Econometría II: Análisis de regresión con datos de series de tiempo

Universidad Centroamericana - Economía Aplicada 2022s2

Alvaro López-Espinoza

S1: Introducción a las series de tiempo

1.1. Datos de la sesión

Fecha de la sesión: 2022-08-29

Lecturas obligatorias:

- Wooldridge - Cap. 10.1-10.2
- Bee Dagum & Bianconcini - Cap. 2.1.

1.2. Naturaleza de las series de tiempo

Existen 4 estructuras básicas de datos:

Revisar: Wooldridge - Cap. 1.3

- Datos de corte tranversal: Muestra de unidades tomadas en algún punto dado en el tiempo.

Base de datos sobre tasas de crecimiento económico y características de un país

obsno	país	cpibrpc	govcons60	second60
1	Argentina	0.89	9	32
2	Austria	3.32	16	50
3	Bélgica	2.56	13	69
4	Bolivia	1.24	18	12
⋮	⋮	⋮	⋮	⋮
61	Zimbabwe	2.30	17	6

- Datos de combinación de cortes transversales: Combinación de dos o más muestras de datos de corte transversal.

Combinación de cortes transversales: precios de la vivienda en dos años diferentes

obsno	año	preciov	impprov	piescuadr	recs	baños
1	1993	85 500	42	1600	3	2.0
2	1993	67 300	36	1440	3	2.5
3	1993	134 000	38	2000	4	2.5
⋮	⋮	⋮	⋮	⋮	⋮	⋮
250	1993	243 600	41	2600	4	3.0
251	1995	65 000	16	1250	2	1.0
252	1995	182 400	20	2200	4	2.0
253	1995	97 500	15	1540	3	2.0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
520	1995	57 200	16	1100	2	1.5

- **Datos de series de tiempo**: Observaciones de una o varias variables a lo largo del tiempo.

Salario mínimo, desempleo y datos relacionados con Puerto Rico

obsno	año	prommin	coverprom	desempl	PNB
1	1950	0.20	20.1	15.4	878.7
2	1951	0.21	20.7	16.0	925.0
3	1952	0.23	22.6	14.8	1015.9
⋮	⋮	⋮	⋮	⋮	⋮
37	1986	3.35	58.1	18.9	4281.6
38	1987	3.35	58.2	16.8	4496.7

- Datos panel: Serie de tiempo por cada unidad de una base de datos de corte transversal.

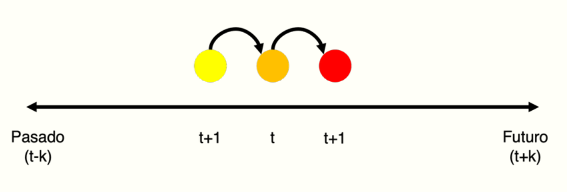
Datos de panel, de dos años, sobre estadísticas de delincuencia urbana

obsno	ciudad	año	homicidios	población	desempl	policía
1	1	1986	5	350 000	8.7	440
2	1	1990	8	359 200	7.2	471
3	2	1986	2	64 300	5.4	75
4	2	1990	1	65 100	5.5	75
⋮	⋮	⋮	⋮	⋮	⋮	⋮
297	149	1986	10	260 700	9.6	286
298	149	1990	6	245 000	9.8	334
299	150	1986	25	543 000	4.3	520
300	150	1990	32	546 200	5.2	493

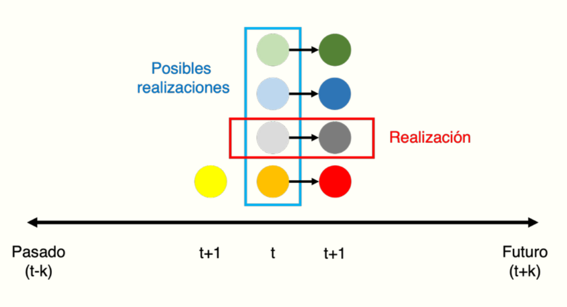
Características de las series de tiempo:

Revisar: Wooldridge - Cap. 10.1 & 1.3.

- Orden temporal: El pasado afecta al futuro.



- Aleatoriedad en las series de tiempo: Los datos como resultados de variables aleatorias.



- Periodicidad de los datos: La frecuencia de recolección de la información.

