Introducción a la econometría

Un enfoque basado en simulación

Guido Ianni

# Clase 1: Instalación e interfaz

1. Overview del problema de la econometría
2. Interfaz de Eviews

* Archivos .wf1 y .prg
* Creación de WorkFiles
* Creación de series
* Objetos de ecuación
* Gráficos
  + Línea
  + Scatter
  + Histogramas
  + Q-QPlots
* Creación de simples
* Codificación de dummys
* Creación de gráficos en línea de comandos

https://eviews.com/Learning/programming\_a.html

# Clase 2: Estadística descriptiva

* Medidas de tendencia central y de dispersión
* Visualización de datos
  + Histograma
  + KDE
  + Q-Q plots
  + Boxplots
* Distribuciones derivadas de la distribución normal
* Ley de Grandes Números
* Teorema Central del limite
* Inferencia (estimación de varianza)

2-1 Descriptiva.prg

1. Medidas de tendencia central y de dispersión
2. Visualización de datos
   1. Fig 1 – Histogramas
   2. Fig2 – Medias y Boxplots
   3. Fig 3 – Q-Q Plots e histogramas asociados

*Actividades*

1. Cambiar forma de algún histograma a KDE

2. Agregar ajuste en el histograma a alguna distribución teórica

3. Hacer la prueba formal

4. Hacer alguna prueba de hipótesis de "media=algo" sencilla.

2-2 distr\_norms.prg

1. Simula distribuciones de probabilidad
2. Calcula los histogramas y los ajustes a la distribución teórica

**2-3 LeyGdesNums.prg**

1. Simula con y con

*Actividades*

Observar que cuando hay autocorrelación se tarda más en converger

**2-4 TCL.prg**

1. Simula observaciones de y calcula
2. Ilustra el Teorema Central del Límite haciendo el paso anterior R veces y viendo el histograma y QQ plot

*Actividades*

Jugar con distintas *distribuciones* y tamaños de muestra para ver qué tan veloz es la convergencia

**2-5 Estima\_varianza.prg**

*Queremos* ver:

Que y son insesgados, pero es sesgado

Que tiene menor varianza que

# Clase 3: Distribuciones bi-variadas

**-** Variaciones categoricas (ANOVA)

- Normal Bivariada

- MRLS. Características

- MRLS. Distribucion de beta\hat y sigma\hat

Distribuciones empiricas de beta\hat cuando e no es normal

- Transformaciones Box-Cox

- Distribución de beta\hat cuando la distribución de las perturbaciones no es normal

FIg 5.1 Bhat no es normal incluso para n grande, aunque normalidad aumenta con n.

fig 6.2 ¿cuán grande tiene que ser n? Para cualquier N, siempre hay una distribución que te caga.

\* y = xb+e con x~N(0,1) y e~DoublePareto (varianza infinita para a=2)

\* MCO beta estandarizado = sqrt(n (a-2)/a)) \* (bhat-beta) ~ N(0,1) asintotica

\* n=100 y a=[3 2.4 2.2 2.2 2.05]

**3-1** **bivariate-categ**.prg

1. Simula
2. Muestra el histograma con KDE y el boxplot agrupado por la categoria

*Actividades*

Plotear contra LOGIT/etc

Crear series e separando a por categoría

Hacer un test de igualdad de medias

Extender el análisis a más de 2 categorías y hacer un test de igualdad de las 3

**3-2** **Normal-bivariada**.prg

Simula drawings de

*Actividades*

Cambiar el valor de rho y analizar las consecuencias

Estimar por MCO

Invertir el modelo anterior

Estimar y comprobar si las estimaciones coinciden, explicar resultados

**3-3** **MRLS**.prg

Simula drawings de y plotea vs

Simula drawings de y plotea “fits” de 4 modelos distintos (dos con @coefs arbitrarios, MCO del modelo lineal y el modelo real)

**3-4** **COV\_BETAS**.prg

MonteCarlo de un MRLS, guarda y reject

Muestra histogramas de y scatters de los mismos para ver las distribuciones conjuntas

*Actividades*

Cambiar errores a otras distribuciones

Interesante con errores no-simétricos (e.g. )

**3-5** **Asimptotics**.prg

Los resultados asintóticos pueden fallar

**3-6** **NON\_LINEAR**.prg

Esta simulación se centra en modelos no lineales. En particular, se simula

Se presentan los SCATTER para XY y XZ con una muestra chica

fit lin-log y poly(2) para XY

fit poly(1) poly(2) y poly(3) para XZ

Actividades

1. Fijar el rango del eje LEFT de cada vista
2. Observar la varianza de y|x a medida que se core la ventana del sample
3. Observar la varianza de cada estimación no-lineal
4. Agregar poly-fit de grados mayores y ver si ajusta mejor o no a los datos
5. Confirmar estimando estos modelos

**3-7** **Box-Cox**.prg

Las distribuciones de probabilidad están determinadas si uno determina [en la normal , pero eso no pasa en todas las distribuciones; al margen de las transformaciones loc, scale].

