

Regresión lineal en series de tiempo

Gustavo A. Caffaro

Instituto Tecnológico de Santo Domingo

Nov 2021 - Ene 2022



Introducción

Contenido:

- Regresión lineal para series de tiempo
- Supuestos de la regresión lineal
- Estimación de los parámetros
- Selección de modelo (bondad de ajuste)



La regresión en el contexto de series de tiempo

Asumamos que tenemos una serie x_t , para $t = 1, \dots, n$, que *depende* de manera lineal de una colección de series *independientes*, $x_{1,t}, x_{2,t}, \dots, x_{q,t}$ fijas y conocidas. Entonces, podemos expresar esta relación mediante el **modelo de regresión lineal**

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

donde los β_0, \dots, β_k son los coeficientes desconocidos de la regresión y w_t es un ruido blanco con media cero y varianza σ_w^2 .



Ej. Curva de Phillips

En macroeconomía, la relación entre la inflación y el desempleo suele llamarse la curva de Phillips. Podríamos caracterizar una curva de Phillips lineal de la siguiente manera:

$$infl_t = a + b \cdot desemp_t + u_t$$

El objetivo de la estimación por regresión lineal es estimar los valores de a y b mediante los estimadores \hat{a} y \hat{b} .



Ej. Indicadores financieros en RD

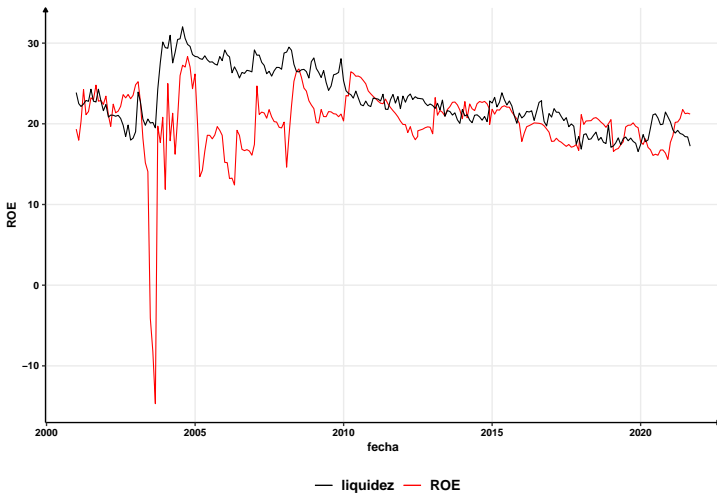
Podríamos describir la relación entre la rentabilidad del sistema financiero y su liquidez.

$$ROE_t = \alpha + \beta \cdot liquidez_t + u_t$$

- ROE: Rentabilidad del patrimonio (ingresos/patrimonio)
- liquidez: disponibilidades/total de captaciones



Ej. Indicadores financieros en RD



Supuestos de la regresión lineal

- 1 Linealidad
- 2 Rango completo
- 3 Exogeneidad de las variables independientes
- 4 Homoscedasticidad
- 5 No correlación serial
- 6 Normalidad



1. Linealidad

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

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

donde u_t es una serie de errores y n el número de observaciones.

Este modelo especifica una relación lineal entre la variable dependiente y las independientes.

La linealidad se refiere a cómo son introducidos los parámetros y el error, no a la relación entre las variables. Por ejemplo, $y = a + b \ln x + e$ sigue siendo lineal.



1. Linealidad

El supuesto de linealidad no es tan descabellado.

Muchas relaciones complejas y no lineales pueden aproximarse con un modelo lineal. La idea es comenzar con una relación no lineal $y = f(x)$, tomar el logaritmo, reexpresar y en función de estos logaritmos, y expandir usando el segundo orden de las series de Taylor, para obtener un modelo lineal en parámetros:

$$\ln y = \beta_0 + \sum_{k=1}^K \beta_k \ln x_k + \frac{1}{2} \sum_{k=1}^K \sum_{l=1}^K \gamma_{kl} \ln x_k \ln x_l + \epsilon.$$

Esta técnica es ampliamente utilizada, y recibe el nombre de *log-linearización*.



