario. A súa varianza faise infinita e indefinida e as autocovarianzas tamén aumentan ao longo do tempo.

Perturbación aleatoria

Término en inglés: **Random disturbance, random error**

A perturbación aleatoria é un elemento non observable do modelo econométrico que intenta recoller efectos difíciles de especificar de maneira individual, pero que deben terse en conta. A existencia da perturbación aleatoria xustifícase pola existencia de múltiples pequenos efectos que afectan ao comportamento da variable dependente, pola existencia de factores nalgúns casos difíciles de medir ou sen datos dispoñibles, e pola inherente aleatoriedade do comportamento humano, que está na base das relacións económicas.

Exemplo: Estúdase o gasto familiar en función dos ingresos da familia. Efectos como os gustos, pautas culturais, variacións no tempo, modas... afectan ao modelo pero de maneira pouco relevante e ademais moi difíciles de cuantificar, polo que estes elementos estarán recollidos na perturbación aleatoria.

Perturbación aleatoria AR(1)

Término en inglés: **AR(1) random disturbance or AR(1) random error**

Dicimos que nun modelo econométrico a perturbación é AR(1), é dicir, presenta autocorrelación de primeira orde cando a perturbación aleatoria asociada ao momento t depende da perturbación aleatoria no momento inmediatamente anterior máis un variable ruído branco. É máis frecuente en modelos con datos temporais anuais.

A autocorrelación AR(1) modelízase tal que:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

sendo u_t unha variable Ruído Branco e ρ é o coeficiente de autocorrelación.

Polinomio de retardos distribuídos de Almon

Termo en inglés: **Almon's polynomial distributed lag**

Na estimación de modelos dinámicos con retardos distribuídos finitos, a proposta de Almon formula que os parámetros do modelo orixinal pódense aproximar a través das ordenadas dunha función polinómica de grao “q”, non moi elevado, e en calquera caso menor que “n”, sendo n o número de retardos. Se denota como PDL, polinomio de retardos distribuídos.

Formalmente:

Sexa o modelo dinámico de retardos distribuídos finitos:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + \dots + \beta_n x_{t-n} + \varepsilon_t$$

a proposta do PDL de Almon supón:

$$\beta_i = \alpha_0 + \alpha_1 i + \alpha_2 i^2 + \dots + \alpha_q i^q$$

Con $n > q$

Predición

Termo en inglés: **Forecast**

A predición dunha variable económica é unha estimación do seu valor futuro, baseada na información pasada e presente, información que está contida nun modelo. Hai diversas técnicas de predición en econometría: mediante modelos estruturais, predición univariante etc. As predicións serven como guías para a adopción de políticas e é igualmente importante como “liña mestra” para a construción e verificación dos modelos econométricos.

Tamén pode referirse a obtención do valor da variable dependente para observacións que non forman parte das observación mostrais, “out-of-sample”.

Proceso autorregresivo de orde p, AR(p)

Termo en inglés: **Autoregressive process**

Un proceso autorregresivo de orde p, AR(p), é un tipo de proceso estocástico estacionario utilizado para a predición que se caracteriza por expresar a variable Y_t en función de valores da variable asociados a momentos anteriores ata Y_{t-p} e unha variable Ruído Branco u_t do momento actual. A orde do proceso virá dada polos retardos de Y_t incorporados

$$Y_t = \delta + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \dots + \phi_p Y_{t-p} + u_t$$
$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p) Y_t = u_t$$

sendo $(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)$ o polinomio autorregresivo de orde p.

Os procesos autorregresivos precisan que se cumpran unhas determinadas condicións para que sexan estacionarios. Estas condicións teñen que ver coas raíces do polinomio autorregresivo.

Exempo: AR(1).

$$Y_t = \delta + \phi_1 Y_{t-1} + u_t$$
$$(1 - \phi_1 L) Y_t = u_t$$

Neste proceso a condición de estacionariedade require que $|\phi_1| < 1$

Para a súa identificación teranse en conta os coeficientes de autocorrelación e os coeficientes de autocorrelación parciais. Os coeficientes de autocorrelación declinan sen chegar a ser 0 e os coeficientes de autocorrelación parcial tomarán un valor alto nos p primeiros coeficientes para anularse no resto.

Proceso ergódico

Término en inglés: **Ergodic process**

No estudo das series temporais, un proceso estocástico dise ergódico se os seus momentos poden estimarse consistentemente a partir dos correspondentes momentos mostrais.

Sabendo que o coeficiente de autocorrelación mide a relación lineal entre dúas variables do proceso, unha condición necesaria para que o proceso sexa ergódico é que os coeficientes de autocorrelación, a medida que nos afastamos no tempo, se aproximen a cero, é dicir:

$$\lim_{k \rightarrow \infty} \rho_k = 0$$

De forma intuitiva podemos dicir que un proceso é ergódico se, a medida que nos afastamos no tempo ($k \rightarrow \infty$), a relación da variable cos seus valores pasados é cada vez máis pequena, é dicir a correlación diminúe.