El objetivo de Box-Cox, de alguna forma, es transformar una VA (generalmente, de la variable de respuesta de un modelo de regresión) para lograr que sea “más o menos constante”

A white sheet with black text and black text

Description automatically generated

# Clase 4: Laboratorio 1

DATASET CPS09mar

Datos de Marzo 2009 de la Current Population Survey (CPS). Esta estratificación extrae individuos con trabajo a tiempo completo (excl. militares). Forma parte de los datasets en Hansen(2015)

**Actividades**

0) Construir la serie wage (earnings/(hours\*weeks)

1) Observar la distribución de wage

- Estadísticas descriptivas

- Histograma (con KDE y Theoretical distribution)

- Boxplot

2) Encontrar una transformación T() que haga que tenga una distribución "suficientemente normal"

4) Analizar la distribución de T(w) según sexo (female) y latino

- Ver histogramas y boxplots

- Hacer una tabla 2x2 con las medias agrupadas por latino y por sexo

- Hacer el test de igualdad de salarios para ambas características (en conjunto, y por separado)

5) Hacer un gráfico para ver cómo evoluciona la media de T(w) como función de los años de educación, separando hombres y mujeres

6) lwage vs experience

- Construir la variables experience = age-education-6

- Hacer un grafico que tenga

i. la media de lwage condicional en experience para individuos con 12 años de educación

ii. el resultado de una proyección lineal

iii. el resultado de una proyección cuadrática

- KDE de T(w) para exp=5 10 25 40

7) Estimar modelos de regresión para

- T(w) ~ c female

- T(w) ~ c female married

- T(w) ~ c female married (c/interacciones)

8) La variable raza = {white black ... other}

- Explorar los value counts

- Estimar modelos T(w) ~ c female married blanco negro (c/interacciones)

9) Ahora nos vamos a centrar en el efecto de la Experiencie

- T(w) ~ c experience

- Dados los resultados de 6.iii) T(w) ~ c experience experience^2

# Clase 5

HAT Matrix | Proyecciones y Heteroscedasticidad

5-1 Prediction Intervals.prg

Los intervalos de regresión proporcionan una medida de la incertidumbre asociada con las estimaciones

es heteroscedástico debido a la incertidumbre de los parámetros

A graph with a line and dots

Description automatically generated

5-2 Heteroscedasticidad.prg

- Detección

i. Breusch-Pagan-Godfrey: e^2 = h(vars, params) | Generalmente Box-Cox

ii. White: e^2 = linear(x, x^2, interacciones)

iii. Harvey: log(e^2) = linear(vars, params)

iv. ARCH e^2 = linear(lags de e^2)

- Remediales: HC consistent

i. Huber-White

Estimaciones Heteroscedasticidad Eviews:

Los “sandwich methods” estiman .

Como , donde es el peso de la observación en la estimación de la matrix de covarianzas

HC0: [White tradicional]

HC1: [Bessel’s correction para sesgo por muestra chica]

HC2: donde viene de la HAT matrix

[La intuición es que incluso si es homoscedástico, no lo es por el parameter uncertainty]. Ponderar por el hat matrix es tener eso en cuenta

Sim 1: y\_con ~ N(mu, sigma^2)

y\_sin ~ N(mu, @mean(sigma)^2)

mu = b1 + b2\*x

sigma = h(x)

Sim 2: log(y\_nl)=c1 + c2 \* log(x) + u

u~N(0, s)

Mostramos

Fig1: Las consecuencias de la heteroscedasticidad

en el scatter de (x, y\_sin) y (x, y\_con)

Fig2: Que la heteroscedasticidad puede aparecer por un error de especificación,

si estimo y\_nl ~ x [el true\_model sería log(y\_nl) ~ log(x) ]

puedo tener heteroscedastacidad (y autocorrelación)

por no haber estimado la forma funcional adecuada

Fig 3: Cómo la presencia de heteroscedasticidad dificulta la estimación

A través de una simulación MonteCarlo obtenemos

la distribucion de beta\_hat en los modelos SIM1 y SIM2

5-3 Outliers.prg

Notación

[Modelo]

Subíndice = leave one out

es heteroscedastico incluso cuando es homoscedastico

debido al parameter uncertainty.

*Leverage plots: vs*

Análogo multivariado del lineplot de vs X.   
Si y se hace el plot de vs

*Influence statistics*

RStudent:

DFFITS:

DRResid

COVRATIO:

HatMatrix

DFBetas

*Actividad*

Dataset CPS09mar regresar para single asian male

Una observación particular tiene un DFFits de 0.29 (29% si se elimina esa obs). La observación exhibe una palanca (leverage) inusualmente alta de 0.33, indicando una influencia desproporcionadamente fuerte en su propio valor ajustado ().

