El Modelo Lineal

Román Salmerón

Departamento de Métodos Cuantitativos para la Economía y la Empresa
Facultad en Ciencias Económicas y Empresariales
http://fccee.ugr.es/
romansg@ugr.es

1. Especificación del Modelo Lineal

Estudio de una variable dependiente a partir de k variables independientes (con constante) a partir de n observaciones.

$$y_{n\times 1} = X_{n\times k}\cdot\beta_{k\times 1} + u_{n\times 1} \quad \begin{cases} E[u_{n\times 1}] = 0_{n\times 1} \text{ (ya que } E[u_t]=0\ \forall t \text{)} \\ Var(u_{n\times 1}) = \sigma^2\cdot Id_{n\times n} \\ \qquad \hookrightarrow Var(u_t)=\sigma^2\ \forall t, \quad Cov(u_t,u_s)=0,\ \forall t\neq s \\ X \text{ no aleatoria con } rg(X)=k \\ \qquad \hookrightarrow X_i, i=1,\dots,k, \text{ linealmente independientes} \\ X \text{ y } u \text{ incorrelados} \end{cases}$$

$$y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \quad X = \begin{pmatrix} 1 & X_{12} & X_{13} & \cdots & X_{1k} \\ 1 & X_{22} & X_{23} & \cdots & X_{2k} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n2} & X_{n3} & \cdots & X_{nk} \end{pmatrix} = (i\ X_2 \cdots X_k), \quad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}.$$

2. Estimación del Modelo Lineal

Estimación de las cantidades constantes del modelo.

♯ $\widehat{\beta} = \left(X^tX\right)^{-1}X^ty$ estimador por MCO de β

◇ Tª Gauss-Markov: $\widehat{\beta}$ es un estimador lineal, insesgado y óptimo (mínima varianza)

◇ $Var(\widehat{\beta}) = \sigma^2\cdot\left(X^tX\right)^{-1}$

◇ Consecuencias estimación MCO: $i^t\cdot e = \sum_{t=1}^{n} e_t = 0$, $X^t\cdot e = 0$, $\overline{Y} = \overline{\widehat{Y}}$, $\widehat{y}^t\cdot e = 0$.

♯ $\widehat{\sigma}^2 = \frac{e^te}{n-k}$ estimador insesgado de σ^2 (e^te es la SCR) $\Rightarrow \widehat{Var(\widehat{\beta})} = \widehat{\sigma}^2\cdot\left(X^tX\right)^{-1}$

◇ $\widehat{\sigma}^2 = \frac{y^ty-\widehat{\beta}^tX^ty}{n-k}$

3. Validación del Modelo Lineal

Herramientas para determinar si la estimación realizada es o no válida.

♯ Coeficiente de determinación (R^2): porcentaje de variabilidad explicada por el ajuste (estimación) realizado del modelo.

◇ $R^2 = \frac{SCE}{SCT} = 1 - \frac{SCR}{SCT} \Rightarrow R^2 = \frac{\widehat{\beta}^tX^ty - n\overline{Y}^2}{y^ty-n\overline{Y}^2} = 1 - \frac{y^ty-\widehat{\beta}^tX^ty}{y^ty-n\overline{Y}^2}$

◇ Siempre que el modelo tenga constante: $0\leq R^2\leq 1$.

◇ Cuanto más próximo a 1 mejor será el ajuste.

◇ El coeficiente de determinación será significativo (es decir, validará el modelo) siempre que sea superior a la siguiente cota:

$$\frac{\frac{k-1}{n-k}\cdot F_{k-1,n-k}(1-\alpha)}{1+\frac{k-1}{n-k}\cdot F_{k-1,n-k}(1-\alpha)}.$$

◇ Coeficiente de determinación corregido: $\overline{R}^2 = 1-(1-R^2)\cdot\frac{n-1}{n-k}$.

♯ Selección de modelos:

$$\mathfrak{L} = -\frac{n}{2}\cdot(1+\ln(2\cdot\pi) - ln(n)) - \frac{n}{2}\cdot\ln(SCR)$$

◇ $AIC = -2\cdot\mathfrak{L} + 2\cdot k$.

◇ $BIC = -2\cdot\mathfrak{L} + k\cdot\ln(n)$.

◇ $HQC = -2\cdot\mathfrak{L} + 2\cdot k\cdot\ln\left(\ln(n)\right)$.