Na elaboración de modelos ARIMA non ámbito da economía, unha das condicións que se establecen é que o proceso sexa ergódico. Tamén se require que o proceso sexa estacionario.

Proceso estocástico

Término en inglés: **Stochastic process**

Un proceso estocástico ou aleatorio é un conxunto de variables aleatorias que corresponden a momentos sucesivos de tempo. Formalmente: $\{Y_t\} = \{Y_{t_1}, Y_{t_2}, Y_{t_3}, \dots, Y_{t_N}\}$

Para cada instante de tempo, $t=t_i$, un proceso estocástico é unha única variable aleatoria cunha determinada distribución de probabilidade. A descrición das características do proceso estocástico pódese facer mediante dúas formas alternativas:

1. A partir das funcións de distribución conxuntas. Trataríase de determinar as funcións de distribución conxuntas par cada conxunto finito de variables do proceso
2. A partir dos momentos de primeira e segunda orde. É unha caracterización máis incompleta, pero se o proceso é normal, quedaría perfectamente caracterizado e falaríamos de proceso Gaussiano.

Proceso estocástico estacionario

Término en inglés: **Stationary stochastic process**

Sabendo que os procesos estocásticos son un conxunto de variables aleatorias que corresponden a momentos sucesivos de tempo, definimos os procesos estocásticos estacionarios como procesos estocásticos nos que a distribución de probabilidade nun instante de tempo fixo e a mesma para todos os instantes de tempo. Nun sentido máis incompleto podemos definir os procesos estocásticos como estacionarios se :

1. A media do proceso mantense estable ao longo do tempo. $E[Y] = \mu$ para todo t .
2. A varianza mantense estable ao longo do tempo $\text{Var}[Y] = \sigma^2 < \infty$ para todo t .
3. A autocovarianza depende unicamente do desfase temporal entre as dúas variables do proceso: $\text{Cov}(Y_t, Y_{t-k}) = \gamma_k$ para calquera t .

Proceso integrado

Término en inglés: **Integrated process**

Na análise estocástica de series temporais, un proceso integrado é aquel que, non sendo estacionario, pode converterse en estacionario aplicando diferenzas. Definimos a orde de integración do proceso como o número de diferenzas precisas para facelo estacionario.

Por exemplo: se o proceso estocástico Y_t non é estacionario pero o proceso Z_t das segundas diferenzas si o é, entón diremos que Y_t é un proceso “integrado de orde dous” denotado por $I(2)$.

$$Z_t = \Delta^2 Y_t = (1-L)^2 Y_t = Y_t - 2Y_{t-1} + Y_{t-2}$$

Un proceso estocástico estacionario denomínase a miúdo un proceso $I(0)$.

Pseudo-R² de McFadden

Térmo en inglés: **McFadden pseudo-R**

O R^2 proposto por McFadden (tamén coñecido como pseudoR de McFadden) emprégase na interpretación dos modelos de elección binaria. Defínese como:

$$R_{\text{McFadden}}^2 = \text{PseudoR}^2 = 1 - \frac{\text{LnL}_{\text{SR}}}{\text{LnL}_{\text{CR}}}$$

Sendo:

LnL_{CR} o valor (máximo) do logaritmo da función de verosimilitud que se obtén baixo a hipótese nula de modelo con restricións, que inclúe unicamente a constante, é dicir: $H_0: \beta_1 = \beta_2 = \dots = \beta_k = 0$

LnL_{SR} o valor (máximo) do logaritmo da función verosimilitud que se obtén baixo a hipótese alternativa de modelo sen restricións:

Se as variables explicativas non teñen ningún valor explicativo, entón a ratio sería igual á unidade e o valor do pseudoR sería cero.

Raíz do erro cadrático medio (RECM)

Térmo en inglés: **Root Mean Squared Error (RMSE)**

A raíz do erro cadrático medio é unha medida da bondade do axuste. Defínese como a raíz cadrada da suma de cadrados de erros entre T .

$$\text{RECM} = \sqrt{\text{SCE}/T}$$

Habitualmente exprésase en porcentaxe (%RECM) respecto a media aritmética de y :

$$\% \text{RECM} = \frac{\text{RECM}}{\bar{Y}} * 100 = \frac{\sqrt{\text{SCE}/T}}{\bar{Y}} * 100\%$$

De este modo a %RECM indica en porcentaxe o erro que se está a cometer ao facer a estimación respecto a un valor medio do regresando. Se este toma un valor igual ou próximo a cero será un bo axuste.

Raíz unitaria

Térmo en inglés: **Unit Root**

Unha raíz unitaria é unha característica dos procesos temporais. Un proceso con raíz unitaria contén unha tendencia estocástica e non sería estacionario.