Notación vectorial

Con la intención de simplificar la notación, podríamos reescribir el modelo de regresión lineal en forma matricial. Digamos que $\mathbf{x}_t = (x_{t1}, x_{t2}, \dots, x_{tk})$ captura el conjunto de todas las variables independientes en el período t .

De igual forma, podríamos llamar \mathbf{X} la matriz que incluye todas las variables independientes en todos los tiempos. \mathbf{X} tiene dimensión $n \times k$. Entonces podríamos reescribir el modelo de regresión lineal como

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{u},$$

donde \mathbf{y} , \mathbf{u} son vectores de dimensión $n \times 1$.

De esta forma es que los programas estadísticos, como R o MATLAB, almacenan las observaciones.



2. Rango completo

Una matriz \mathbf{X} de dimensión $n \times K$ tiene rango completo si su rango es K . Matemáticamente:

$$\text{rango}(X) = \dim(X)$$

Este supuesto nos dice que ninguna variable independiente es una constante o que no existen relaciones lineales exactas entre las variables independientes.

Si este supuesto se viola, el modelo es incalculable (luego veremos por qué).

Adicionalmente, si estamos numéricamente cerca de violar este supuesto, tenemos el problema de multicolinealidad.



3. Exogeneidad de las variables independientes

Este supuesto asume que los errores u tienen valor esperado condicional igual a cero para *cada* observación:

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

¡Esta no es una condición trivial! De hecho, es una (si no *la*) condición empíricamente más sencilla de violar:

- ej. error de muestreo o sesgo de variable omitida.

Este supuesto nos dice que el valor esperado de cada error en cada período t , no está relacionado con *ninguna* variable x_i en *ningún* período.

O sea, que u_t no está relacionado con ninguna de las x s ni en el tiempo t , ni $t - 1$, ni $t + 1$, etc. . .



3. Exogeneidad de las variables independientes

Intuitivamente, este supuesto nos dice que ninguna variable tiene ningún tipo de información respecto al valor esperado del error u en ningún período.

¿Podrían pensar en un ejemplo donde esto no sea cierto?

Si $E(u_t) = 0$ y u_t y x_{tj} son independientes, entonces automáticamente se cumple este supuesto.

Finalmente, si el modelo no incluye una constante, asumir que $E(u_t) = 0$ es muy restrictivo. En consecuencia (a menos que haya una razón teórica muy fuerte), ¡los modelos de regresión *siempre* deben incluir una constante!



Estimadores insesgados

En conjunto, los supuestos 1-3 garantizan que los estimadores de la regresión lineal, $\hat{\beta}_j$, no sean sesgados.

Un estimador insesgado $\hat{\beta}$ es aquel cuyo valor esperado corresponde con el parámetro poblacional de dicho estimador. Es decir, que $E(\hat{\beta}) = \beta$.

Ej.

Digamos que queremos tener estimaciones de los parámetros a y b que definen la relación real entre dos variables, y y x , descrita por $y = a + bx + u$. Los *estimadores* insesgados de a y b , los cuales llamamos \hat{a} y \hat{b} , serían aquellos cuyo valor esperado es a y b , respectivamente:

$$E(\hat{a}) = a, \quad E(\hat{b}) = b.$$



4. Homoscedasticidad

El cuarto supuesto de la regresión lineal tiene que ver con las varianzas y covarianzas de los errores.

En particular, se asume que, condicionado en \mathbf{X} , la varianza de u_t es la misma para todos los tiempos t :

$$\text{Var}(u_t | \mathbf{X}) = \text{Var}(u_t) = \sigma^2, \quad t = 1, 2, \dots, n$$

En la práctica, este supuesto se viola con regularidad. ¿CUáles de los ejemplos vistos en clase no cumplen con este supuesto?