♯ Distribuciones:

$$\widehat{\beta}\sim N\left(\beta,\sigma^2\cdot\left(X^tX\right)^{-1}\right) \rightarrow \left(R\widehat{\beta}-R\beta\right)^t\cdot\left[R\cdot\left(X^tX\right)^{-1}\cdot R^t\right]^{-1}\cdot\left(R\widehat{\beta}-R\beta\right)\sim\chi^2_q$$

$$\downarrow$$

$$\frac{(n-k)\cdot\widehat{\sigma}^2}{\sigma^2}\sim\chi^2_{n-k} \rightarrow \left(R\widehat{\beta}-R\beta\right)^t\cdot\frac{\left[R\cdot\left(X^tX\right)^{-1}\cdot R^t\right]^{-1}}{q\cdot\widehat{\sigma}^2}\cdot\left(R\widehat{\beta}-R\beta\right)\sim F_{q,n-k}$$

♯ Contrastes de hipótesis:

◇ Rechazo $H_0: R\beta = r$ si $\left(R\widehat{\beta}-r\right)^t\cdot\frac{\left[R\cdot\left(X^tX\right)^{-1}\cdot R^t\right]^{-1}}{q\cdot\widehat{\sigma}^2}\cdot\left(R\widehat{\beta}-r\right) > F_{q,n-k}(1-\alpha)$.

◇ Rechazo $H_0:\beta_i = b_i$ si $\left|\frac{\widehat{\beta}_i-b_i}{\widehat{\sigma}\sqrt{w_i}}\right| > t_{n-k}\left(1-\frac{\alpha}{2}\right)$, w_i elemento (i,i) de $\left(X^tX\right)^{-1}$.

♯ Análisis de la varianza (ANOVA):

◇ Rechazo $H_0:\beta_2=\beta_3=\cdots=\beta_k=0$ si $\frac{\frac{SCE}{k-1}}{\frac{SCR}{n-k}} = \frac{\frac{R^2}{k-1}}{\frac{1-R^2}{n-k}} > F_{k-1,n-k}(1-\alpha)$.

♯ Intervalos de confianza:

◇ Para $R\beta$: $\left(R\widehat{\beta}-r\right)^t\cdot\left[R\cdot\left(X^tX\right)^{-1}\cdot R^t\right]^{-1}\cdot\left(R\widehat{\beta}-r\right)\leq q\cdot\widehat{\sigma}^2\cdot F_{q,n-k}(1-\alpha)$.

◇ Para β_i: $\widehat{\beta}_i\pm t_{n-k}\left(1-\frac{\alpha}{2}\right)\cdot\widehat{\sigma}\cdot\sqrt{w_i}$.

◇ Para σ^2: $\left[\frac{(n-k)\cdot\widehat{\sigma}^2}{\chi^2_{n-k}\left(1-\frac{\alpha}{2}\right)}, \frac{(n-k)\cdot\widehat{\sigma}^2}{\chi^2_{n-k}\left(\frac{\alpha}{2}\right)}\right] = \left[\frac{SCR}{\chi^2_{n-k}\left(1-\frac{\alpha}{2}\right)}, \frac{SCR}{\chi^2_{n-k}\left(\frac{\alpha}{2}\right)}\right]$.

4. Explotación del Modelo Lineal

¿Qué ocurre para nueva información recogida en x_0?

♯ Predictor puntual: $p_0 = x_0^t\cdot\widehat{\beta}$.

$\hookrightarrow$ lineal, insesgado ($E[p_0] = x_0^t\cdot\beta$) y óptimo (mínima varianza).

♯ Predictor por intervalo:

◇ Para el valor esperado: $x_0^t\cdot\widehat{\beta}\pm t_{n-k}\left(1-\frac{\alpha}{2}\right)\cdot\widehat{\sigma}\cdot\sqrt{x_0^t\cdot\left(X^tX\right)^{-1}\cdot x_0}$.

◇ Para la permanencia estructural: $x_0^t\cdot\widehat{\beta}\pm t_{n-k}\left(1-\frac{\alpha}{2}\right)\cdot\widehat{\sigma}\cdot\sqrt{1+x_0^t\cdot\left(X^tX\right)^{-1}\cdot x_0}$.

5. Estimación con información a priori

¿Cómo estimar β sabiendo que verifica que $R\cdot\beta = r$ (q restricciones)?

♯ Mínimos Cuadrados Restringidos

◇ $\widehat{\beta}_R = \widehat{\beta} + \left(X^t\cdot X\right)^{-1}\cdot R^t\cdot\left[R\cdot\left(X^t\cdot X\right)^{-1}\cdot R^t\right]^{-1}\cdot\left(r-R\cdot\widehat{\beta}\right)$.

$\hookrightarrow$ insesgado (siempre que $r = R\cdot\beta$) y óptimo $\left(Var\left(\widehat{\beta}_R\right)\leq Var\left(\widehat{\beta}\right)\right)$.

♯ Consecuencias:

◇ $R_R^2\leq R^2$.

◇ $SCR = e^te\leq e_R^te_R = SCR_R$.

◇ $\widehat{\sigma}_R^2 = \frac{e_R^te_R}{n-k+q}$.

◇ Rechazamos $H_0: R\cdot\beta = r$ si $\frac{(e_R^te_R-e^te)/q}{e^te/(n-k)} > F_{q,n-k}(1-\alpha)$.

6. Multicolinealidad

El problema de multicolinealidad consiste en la existencia de relaciones lineales entre dos o más variables independientes del modelo lineal uniecuacional multiple.