Existirá no proceso unha raíz unitaria si o valor 1 é unha raíz da ecuación característica do proceso. Neste caso o proceso é non estacionario.

Exemplo:

No proceso AR(1)

$$Y_t = 0,4Y_{t-1} + u_t$$

$$(1 - 0,4L)Y_t = u_t$$

A raíz da ecuación característica é menor que 1, en termos absoluto, e o proceso é estacionario (alternativamente, a raíz do polinomio autorregresivo $1-0,4L=0 \rightarrow |L|=|1/0,4|=2,5 > 1$ cae fora do círculo unitario)
 En cambio, no proceso

$$Y_t = Y_{t-1} + u_t$$

$$(1-L)Y_t = u_t$$

no polinomio $1-L=0$ temos que a raíz $L=1$, entón determinamos que existe unha raíz unitaria de forma que o proceso non é estacionario. Esta ecuación representa un paseo aleatorio.

Determinar si o proceso é ou non estacionario (si ten ou non raíz unitaria) é importante para a estimación, pois o uso de MCO en variables non estacionarias pode dar lugar a estimacións con parámetros erróneos na relación entre as variables (relacións espurias).

Regresión espuria

Termo en inglés: **Spurious Regression**

O problema da regresión espuria fórmase na modelización con series temporais. Ao analizar a relación entre dúas series temporais non estacionarias con tendencias estocásticas, as variables poden ter un comportamento similar no tempo e polo tanto pode aparecer entre elas unha relación matemática que en realidade non existe.

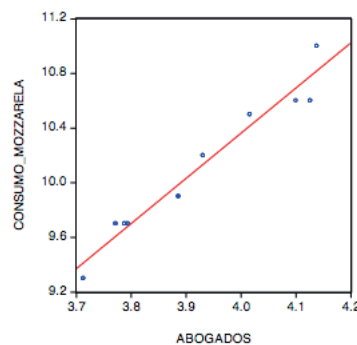
Ao presentar esta tendencia común, é moi posible que os modelos que relacionan ambas variables presenten unha elevada bondade do axuste (R^2 elevado).

Nestes modelos é frecuente atopar un valor do estatístico Durbin-Watson moi pequeno (dw -0), indicación que as perturbacións están fortemente autocorrelacionadas.

Exemplo: Con datos de 2000 a 2009 regresamos o consumo per capita de queixo mozzarella sobre o número de avogados en Hawaii.

Gráfico 12. Exemplo regresión espuria

Variable	Coefficient	Std. Error	t-Statistic	Prob.
ABOGADOS	3.306863	0.250211	13.21629	0.0000
C	-2.864067	0.983163	-2.913114	0.0195
R-squared	0.956205	Mean dependent var	10.12000	
Adjusted R-squared	0.950731	S.D. dependent var	0.541192	



Fonte da imaxe: Elaboración propia a partir dos datos extraídos de <http://tylervigen.com/spurious-correlations>

Esta relación non ten ningún sentido, sen embargo, como teñen un comportamento similar no tempo, parece que si están relacionadas. Non é unha verdadeira relación causa-efecto, se non unha relación espuria.

Regressor significativo

Termo en inglés: **Significant regressor**

Nun modelo econométrico dise que unha variable explicativa ou regresor é significativo ou ten significatividade se unha variación ou cambio nela supón un cambio na variable dependente. A significatividade estatística dun regresor analízase mediante o contraste de nulidade individual do parámetro que recolle o seu efecto. Son os chamados contrastes de significatividade dos coeficientes e se se rexeita a hipótese nula o coeficiente será estatisticamente distinto de cero e a variable explicativa terá un efecto estatisticamente significativo sobre o regresando.

Regresores estocásticos

Término en inglés: **Stochastic regressors**

Na modelización econométrica, é frecuente asumir o suposto de que os regresores son non estocásticos. Sen embargo, este suposto non é sempre plausible, hai regresores que son estocásticos por natureza. O termo regresor estocástico significa que os regresores, é dicir, as variables explicativas, son aleatorias. A consideración de aleatorios fai que a relación do regresor estocástico coa perturbación aleatoria sexa unha cuestión crucial á hora de determinar o método adecuado de estimación.

Exemplos de modelos nos que algún regresor ten necesariamente a consideración de estocástico :

- Modelos de ecuacións simultáneas
- Modelos autorregresivos
- Modelos con erros aleatorios nas variables explicativas.

Regresores non estocásticos

Término en inglés: **Non stochastic regressors**

Unha das hipóteses dos modelos econométricos refírese ao comportamento da matriz de regresores. Fálase de matriz de regresores non estocástica cando consideramos que os regresores son variables que toman valores fixos para distintas mostras do mesmo tamaño. Si isto se verifica entón no modelo hai incorrección total entre os regresores e a perturbación, e dicir, non hai relación entre os regresores e a perturbación.