Sin embargo, como veremos más adelante, este supuesto puede ser relajado de manera relativamente sencilla, por lo que no presenta un problema tan serio.



5. No correlación serial

Condicional en \mathbf{X} , los errores de dos períodos distintos no tienen correlación:

$$\text{Corr}(u_t, u_s \mid \mathbf{X} = 0), \quad \forall t \neq s.$$

Si ignoramos la condicional en \mathbf{X} tenemos que

$$\text{Corr}(u_t, u_s) = 0, \quad \forall t \neq s,$$

lo que nos dice que los errores no guardan correlación en el tiempo.

Cuando este supuesto no se cumple, los errores u_t sufren de **correlación serial**, o **autocorrelación**.



Insesgo de $\hat{\sigma}^2$

Si los supuestos 1-5 se mantienen, se tiene que

- el estimador de la varianza de los errores, $\hat{\sigma}^2$, es insesgado: $E(\hat{\sigma}^2) = \sigma^2$
- la varianza de los estimadores $\hat{\beta}_j$, condicionada en \mathbf{X} , viene dada por

$$\text{Var}(\hat{\beta}_j | \mathbf{X}) = \frac{\sigma^2}{SST_j(1 - R_j^2)},$$

donde SST_j es la suma total de cuadrados de x_{ij} y R_j^2 es el R -cuadrado de la regresión de x_j en las demás variables independientes.

- $SST_j = \sum_{t=1}^n (x_j - \bar{x}_j)$
- El R^2 mide qué tan bien proyecta nuestro modelo la variable objetivo. Más en breve.



6. Normalidad

El último supuesto de la regresión lineal es el de la normalidad de los errores.

Formalmente, este supuesto nos dice que los errores u_t son independientes de \mathbf{X} y son independientes e idénticamente distribuidos normalmente:

$$u_t \stackrel{iid}{\sim} N(0, \sigma^2)$$

Este supuesto implica los supuestos 3-5. ¿Por qué?

Este supuesto nos dice que, condicionando en \mathbf{X} :

- los $\hat{\beta}_j$ son normales
- podemos hacer inferencia sobre los estimadores (ej. podemos construir intervalos de confianza, estadísticas t)



Recapitulando supuestos RL

Los supuestos 1-3 aseguran que los parámetros de la regresión lineal que queremos estimar (e.j. los $\hat{\beta}_j$) son insesgados.

Los supuestos 1-5 aseguran que

- el estimador de la varianza del modelo, $\hat{\sigma}^2$, es insesgado.
- podemos tener una estimación de la varianza de los $\hat{\beta}_j$.

Los supuestos 1-6 aseguran que:

- los $\hat{\beta}_j$ están distribuidos normalmente
- podemos realizar inferencia sobre los estimadores del modelo



Estimando β_j por mínimos cuadrados

Vimos que estamos interesados en estimar e inferir sobre los parámetros β . Existen varias metodologías para hacer esta estimación. Aquí veremos una de las más sencillas: la estimación por **mínimos cuadrados** (least squares). El objetivo de mínimos cuadrados busca minimizar la suma del cuadrado de los errores:

$$\min_{\beta} \sum_{t=1}^n u_t^2 = \min_{\beta} \sum_{t=1}^n (y_t - \beta_0 - \beta_1 x_{t1} - \cdots - \beta_k x_{tk})^2,$$

el cual es sencillo resolver utilizando técnicas de optimización.