Este individuo, de 65 años con 8 años de educación, tiene una experiencia potencial de 51 años, siendo la más alta en la muestra.

5-4 Autocorrelacion.prg

Montecarlo de

Ilustramos los momentos de un proceso AR(1) estable en el cual la distribución de es sesgada pero consistente, y su distribución no es normal (de hecho, no existe una solución analítica sencilla para esta distribución).

Se estiman 3 modelos:

Ignorar autocorrelación y hacer MCO)

Estimar un modelo AR usando las funcionalidades de Eviews

Estimar el modelo por MCG

Se muestran las distribuciones empíricas para cada modelo de y

*Actividades*

Diagnosticar la autocorrelación

Recalibrar los parámetros a

N=25 y R=100 (repeticiones insuficientes)

N=25 y R=10.000 (normalidad y alto sesgo)

N=250 y R=10.000 (consistencia)

En las simulaciones, los modelos 1 y 3 coinciden en su estimación, debido a que no existen otros regresores. Mostrarlo agregando otro regresor al modelo.

Ajustar el código para obtener:

Estimaciones de según AR(1) donde cada repetición tiene una observación más que la anterior.

Hacer el lineplot de a medida que aumenta

Ilustrar la convergencia de una forma alternativa.

Para clase8 woolridge.wf1 de eviews (User Guide II – p.35)

# Clase 6

6 - 1 Multicolinealidad.prg

Montecarlo de dos modelos (Alta y Baja), uno con alta y otro con baja multicolinealidad

Ambos modelos difieren en la correlación ρ que existe entre X1 y X2 (baja tiene la ρ/2). X3 es independietne de X1 y X2 sigmay se calibra a partir de al correlacion deseada entre (μy, y). Todas las series tienen media cero y varianza unitaria

Además de los modelos de reresión ALTA y BAJA se estiman dos modelos más:

PCA2 realiza la proyección de las X sobre los de 2 primeros componentes principales

Bias estima un modelo omitiendo uno de los regresores colineales (x2)

FIGURAS

FIG1: Scatter Matrix de regresores en ambos modelos (altima rep)

FIG2: Elipses de confianza para las estimaciones puntuales

con valores verdaderos señalados (altima rep)

FIG3: Histograma de los estimadores MonteCarlo de los 4 modelos de regresión

FIG4: Montecarlo distribuciones de b1 b2 y b3 y RB\_true de modelo Alta

*Actividades*

1. Observar los efectos de la correlación de las X en

R2 | t-statistics (p-vals) | F-Statistics

2. Obtener los VIFs de los coeficientes cada ecuación

3. Realizar la descomposición de la varianza basada en SVD

4. Mínimos Cuadrados restringidos (MCR)

Restricciones verdaderas y aproximadamente verdaderas

Test de Wald

6 - 2 Overfit y Regularizacion

Nosotros no observamos μ sino (x y). Quizás no hay forma de "adivinar" que el verdadero modelo no es lineal.

Una forma de resolverlo puede ser entrenar un modelo polinomial para captar no-linealidades en los datos. Aumentar los regresores siempre aumenta el R2 a medida que aumenta K (poly degree). El riesgo que existe es sobreajustar a los datos y no aprenderse el modelo subyacente.

¿Cómo evitar el sobreajuste? Train\_test\_split |

RMSE en conjuntos de entrenamiento y testeo

Graficas de los modelos

Una opción: elegir el K que minimice el ECM en Testeo

Otra opción: agregar una penalidad por tener coeficientes grandes

Eso da lugar a técnicas de regularización

*Actividades*

1. Ver los coefficientes de las regresiones

Como me crashea Eviews cuando hago coefevol de ElasticNet,

hacer el grafico de CoefEvolution (escalas de ejes logaritmicas)

y el diagnostico de train/test metrics

2. Chequear que VarSelect no hace regularizacion

3. Chequear que otra alternativa es PCA

6 - 3 Endogeneidad

Se lustran los efectos de

Incorporar regresores irrelevantes [ineficiencia]

Omitir regresores relevantes

Que correlacionan con las X [sesgado e inconsistencia]

Que NO correlacionan con las X [consitencia]

Montecarlo de

y = μ + e ~ N(μ, σ^2)

μ = b1 x1 + b2 x2 + b3 x3

corr(x1, x2) = rho2

corr(x1, x3) = 0

corr(x1, x4) = rho4 [que no está en el modelo]

El resto de las explicativas son ortogonales

Se estiman 4 modelos:

TRUE

OMITX2

OMITX3

ADDX4