Formalmente: $\text{Cov}(x_t, \varepsilon_{t'}) = 0$ para todo t e t'

Esta hipótese, certamente moi restritiva, pode ser substituída por regresores estocásticos pero independentes da perturbación aleatoria.

Retardo temporal

Término en inglés: **Lag**

Nos modelos econométricos dinámicos, denomínase “retardo” ao período de tempo que transcorre entre a causa e o efecto, é dicir na reacción na variable dependente ante un cambio producido en unha ou varias variables explicativas. Asímesse que este retardo está suxeito a unha temporalidade, polo que normalmente fálase de período.

Exemplo:

$$y_t = \alpha + \beta x_{t-3} + \varepsilon_t$$

Isto quere dicir que o cambio na variable X no momento “t” terá efecto sobre a variable dependente transcorridos tres períodos.

Ruído branco

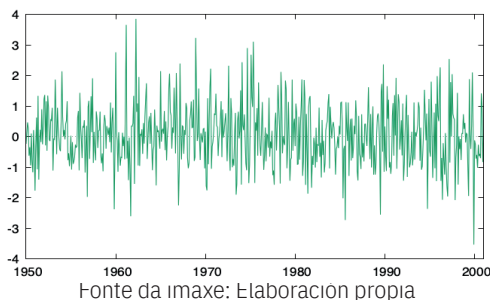
Término en inglés: **White noise**

Antes de falar de proceso “ruído branco” debemos entender o concepto de “proceso estocástico”. Un proceso estocástico é un conxunto ou familia de variables aleatorias ordenadas segundo outra variable que adoita ser o tempo. Neste contexto, un proceso estocástico “Ruído Branco” $A(t)$, será un proceso estocástico que se caracteriza por:

1. Esperanza matemática constante é igual a cero. $E[A_t] = 0$ para todo t .
2. Varianza constante e finita, $\text{Var}[A_t] = E[A_t]^2 = \sigma^2$, para todo t .
3. Covarianzas entre as variables aleatorias asociadas a distintos momentos de tempo nulas, $\text{Cov}[A_t, A_{t'}] = 0$ para todo $t \neq t'$.

Sendo o proceso gaussianos (distribución normal multivariante) as variables aleatorias do proceso serán independentes. No modelo de regresión clásico suponse que a perturbación aleatoria se comporta como “Ruído Branco”.

Na seguinte imaxe móstrase a representación gráfica correspondente a unha serie “Ruído Branco”.



Fonte da imaxe: Elaboración propia

Serie temporal

Término en inglés: **Time series**

Unha serie temporal é unha sucesión de observacións dunha variable referidos a distintos momentos sucesivos de tempo. Estas observacións están ordenadas e corresponden a instantes equidistantes de tempo.

Na metodoloxía ARIMA, unha serie temporal concíbese como unha realización dun proceso estocástico.

Exemplo:

No gráfico móstrase a evolución do IBEX35 ao longo do ano 2020 (ata 07-09-2020) o período de tempo comprendido entre 2013 ata 2017.



Fonte da imaxe: Elaboración propia a partir dos datos do IBEX ofrecidos por Google.

Serie temporal non estacionaria

Término en inglés: **Non-stationary time series**

Enténdese que un proceso estocástico non é estacionario cando a distribución de probabilidade conxunta cambia no tempo. En sentido menos completo, se a media e/ou a varianza depende de t , e/ou as autocovarianzas dependen tamén de t , o proceso estocástico (serie) sería non estacionario.

Formalmente:

$$E[Y_t] = \mu_t \text{ para todo } t \text{ e/ou}$$

$$\text{Var}[Y_t] = \sigma_t \text{ para todo } t \text{ e/ou}$$

$$\text{Cov}[Y_t, Y_{t+h}] \neq [Y_s, Y_{s+h}], t \neq s, \text{ para todo } t, s.$$

Suma de cadrados dos erros (SCE)

Termo en inglés: **Sum squared resid**

A SCE na estimación MCO dun modelo econométrico defínese como a suma dos cadrados dos residuos mínimo cadráticos ordinarios (e_t), é dicir as distancias entre os valores observados e axustados por MCO da variable dependente:

$$SCE = \sum_{t=1}^T e_t^2 = \sum_{t=1}^T (y_t - \hat{y}_t)^2 = \sum_{t=1}^T \left(y_t - (b_0 + b_1 x_{1t} + \dots + b_k x_{kt}) \right)^2$$

Tendencia

Termo en inglés: **Trend**

Na análise de series temporais, a tendencia, como concepto teórico, consiste no comportamento xeral ao longo prazo dunha serie temporal. Proporciona patróns ou pautas con respecto ao seu posible movemento no futuro.