1.3. Componentes de las series de tiempo

Los componentes latentes de una serie de tiempo son:

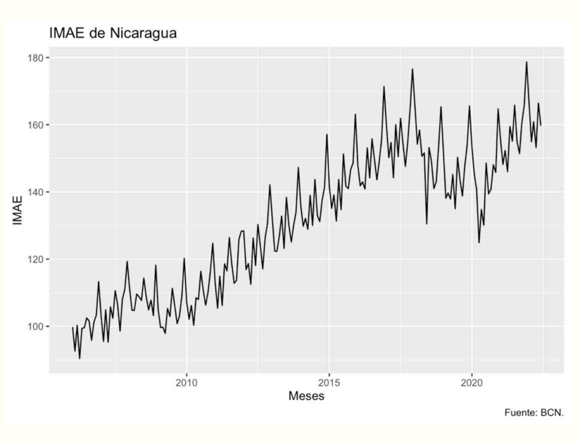
Revisar: Bee Dagum & Bianconcini - Cap. 2.1.

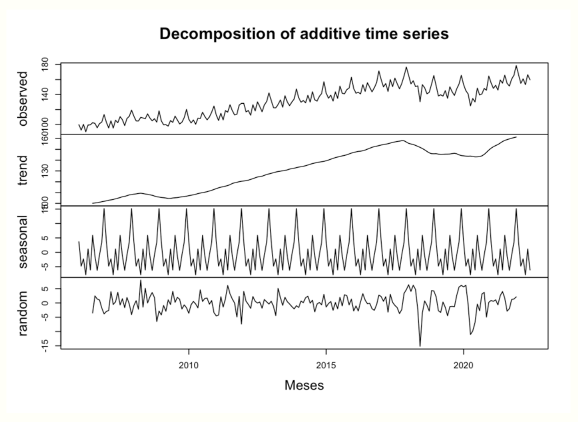
1. Tendencia (T): El valor esperado de largo plazo (tendencia secular).
2. Ciclo (C): Movimientos sobrepuestos a lo largo de la tendencia de largo plazo.
3. Estacional (S): Movimientos de corto plazo (menos de un año) sistemáticos a lo largo de la serie de tiempo.
4. Irregular (I): Componente aleatorio de la serie.

Los componentes pueden ser aditivos o multiplicativos:

- Aditivo: $y_t = T_t + C_t + S_t + I_t$
- Multiplicativo: $y_t = T_t \times C_t \times S_t \times I_t$

Ejemplo con el IMAE de Nicaragua:





1.4. Modelos de regresión de series de tiempo

Algunos modelos de series de tiempo:

- Modelos estáticos. Modeliza la relación contemporánea entre las variables.

$y_t = \beta_0 + \beta_1 z_t + u_t,\ t = 1, 2, \dots, n$

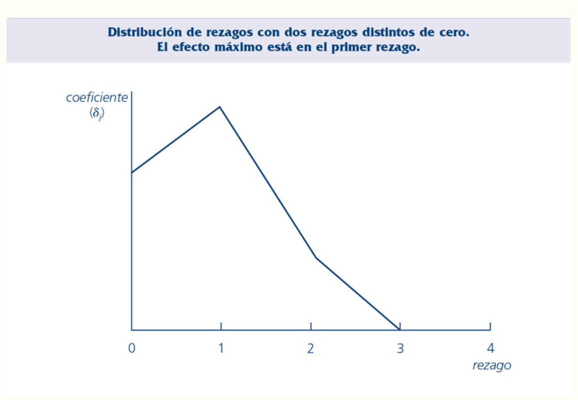
- Modelos con rezagos distribuidos finitos (FDL). Modeliza la relación de cuando una o más variables influyen en y en forma rezagada.

$y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + u_t,\ t = 1, 2, \dots, n$

Generalizando: $y_t = \alpha_0 + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \dots + \delta_q z_{t-q} + u_t$

Propensión de impacto (δ_0): Cambio inmediato en y debido al aumento de 1 unidad de z.

Distribución de rezagos (δ_j):



Propensión de largo plazo (LRP): El cambio en y debido a un cambio permanente de z. Es la suma de los coeficientes de la z actual y sus rezagos ($\delta_0 + \delta_1 + \delta_2$).

S2: Propiedades de muestras finitas de MCO bajo los supuestos clásicos

2.1. Datos de la sesión

Fecha de la sesión: 2022-08-31

Lecturas obligatorias:

- Wooldridge - Cap. 10.3

Lecturas opcionales:

- Wooldridge - Cap. 2-4

2.2. Insesgamiento de MCO

Supuesto ST.1: Lineal en los parámetros

Revisar: Supuestos de la Regresión Lineal Múltiple (RLM) en Wooldridge - Cap. 3.3.

El proceso estocástico $\{(x_{t1}, x_{t2}, \dots, x_{tk}, y_t) : t = 1, 2, \dots, n\}$ sigue el modelo lineal:

$y_t = \beta_0 + \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + u_t,$

donde $\{u_t : t = 1, 2, \dots, n\}$ es la secuencia de errores, y n es el número de observaciones.