♯ Multicolinealidad perfecta o exacta: la relación lineal es perfecta.

◇ La matriz X no es de rango completo por columnas, esto es, $rg(X) < k$. Entonces es imposible obtener una solución única para $\widehat{\beta}$.

♯ Multicolinealidad aproximada: la relación lineal es aproximada.

◇ No se incumplirá la hipótesis básica de que la matriz X sea completa por columnas. Sin embargo, el determinante de X^tX será muy próximo a cero, por lo que $(X^tX)^{-1}$ tenderá a tener valores altos y se obtendrán estimaciones inestables.

♯ Causas:

◇ relación causal entre variables explicativas del modelo.

◇ escasa variabilidad en las observaciones de las variables independientes.

◇ reducido tamaño de la muestra.

♯ Soluciones:

◇ mejora del diseño muestral estrayendo la información máxima de la variables observadas.

◇ eliminación de las variables que se sospechan son causantes de la multicolinealidad.

◇ en caso de disponer de pocas observaciones, aumentar el tamaño de la muestra.

7. Heteroscedasticidad

Cuando se incumple el supuesto de homocedasticidad, es decir, la varianza no es constante sino que varía con la observación ($E[u_t^2] = \sigma_t^2 = \sigma^2\cdot w_t, \forall t$), se dice que hay heteroscedasticidad.

♯ Este problema aparece especialmente cuando se disponen de datos de sección cruzada.

♯ La consecuencia de la presencia de heteroscedasticidad en un modelo lineal es que los estimadores obtenidos aunque serán lineales e insesgados no serán óptimos.

♯ Procedimientos de Detección:

◇ Gráfico de los residuos.

◇ Gráficos de dispersión de los residuos frente a la variable que se sospecha provoca la heteroscedasticidad.

◇ Tests de Glesjer y de Goldfeld-Quandt (para muestras pequeñas y heteroscedasticidad provocada por una variable).

◇ Tests de White y de Breusch-Pagan (muestras grandes y no se especifica la variable que provoca la heteroscedasticidad).

♯ Estimación bajo heteroscedasticidad: Mínimos Cuadrados Ponderados.

◇ Transformar los datos según:

$$Y_t^* = \frac{Y_t}{\sqrt{w_t}}, \quad X_{it}^* = \frac{X_{it}}{\sqrt{w_t}}, \quad i=1,\dots,k, \quad t=1,\dots,n.$$

◇ Estimar el modelo transformado por MCO.

8. Autocorrelación

Cuando se incumple el supuesto de incorrelación, es decir, la covarianza de la perturbación aleatoria es no nula para dos instantes de tiempo distintos ($E[u_i\cdot u_j]\neq 0, \forall i\neq j$ o, equivalentemente, $E[u_t\cdot u_{t-k}]\neq 0, \forall k>0$), se dice que hay autocorrelación.

♯ Este problema aparece especialmente cuando se disponen de datos de series temporales.

♯ La consecuencia de la presencia de autocorrelación en un modelo lineal es, como es sabido, que los estimadores obtenidos aunque serán lineales e insesgados no serán óptimos.

♯ Suponemos que la estructura de autocorrelación viene marcada por un proceso autorregresivo de orden 1, esto es:

$$u_t = \rho\cdot u_{t-1} + v_t.$$

♯ Procedimientos de Detección:

◇ Gráfico temporal de los residuos.

◇ Gráficos de dispersión de los residuos frente algún retardo suyo.

◇ Funciones de autocorrelación (ACF) y autocorrelación parcial (ACP).

◇ Contrastes de Durbin-Watson y de Ljung-Box.

♯ Estimación bajo autocorrelación: Prais-Winsten (muestras pequeñas).

◇ Estimar el modelo original por MCO.

◇ Utilizar los residuos del ajuste anterior para estimar ρ.

◇ Transformar los datos según:

$$Y_t^* = \begin{cases} \sqrt{1-\rho^2}\cdot Y_1 & t=1 \\ Y_t - \rho\cdot Y_{t-1} & t>1 \end{cases}, \quad X_{it}^* = \begin{cases} \sqrt{1-\rho^2}\cdot X_{i1} & t=1 \\ X_{it}-\rho\cdot X_{it-1} & t>1 \end{cases},$$

con $i=1,\dots,k$.

◇ Estimar el modelo transformado por MCO volviendo al primer paso.

◇ Repetir este proceso hasta que la diferencia entre dos estimaciones consecutivas de ρ sea menor que un valor prefijado (de orden 10^{-3} normalmente).

♯ Estimación bajo autocorrelación: Cochrane-Orcutt (muestras grandes).

◇ Cuando el número de observaciones es suficientemente grande se puede despreciar la primera observación (perdiéndola) transformando los datos según

$$Y_t^* = Y_t - \rho\cdot Y_{t-1}, \quad X_{it}^* = X_{it} - \rho\cdot X_{it-1},$$

para $t>1$ e $i=1,\dots,k$.

◇ Proceso iterativo análogo al anterior.