De xeito analítico, na análise clásica de series temporais, trátase de illar esta compoñente para ver cal é o seu efecto na serie temporal. Para isto, utilízanse diversos métodos como poden ser o método do axuste analítico, o método de medias móbeis etc.

Teorema de Gauss-Markov

Termo en inglés: **Gauss-Markov theorem**

O teorema de Gauss-Markov permite demostrar que o estimador MCO (Mínimos Cadrados Ordinarios) nun modelo econométrico que cumpre os supostos asumidos no modelo de regresión lineal clásico (MRLC) serán os estimadores de menor varianza entre os lineais e inesgados (ELIO).

Test da razón de verosimellanza para heterocedasticidade

Termo en inglés: **Likelihood ratio test for heteroscedasticity**

Utilízase para analizar a heterocedasticidade por grupos. Neste caso, suponse que a mostra pódese dividir en “p” grupos de forma que cada grupo está formado polas observacións na que algunha variable explicativa toma valores similares ou iguais. O tamaño de cada grupo non ten porque ser o mesmo. O punto de partida é a heterocedasticidade entre grupos e a homocedasticidade dentro de cada grupo.

H_0 : Non existe heterocedasticidade por grupos

H_1 : Existe heterocedasticidade por grupos

Para realizar o test danse os seguintes pasos previos:

Dividir a mostra en “p” grupos e estimar por MCO co conxunto das T de observacións e coas observacións correspondentes a cada grupo i, T_i , con $i=1,2,\dots,p$.

A partir das estimacións realizadas calcúlase o estimador consistente da varianza da perturbación para o conxunto de observacións, $\hat{\sigma}^2 = \frac{SCE}{T}$, e para cada grupo de observacións $\hat{\sigma}_i^2 = \frac{SCE_i}{T_i}$

O estatístico de proba é:

$$\mu = -2 \ln(\lambda) \rightarrow \chi_{p-1}^2, \mu = -2 \left(\sum_{i=1}^p T_i \ln \hat{\sigma}_i - T \ln \hat{\sigma} \right)$$

Test de Breusch-Godfrey

Termo en inglés: **Breusch-Godfrey Test**

O test de Breusch-Godfrey sérvenos para analizar a autocorrelación nun modelo econométrico. Pertence á familia dos tests asintóticos coñecidos como tests do Multiplicador de Lagrange. Ao contrario que outros tests coñecidos como o Durbin-Watson, este test permite contrastar diversos tipos ou estruturas de autocorrelación.

As hipóteses son as seguintes:

H_0 : Non existe autocorrelación

H_1 : Existe autocorrelación ata de orde p AR(p) (ou outra estrutura como medias móbeis)

O test será $T \cdot R^2$ que, baixo hipótese nula certa, segue unha distribución χ_p^2 .

O R^2 obtense da regresión auxiliar dos residuos MCO (e_t) sobre as variables explicativas do modelo econométrico e os residuos MCO retardados ata p períodos.

Exemplo:

Nun modelo econométrico simple, para contrastar a existencia de AR(3) na perturbación aleatoria a regresión auxiliar a estimar sería:

$$e_t = \alpha_0 + \alpha_1 x_t + \delta_1 e_{t-1} + \delta_2 e_{t-2} + \delta_3 e_{t-3} + u_t$$

Test de Dickey-Fuller (DF)

Termo en inglés: **Dickey-Fuller test**

O test de Dickey-Fuller é un test de raíces unitarias. Dito test quere contrastar a non estacionariedade dunha serie temporal, é dicir, a existencia de raíz unitaria no polinomio autorregresivo dun paseo aleatorio.

Concretamente o test formula a seguinte ecuación:

$$\Delta Y_t = (\varphi - 1)Y_{t-1} + u_t = \gamma Y_{t-1} \phi + u_t$$

As hipóteses son as seguintes:

$H_0: \gamma = 0$ (existe raíz unitaria, serie non estacionaria)

$H_1: \gamma < 0$ (non existe raíz unitaria, serie estacionaria)

Para resolver o contraste débense utilizar as distribucións empíricas deste estatístico con valores críticos tabulados, que dependen do número de observacións e do nivel de significación co que se efectúa o contraste.

— Se $t(\gamma) < \tau$ rexeitase H_0 , proceso estacionario.

— Se $t(\gamma) > \tau$ non se rexeita a H_0 , proceso non estacionario.

Débese indicar que se o modelo contén deriva e/ou tendencia determinista, os valores críticos cos que se compara o ratio t asociado ao parámetro γ serán diferentes.

Test de Dickey-Fuller aumentado (ADF)

Termo en inglés: **Augmented Dickey-Fuller test**

O test ADF emprégase en econometría para analizar se unha serie temporal presenta ou non unha raíz unitaria. Este test é unha versión aumentada da proba Dickey-Fuller, lévase a cabo co mesmo procedemento. A diferenza entre ambos consiste en que o test ADF elimina da serie temporal os problemas de autocorrelación.