Mínimos cuadrados

Utilizando notación matricial, el objetivo se escribiría como

$$\min_{\beta} \mathbf{u}'\mathbf{u} = \min_{\beta} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta).$$

Si expandimos y optimizamos igualando la 1ra derivada a cero, tenemos que

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

el cual puede ser resuelto si y solo si \mathbf{X} tiene rango completo:

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



Bondad de ajuste

Definamos los siguientes conceptos que miden, respectivamente, la **suma total de cuadrados (SST)**, la **suma explicada de cuadrados (SSE)**, y la **suma de residuales cuadrados (SSR)**:

$$SST = \sum_{t=1}^n (y_t - \bar{y})^2$$

$$SSE = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2$$

$$SSR = \sum_{t=1}^n (y_t - \hat{y}_t)^2 = \sum_{t=1}^n \hat{u}_t^2$$

Se podría probar que

$$SST = SSE + SSR$$



El coeficiente de determinación, R^2

Dividiendo la ecuación anterior por SST, tenemos que

$$\frac{SSE}{SST} + \frac{SSR}{SST} = \frac{SST}{SST} = 1$$

Así obtenemos el R -cuadrado:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$$

el cual se interpreta como la proporción de la variación de la muestra en y_t explicada por la regresión lineal. En palabras llanas, el R^2 mide qué tan bien el modelo se ajusta a la data.

Por definición:

$$0 \leq R^2 \leq 1$$



Overfitting

Resulta ser que con cada variable que agregamos a nuestro modelo, el R^2 aumenta.

Esto sugeriría agregar tantas variables como podamos al modelo.

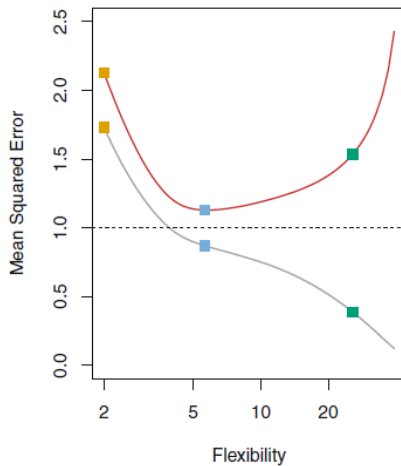
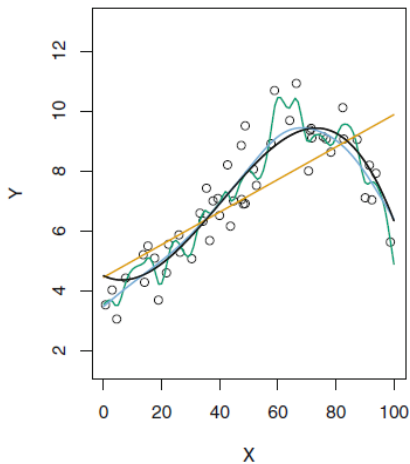
- Sin embargo, agregar variables indiscriminadamente afecta la calidad de las proyecciones.

Este fenómeno se conoce como **overfitting**.

Intuitivamente, el modelo abusa de la cantidad de variables disponibles (grados de libertad) para tratar de ajustarse lo más posible a las observaciones de y .



Overfitting



Ajuste vs parsimonia

Dicho todo esto, ¿cómo sabemos cuáles variables incluir en el modelo y cuáles no?

Existen criterios ampliamente utilizados para responder esta pregunta. Algunos:

- criterio de información de Akaike (AIC)
- criterio de información bayesiano (BIC)



Criterio de información de Akaike (AIC)

El **Criterio de Información de Akaike (AIC)**, viene dado por

$$AIC = T \ln SSR + 2k,$$

donde k es el número de parámetros del modelo.

El modelo que minimiza el AIC es, de acorde a este criterio, el *mejor* modelo.

El objetivo del AIC es *penalizar* la incorporación de cada nueva variable.



Criterio de información bayesiano (BIC)

El **criterio de información bayesiano (BIC)** viene dado por

$$BIC = T \ln SSR + k \ln T$$

El BIC es también llamado Criterio de información de Schwarz

Observa que a partir de alrededor de 8 observaciones, la penalidad del BIC es mayor que la del AIC:

- $k \ln T \geq 2k, T \geq e^2 \approx 7.4$

⇒ El BIC tiende a escoger modelos más parsimoniosos (pequeños).

Varios estudios sugieren usar BIC cuando el número de observaciones es grande, y el AIC cuando es pequeño.

