Econometría Financiera

2023

Programa y Unidad I

Rodrigo Ortiz

Universidad Adolfo Ibáñez

Chile, 2023

1. Introducción
2. Evaluaciones
3. Bibliografía: Esencial
4. Bibliografía: Avanzada
5. Repaso
    - Tipos de Variables
    - Análisis Univariante
    - Modelo de Regresión Lineal Múltiple
    - Estimación por Mínimos Cuadrados Ordianarios

Introducción

El objetivo de este curso es introducir al alumno a las propiedades de los modelos de series de tiempo, así como a las aplicaciones empíricas. Al final del curso, el alumno deberá ser capaz de identificar las herramientas más apropiadas para abordar un problema financiero de carácter estadístico.

El curso se estructurará de manera tal que el alumno se vea expuesto a la teoría y a su aplicación, utilizando para tal fin programas estadísticos de uso amplio, tales como STATA y R-Project.

Objetivos del curso

Al finalizar el curso se espera que cada participante sea capaz de:

- Conocer y aplicar herramientas estadísticas de descripción de datos.
- Identificar las herramientas econométricas de series de tiempo más apropiadas para abordar un problema financiero.
- Utilizar con destreza un paquete econométrico, tal como STATA y R-Project.

Evaluaciones

- 1 prueba (40 %)
- 1 trabajo grupal (30 %)
- 1 examen (30 %)

Importante:

- No se elimina ninguna nota.
- A fin de facilitar la corrección, las guías y el trabajo se deben realizar en grupos de 4-5 alumnos. No se aceptarán entregas individuales.
- Se recuerda que, por reglamento, el alumno debe contar con una asistencia de, al menos, un 75 %.

Bibliografía: Esencial

- Brooks, C. (2014). Introductory Econometrics for Finance. Tercera edición. Cambridge University Press.
- Fabozzi, F, S. Focardi, S. Rachev & B. Arshanapalli (2014). The Basics of Financial Econometrics: Tools, Concepts and Asset Management Applications. Wiley.
- González-Rivera, G (2012). Forecasting for Economics and Business. Primera edición. Routledge.
- Hanke, J. & D. Wichern (2010). Pronósticos en los negocios. Novena edición. Pearson.
- Newbold, P., W. Carlson, & B. Thorne (2013). Estadística para administración y economía. Octava edición. Pearson.
- Novales, A. (1993). Econometría. Segunda edición. McGraw-Hill.
- Walpole. R., R. Myers, S. Myers, K. Ye (2012). Probabilidad y estadísticas para ingeniería y ciencias. Novena edición. Pearson.
- Wooldridge, J. (2015). Introducción a la Econometría. Quinta edición. Cengage Learning.

Bibliografía: Avanzada

- Enders, W. (2015). Applied Econometric Times Series. Cuarta edición. Wiley Series in Probability and Statistics.
- Mills, T. and Markelos, R. (2008). The Econometric Modelling of Financial Time Series. Cambridge University Press.

Tipos de Variables

- Variables cuantitativas: edad, estatura, renta
  - Continuas o de intervalo (estatura)
  - Discretas (número de hermanos)
- Variables cualitativas, atributo o categoría: color, género, municipio de nacimiento
  - Binarias, dos valores disponibles (género, se convierte en numérica)
  - Generales (color de ojos, podemos transformarlas en binarias)

Tipos de Variables

- Variables cualitativas, atributo o categoría:
- Tenemos p variables numéricas en un conjunto de n elementos.
- Cada una de estas p variables se denomina una variable **escalar o univariante**.
- El conjunto de p variables forman una variable **vectorial o multivariante**

$$X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1p} \\ x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix}$$

- p variables escalares en cada uno de los n elementos pueden representarse en una matriz X, de dimensiones (n, p), la **matriz de datos**.

La variable escalar x_j, la media muestral

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$$

que para una variable binaria es la frecuencia relativa de aparición del atributo y para una numérica es el centro de gravedad o geométrico de los datos.
Medida de variabilidad con relación a la media, la **desviación típica muestral**

$$s_j = \sqrt{\frac{1}{n}\sum_{i=1}^{n} d_{ij}} = \sqrt{\frac{1}{n}\sum_{i=1}^{n} (x_{ij} - \overline{x})^2}$$

Análisis univariante

El cuadrado de la desviación típica, la **varianza muestral**

$$s_j^2 = \frac{1}{n}\sum_{i=1}^{n} dij = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x})^2$$

Relación lineal entre dos variables, la **covarianza muestral** entre x_j y x_k

$$s_{jk} = \frac{1}{n}\sum_{i=1}^{n}(x_{ij} - \overline{x_j})(x_{ij} - \overline{x_k})$$