A ecuación para a realización do test máis xeral sería a seguinte (non en todos os casos a ecuación inclúe todos os elementos):

$$\Delta Y_t = \alpha + \tau \text{TREND}_t + \gamma Y_{t-1} + \sum_{i=1}^s \Delta Y_{t-i} + u_t$$

O test ADF contrasta a hipótese de que o coeficiente γ é nulo (a serie ten raíz unitaria) fronte a que sexa negativo (a serie non ten raíz unitaria), polo tanto formulamos as seguintes hipóteses:

$H_0: \gamma = 0$ (existe raíz unitaria, serie non estacionaria)

$H_1: \gamma < 0$ (non existe raíz unitaria, serie estacionaria)

E o seu estatístico é: $t_\gamma = \frac{\hat{\gamma}}{s_\gamma}$

O contraste realízase con valores críticos tabulados que dependen da inclusión na ecuación do test dun termo constante ϵ /ou tendencia determinista.

Test de Durbin-Watson

Termos en inglés: **Durbin-Watson test**

Trátase dun test que se usa para estudar a presenza de autocorrelación de tipo AR(1) na perturbación aleatoria nun modelo econométrico. O estatístico determina se a correlación entre os termos da perturbación adxacentes é ou non é igual a cero. O estatístico emprega os residuos MCO na súa construción.

Hipóteses:

$H_0: \rho = 0$ incorrelación;

$H_1: \rho \neq 0$ autocorrelación AR(1)

$$dw = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2}$$

Estatístico:

O estatístico non segue ningunha distribución coñecida, polo que para tomar a decisión respecto á hipótese nula empréganse uns valores tabulados, cotas inferior (d_L) e superior (d_U), que dependen do tamaño mostral e do número de regresores do modelo.

Regras de decisión:

- se $dw < d_L$ rexeitamos H_0 , aceptando a hipótese alternativa de autocorrelación AR(1). Ademais sabemos que dita autocorrelación será positiva, é dicir $\rho > 0$.
- se $dw > 4 - d_L$, rexeitamos H_0 , aceptando a hipótese alternativa de autocorrelación AR(1). Ademais sabemos que dita autocorrelación será negativa, é dicir $\rho < 0$.
- se $d_U < dw < 4 - d_U$ non rexeitamos H_0 , habería incorrelación.
- se $d_L < dw < d_U$ ou $4 - d_U < dw < 4 - d_L$ o test non decide.

Test de Goldfeld e Quandt

Térmo en inglés: Goldfeld-Quandt test

En Econometría, o test de Goldfeld e Quandt aplícase cando sospeitamos que a varianza do erro varía conforme aos valores dunha variable explicativa identificable x_i . Tamén adecuado para cando podemos supoñer dous grupos de observacións con diferente varianza (por exemplo, con variables ficticias).

No primeiro caso o contraste podemos formulalo como:

H_0 : Homocedasticidade ($\sigma_t = \sigma$)

H_1 : Heterocedasticidade ($\sigma_t = f(x_t)$)

O estatístico é $R = SCE_2 / SCE_1$ sendo SCE_2 a SCE que se obtén a partir da estimación MCO do modelo coas observacións nas que se supón que a varianza é maior, e SCE_1 a obtida coas residuos MCO da estimación realizada coas observacións nas que a varianza se considera que debe ser máis pequena. Previamente requírese que as observacións estean ordenadas de forma crecente segundo os valores da variable que se sospeita está relacionada coa posible heterocedasticidade. Ademais suponse o mesmo número de observacións en cada submostra. De non ser así na obtención do estatístico debería dividirse cada SCE polos seus respectivos graos de liberdade.

Este estatístico, baixo hipótese nula certa, segue unha distribución F_{m_1, m_2} , sendo m_1 e m_2 o número de observacións menos os parámetros estimados en cada conxunto de observacións.

Test de Jarque-Bera

Térmo en inglés: **Jarque-Bera Test**

O test de Jarque-Bera permite contrastar se unha determinada variable ten unha distribución de probabilidade normal.

No estudo econométrico, empregamos este test para analizar a normalidade da perturbación aleatoria do modelo, empregando para ilo os residuos MCO. Contrastamos a asimetría e o exceso de curtose, sabendo que no caso de normalidade a asimetría tomará o valor 0 e a curtose o valor 3.

As hipóteses nula e alternativa que se formulan son:

$H_0 = \varepsilon_t$ segue unha distribución normal

$H_1 = \varepsilon_t$ non segue unha distribución normal.

O estatístico empregado defínese como:

con, S = índice de asimetría e K = índice de curtose.

$$\text{Jarque-Bera} = \frac{T}{6} \left(S^2 + \frac{(K-3)^2}{4} \right)$$

Este estatístico segue unha distribución χ^2_2 .