Supuesto ST.2: No hay colinealidad perfecta

En la muestra no hay variables independientes que sean constantes ni que sean una combinación lineal perfecta de las otras.

Supuesto ST.3: Media condicional cero

Para cada t, dadas las variables explicativas para todos los periodos, el valor esperado del error u, es cero. Matemáticamente:

$\mathsf{E}(u_t|\mathbf{X}) = 0,\ t = 1, 2, \dots, n.$

Teorema ST.1: Insesgamiento de los estimadores de MCO

Bajo los supuestos ST.1, ST.2, y ST.3, los estimadores de MCO son insesgados condicionales sobre $\mathbf{X}$, y por tanto también incondicionalmente: $\mathsf{E}(\widehat{\beta}_j) = \beta_j,\ j = 0, 1, \dots, k.$

2.3. Teorema Gauss-Markov

Supuesto ST.4: Homocedasticidad

Revisar: Supuestos de Gauss-Markov en Wooldridge - Cap. 3.4-3.5.

La varianza de u_t, condicional en $\mathbf{X}$, es la misma para cualquier $t: \mathsf{Var}(u_t|\mathbf{X}) = \mathsf{Var}(u_t) = \sigma^2,\ t = 1, 2, \dots, n.$

Supuesto ST.5: No hay correlación serial

Los errores, condicionales sobre $\mathbf{X}$, en dos periodos distintos, no están correlacionados: $\mathsf{Corr}(u_t, u_s|\mathbf{X}) = 0$, para cualquier $t \neq s$.

Teorema ST.2: Varianzas de muestreo de los estimadores de MCO

Con base en los supuestos ST.1 y ST.5 de Gauss-Markov para las series de tiempo, la varianza de $\widehat{\beta}_j$, condicional sobre $\mathbf{X}$, es:

$\mathsf{Var}(\widehat{\beta}_j|\mathbf{X}) = \frac{\sigma^2}{STC_j(1-R_j^2)},\ j = 1, \dots, k,$

donde STC_j es la suma total de cuadrados de x_{tj} y R_j^2 es la R-cuadrada de la regresión de x_j sobre las otras variables independientes.

Teorema ST.3: Estimación insesgada de σ^2

Bajo los supuestos ST.1 a ST.5, el estimador $\widehat{\sigma^2} = \frac{SRC}{gl}$ es un estimador insesgado σ^2, donde $gl = n - k - 1$.

Teorema ST.4: Teorema Gauss-Markov

Bajo los supuestos ST.1 a ST.5, los estimadores de MCO son los mejores estimadores lineales insesgados condicionales sobre $\mathbf{X}$.

2.4. Inferencia

Supuesto ST.6: Normalidad

Revisar: Supuestos de distribución de muestreo en Wooldridge - Cap. 4.1.

Los errores u_t son independientes de $\mathbf{X}$ y son independientes e idénticamente distribuidos como $\mathsf{Normal}(0, \sigma^2)$.

Teorema ST.5: Distribuciones de muestreo normales

Bajo los supuestos ST.1 a ST.6, los supuestos del MCL para series de tiempo, los estimadores de MCO se distribuyen de forma normal, condicionales sobre $\mathbf{X}$.

Bajo la hipótesis nula, cada estadístico t tiene una distribución t, y cada estadístico F tiene una distribución F. También es válida la construcción usual de los intervalos de confianza.

S3: Introducción a R

Les comparto el script de la clase de hoy.

Tomado de: Healy, K. (2008). Data Visualization. A practical introduction (https://socviz.co).

```
# Instalacion de paquetes
my_packages <- c("tidyverse", "broom", "coefp
                 "gapminder", "GGally", "ggre
                 "here", "interplot", "margir
                 "mapdata", "MASS", "quantreg
                 "survey", "srvyr", "viridis"

install.packages(my_packages, repos = "http:/

## Uso de librerias
library(devtools)
install_github("kjhealy/socviz")
devtools::install_github("kjhealy/socviz")

# Introduccion
## Librerias
library(tidyverse)
library(socviz)
library(gapminder)

## Objetos
c (1, 2, 3, 1, 3, 5, 25)
my_numbers <- c (1, 2, 3, 1, 3, 5, 25)
your_numbers <- c(5, 31, 29, 4, 89, 43)

my_numbers
your_numbers

## Funciones
### Help
? mean

mean(my_numbers)
mean_my_numbers <- mean(my_numbers)

# Dataframes
titanic
class(titanic)
class(my_numbers)
class(my_packages)
titanic$percent

## Crear un data frame
url <- "https://cdn.rawgit.com/kjhealy/viz-or
organs <- read_csv(file = url)

# Plots
gapminder
p <- ggplot(data = gapminder,
       mapping  = aes(x = gdpPercap, y = life
p + geom_point()
```