Análisis univariante

Medida de relación lineal entre x_j y x_k, el **coeficiente de correlación muestral**

$$r_{jk} = \frac{s_{jk}}{\sqrt{s_{jj}}\sqrt{s_{kk}}} = \frac{\sum_{i=1}^{n}(x_{ij}-\overline{x_j})(x_{ik}-\overline{x_k})}{\sqrt{\sum_{i=1}^{n}(x_{ij}-\overline{x_j})^2}\sqrt{\sum_{i=1}^{n}(x_{ik}-\overline{x_k})^2}}$$

$$-1 \leq r \leq 1$$

Coeficiente de variación: magnitud del error promedio de medición como porcentaje de la cantidad medida.

$$CV_j = \sqrt{\frac{s_j^2}{\overline{x_j}^2}}$$

Coeficiente de asimetría

$$A_j = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_{ij} - \overline{x_j})^3}{s_j^3}$$

Coeficiente de homogeneidad: mide la relación entre la variabilidad de las desviaciones y la desviación media

$$H_j = \frac{\frac{1}{n} \sum_{i=1}^{n} (d_{ij} - s_j^2)^2}{s_j^4} \geq 0$$

Análisis univariante

Coeficiente de kurtosis: Mide la relación entre la variabilidad de las desviaciones y la desviación media.

$$K_j = \frac{1}{n}\sum_{i=1}^{n}\frac{(x_{ij}-\overline{x_j})^4}{s_j^4}$$

Modelo de Regresión Lineal Múltiple

- x_k: variables predictoras o independientes, $k = 1, \ldots, p$
- y: variable respuesta o dependiente
- n: observaciones independientes de y
- β_k coeficientes de regresión, **efecto marginal** (cambio esperado) en y por cambio unitario en x_k cuando el resto de las variables permanece constante
- ϵ: **efecto** de todas las variables no incluidas en el análisis

$$
\begin{aligned}
y_1 &= \beta_0 + \beta_1 x_{11} + \beta_2 x_{12} + \cdots + \beta_p x_{1p} + \epsilon_1 \\
y_2 &= \beta_0 + \beta_1 x_{21} + \beta_2 x_{22} + \cdots + \beta_p x_{2p} + \epsilon_2 \\
&\vdots \\
y_n &= \beta_0 + \beta_1 x_{n1} + \beta_2 x_{n2} + \cdots + \beta_p x_{np} + \epsilon_n
\end{aligned}
$$

$$
X = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots & \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \epsilon_0 \\ \epsilon_1 \\ \vdots \\ \epsilon_p \end{bmatrix}
$$

$$
y_{(nx1)} = X_{(nx(p+1))} \beta_{((p+1)x1)} + \epsilon_{(nx1)}
$$

$$y_{(nx1)} = X_{(nx(p+1))}\beta_{((p+1)x1)} + \epsilon_{(nx1)}$$

- E(ϵ)=0_{nx1}
- var(ϵ)=E($\epsilon\epsilon$')
- cov(ϵ_j,ϵ_k)=E($\epsilon_j\epsilon_k$')=0 (independientes entre sí)
- $\epsilon \sim NM_m(0, \sigma^2 I_n)$
- β y σ^2 **son parámetros desconocidos**

Modelo de Regresión Lineal Múltiple

Tenemos $p{+}1$ parámetros β y σ^2. Necesitamos más datos que los parámetros a estimar. Hipóyesis adicionales:

- Número mínimo de observaciones igual a $p+1$
- Variables x_k linealmente independientes
- Para cada conjunto fijo de valores de x_k, la distribución de **y** tiene media $E[\mathbf{y}] = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$
- Media como función **lineal** de los parámetros no conocidos $\beta_o, \beta_1, \ldots, \beta_p$
- $var(\mathbf{y})$ es constante, no depende de los valores de x_k
- La variable respuesta $\mathbf{y} \sim N(\mathbf{X}\beta, \sigma^2 I_n)$

Estimación por Mínimos Cuadrados Ordinarios

X rango máximo $p+1 \leq n$. Se desea encontrar β tal que se minimice la diferencia.

$$
\begin{aligned}
L &= \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \cdots - \beta_p x_{ip}) \\
&= \sum_{i=1}^{n} \epsilon_i^2 \\
&= \epsilon' \epsilon \\
&= (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) \\
&= \mathbf{y}'\mathbf{y} - \mathbf{y}'\mathbf{X}\beta - \beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta \\
&= \mathbf{y}'\mathbf{y} - 2\beta'\mathbf{X}'\mathbf{y} + \beta'\mathbf{X}'\mathbf{X}\beta
\end{aligned}
$$