Test de raíces unitarias

Térmo en inglés: **Unit root tests**

Os tests de raíces unitarias úsanse para contrastar a estacionariedade dunha serie temporal. Os test formulan como hipótese nula a hipótese de que as series teñen tendencia estocástica (existe unha raíz unitaria) fronte a alternativa de que non hai tendencia estocástica (non existe raíz unitaria).

Nun proceso AR(1) que podemos escribir como

$$(1 - \phi_1 L) Y_t = u_t$$

a ecuación característica do seu polinomio de retardos autorregresivo sería:

$$1 - \phi_1 L = 0$$

A raíz do polinomio sería aquel valor que fai que se cumpra a igualdade, neste caso sería o valor $1/\phi_1$ e para que o proceso sexa estacionario esta raíz debe, en termos absolutos ser maior que a unidade, $|1/\phi_1| > 1$ (caer fora do círculo unitario), o que require que $|\phi_1| < 1$. Se $\phi_1 = 1$ entón a raíz do proceso sería igual a 1, o proceso tería unha raíz unitaria, sería un proceso con tendencia estocástica, sería un proceso non estacionario.

Test de White

Término en inglés: **White test**

O test de White é un test asintótico e de aplicación xeral que nos permite coñecer se existe heterocedasticidade na perturbación aleatoria nun modelo econométrico de regresión lineal.

A hipótese que se formula é:

H_0 : Existe homocedasticidade

H_1 : Existe heterocedasticidade

O test defínese como $T \cdot R^2$ onde T é o tamaño da mostra, R^2 o coeficiente de determinación dunha regresión auxiliar na que se fai depender o cadrado dos residuos MCO (e_t) das variables explicativas do modelo econométrico, os seus cadrados e os produtos cruzados.

O test, baixo a hipótese nula, segue unha distribución χ_p^2 , sendo p o número de regresores da regresión auxiliar que se formula (excluíndo a ordenada na orixe).

Por exemplo, no caso dun modelo econométrico con dous regresores, a ecuación auxiliar sería:

$$e_t^2 = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + \alpha_3 x_{1t}^2 + \alpha_4 x_{2t}^2 + \alpha_5 x_{1t} x_{2t} + v_t$$

Test Q^* de Ljung e Box

Término en inglés: **Ljung-Box Test**

Este contraste é un tipo de proba estatística que na metodoloxía ARIMA pódese empregar para analizar a independencia dos residuos na fase de validación na elaboración dun modelo ARIMA.

A hipótese nula no contraste será que a serie de residuos é unha serie ruído branco (todos os coeficientes de autocorrelación nulos) fronte a hipótese alternativa de que non o sexa

H_0 : Residuos (μ_t) son Ruído Branco.

H_1 : μ_t non son Ruído Branco.

O estatístico de Ljung e Box obtense segundo a seguinte expresión:

Na súa aplicación a unha serie de residuos dun modelo ARMA(p,q) ten unha distribución X^2 con $m-p-q$ grados de liberdade baixo a hipótese nula de que os residuos son independentes.

$$Q^* = T(T+2) \sum_{k=1}^M (T-k)^{-1} r_k^2$$

Trampa das variables ficticias

Término en inglés: **Dummy variable trap**

A denominada “trampa das variables ficticias” prodúcese cando se inclúen nun modelo econométrico demasiadas variables ficticias para describir o mesmo grupo. Provocará un problema de multicolinealidade perfecta, é dicir, o rango da matriz de regresores será menor que o número de parámetros a estimar. Supoñendo que a característica cualitativa que se quere incorporar ao modelo ten p modalidades ou categorías ocorrerá ao incluír simultaneamente o regresor ficticio e as p variables ficticias aditivas correspondentes as p modalidades. Para evitar esta situación, débese excluír no modelo unha das variables ficticias, ou ben, excluír o regresor ficticio incluíndo as p variables ficticias. Esta segunda opción tería o problema de que o modelo carece de regresor ficticio.

Exemplo:

Consideremos unha ecuación que relaciona os salarios dos individuos (Sal) co seu nivel de estudos (NE)

$$\text{Sal}_t = \beta_0 + \beta_1 \text{NE}_t + \varepsilon_t$$

Se queremos considerar o efecto sobre o salario do feito de que o perceptor sexa home ou muller deberemos introducir variables ficticias de forma aditiva se supoñemos que tal efecto repercute na ordenada na orixe, distinta para homes e mulleres. Neste exemplo caeríamos na trampa das variables ficticias se na ecuación do modelo incorporáramos dúas variables ficticias para recoller a variable cualitativa home/muller, é dicir, se especificáramos o modelo como:

$$\text{Sal}_t = \beta_0 + \beta_1 \text{NE}_t + \beta_2 \text{Muller}_t + \beta_3 \text{Home}_t + \varepsilon_t$$

xa que a suma das columnas correspondentes as variables ficticias “Muller” e “Home” sería unha columna de valores “1” a cal sería coincidente co columna correspondente ao regresor ficticio ao ter o modelo ordenada na orixe.

Variable aleatoria

Termo en inglés: **Random variable**

Unha variable aleatoria é unha función que asigna un número real a cada un dos resultados dun experimento aleatorio. Este termo está moi ligado ao termo de distribución da probabilidade (función que establece como se distribúe a probabilidade dunha variable aleatoria). Existen dous tipos de variables aleatorias:

- Variable aleatoria discreta: Unha variable aleatoria é discreta se o conxunto de todos os valores que pode tomar é un conxunto finito o numerable.
- Variable aleatoria continua: Unha variable aleatoria é continua se toma un número infinito non numerable de valores.

Variable endóxena

Termo en inglés: **Endogenous variable**

Nun modelo econométrico, a variable endóxena, tamén chamada explicada ou regresando, é aquela cuxo valor está determinado polas relacións establecidas dentro do modelo econométrico no que está incluída. Dito doutra forma, estas variables explícanse dentro dun modelo econométrico a partir das súas relacións cun conxunto de variables explicativas.

Dentro do modelo pode haber variables endóxenas correntes (Y_t) cando estas variables están explicadas no momento actual t , e endóxenas retardadas (Y_{t-1}), as cales son variables que foron explicadas polo modelo nun período anterior e que se incorpora a este como unha variable explicativa máis.

Variable esóxena

Termo en inglés: **Exogenous variable**

Nun modelo econométrico, dentro das variábeis observables podemos distinguir entre endóxenas correntes (variables cuxo comportamento vén explicado polo modelo) e predeterminadas, dentro das cales podemos diferenciar entre esóxenas e endóxenas retardadas.

As variables esóxenas son regresores non explicados polo modelo en ningún momento e poden clasificarse en:

- Controlables: aquelas cuxo valor depende da vontade do usuario do modelo. Á hora de realizar predicións pódese coñecer anticipadamente os valores que terán estas variables nun futuro.
- Non controlables: son aquelas cuxo valor non depende da vontade do usuario do modelo.

Variable ficticia

Término en inglés: **Dummy variable**

Denomínase variable ficticia (dummy, dicotómica ou binaria) a unha variábel que toma unicamente dous valores, 0 e 1. Ditos valores atribúense discrecionalmente, aínda que o normal é dar o valor 1 a presenza dunha determinada característica e 0 a súa ausencia. Nos modelos econométricos inclúense como regresores para recoller o efecto sobre a variábel endóxena de factores cualitativos, non cuantificables, entre outras razóns, pero tamén poden actuar como variables endóxenas.

Exemplo: Se consideráramos a inclusión no modelo da variable cualitativa “nivel de estudos dos individuos”, créanse as seguintes variables ficticias, supoñendo que a clasificación que adoptamos é estudos primarios, secundarios e terciarios:

$$D_1 = \begin{cases} 1 & (\text{se a persoa ten estudos primarios}) \\ 0 & (\text{noutro caso}) \end{cases}$$

$$D_2 = \begin{cases} 1 & (\text{se a persoa ten estudos secundarios}) \\ 0 & (\text{noutro caso}) \end{cases}$$

$$D_3 = \begin{cases} 1 & (\text{se a persoa ten estudos terciarios}) \\ 0 & (\text{noutro caso}) \end{cases}$$

Especificaríamos o modelo:

$$y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \delta_1 D_{1t} + \delta_2 D_{2t} + \epsilon_t$$

Sendo D_1 e D_2 as variábel ficticias. Neste caso, o nivel de “Estudios terciarios” considérase categoría de referencia ou grupo base, ao ser a variable ficticia omitida.

Variable retardada

Término en inglés: **Lagged variable**

Unha variable retardada é unha variable cuxo valor vén dado nun momento anterior no tempo en relación ao momento actual.

En Econometría, os modelos econométricos dinámicos inclúen entre os regresores, variables retardadas, ben valores retardados das variables exógenas, X_t , ben valores retardados da variable endóxena Y . Estes últimos son os denominados modelos autorregresivos.

Varianza

Término en inglés: **Variance**

Medida da dispersión dunha distribución probabilística. No caso da varianza dunha variable aleatoria (Y), exprésase como σ^2 , que é o valor esperado do cadrado das desviacións dos valores de Y respecto á súa media. É dicir, canto varía en termos medios o cadrado do valor real da variable Y con respecto á media da mesma.

A súa unidade de medida será a unidade de Y ao cadrado, polo que a súa interpretación será complicada. Por este motivo, é frecuente utilizar a desviación típica, que non é máis que a raíz cadrada da varianza, cuxa unidade de medida é a mesma que a de Y .

Tanto a varianza como a desviación típica son positivas. Ademais, a varianza non varía ante cambios de orixe, pero si ante cambios de escala.