Estimación por Mínimos Cuadrados Ordinarios

$$\frac{\partial L}{\partial \beta}|_{\widehat{\beta}} = -2\mathbf{X}'\mathbf{y} + 2\mathbf{X}'\mathbf{X}$$

$$\mathbf{X}'\mathbf{X}\widehat{\beta} = \mathbf{X}'y$$

El **estimador de mínimos cuadrados de** β es

$$\widehat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$$

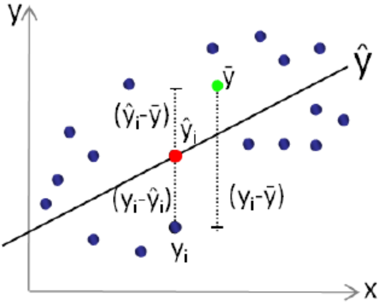
El **modelo de regresión ajustado** es:

$$\widehat{\mathbf{y}} = \mathbf{X}\widehat{\beta}$$

$\widehat{\beta}$ es el estimador lineal insesgado de mínima varianza.

$$SC_T = SC_R + SC_E$$

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (\widehat{y}_i - \overline{y})^2 + \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2$$



$$SC_T = SC_R + SC_E$$

$$\sum_{i=1}^{n} (y_i - \overline{y})^2 = \sum_{i=1}^{n} (\widehat{y_i} - \overline{y})^2 + \sum_{i=1}^{n} (y_i - \widehat{y_i})^2$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \widehat{\epsilon}^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2} = \frac{\sum_{i=1}^{n} (\widehat{y_i} - \overline{y_i})^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}$$

La descomposición de sumas al cuadrado sugiere que la **calidad del ajuste de los modelos** puede ser medida por el **coeficiente de determinación**:

$$R^2 = \frac{\text{Variabilidad explicada}}{\text{Variabilidad total}}$$

La cantidad R^2 entrega la proporción de la variabilidad total de las y_i explicada por las variables predictoras $x_1, x_2, \ldots, x_p$.

Contraste de significación global de la regresión:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$
$$H_1 : \beta_k \neq 0 \text{ para algún } k$$

El estadístico:

$$F_0 = \frac{\frac{SC_R}{p}}{\frac{SC_E}{n-p-1}} = \frac{MC_R}{CM_E}$$

- Rechazamos H_0 si $F_0 \geq F_{\alpha,p,n-p-1}$
- Aceptamos H_0 si $F_0 < F_{\alpha,p,n-p-1}$

Análisis de la varainza (ANOVA) para la significanción de la regresión en la regresión múltiple

Fuente de variación	Suma de cuadrados	Grados de libertad	Media cuadrática	F_0
Regresión	SC_R	p	MC_R	$\frac{MC_R}{MC_E}$
Error o residuo	SC_E	$n-p-1$	CM_E	
Total	SC_T	$n-1$		

Contraste de significación individual:

$$H_0 : \beta_k = 0$$
$$H_1 : \beta_k \neq 0$$

- Rechazamos H_0 si $|t_0| \geq t_{\frac{\alpha}{2},n-p}$
- Aceptamos H_0 si $|t_0| < t_{\frac{\alpha}{2},n-p}$

Multicolinealidad entre las variables predictoras:

- Alta correlación entre las variables predictoras x_k
- La varianza de los coeficientes de regresión aumenta
- Estimaciones poco confiables

Indicadores de multicolinealidad:

- Factor de inflación de varianza, $FIV(\widehat{\beta}) = \frac{1}{R_j^2} > 10$
- Determinante de la matriz de correlaciones cercano a cero
- Uno o más valores propios de **X'X** iguales a cero
- Cociente $\frac{\lambda_{max}}{\lambda_{min}} > 10$
- Prueba F significativa y pruebas de significancia individual no significativas.

Solución: Otro tipo de regresiones, por ejemplo **regresión PLS**

Otros problemas que afectan a los modelos de regresión

- **Autocorrelación** de los errores ϵ_i, por ejemplo variables medidas a través del tiempo
  - Prueba de Durbin-Watson
  - Solución: **modelos de series de tiempo**
- **Heterocedasticidad**, varianza de los errores no constante, puede indicar
  - Efecto de interacción entre una variable predictora presente en el modelo y otra que no esta presente
  - Algunas variables independientes son sesgadas mientras otras no
  - Solución: **regresión WLS**
- Presencia de datos atípicos