Università di Pavia

Il modello di regressione lineare multipla con regressori stocastici

Eduardo Rossi

Modellare l'esperimento casuale bivariato nel quale le variabili casuali (Y, X) sono rivelate all'osservatore in modo sequenziale, prima X dopo Y. Se $E(Y)$ rappresenta la previsione di Y prima che l'esperimento cominci, allora $E[Y|x]$ potrebbe essere definito come la previsione di Y dopo il verificarsi dell'evento $\{X = x\}$. Prima dell'osservazione di X l'osservatore può assegnare una distribuzione alle previsioni condizionali di Y, che saranno formulate dopo l'osservazione della realizzazione di X. Questo significa trattare $E[Y|X]$ come una v.c.

Se X e Y sono v.c. continue, la media di Y condizionata a $X = x$ si definisce

$$E[Y|X = x] = \int y f(y|X = x) dy$$

come la media di una v.c. con densità di probabilità $f(y|X = x)$.

Se $E|g(Y)| < \infty$, allora la media di $g(Y)$ condizionata a $X = x$

$$E[g(Y)|X = x] = \int g(y) f(y|X = x) dy$$

La media condizionale di $E[g(Y)|X = x]$ è una variabile casuale, possiamo indicarla con $h(X)$.

$$g(y, x) \equiv [y - E(Y|X = x)]^2$$

allora

$$E[g(Y, X)|X = x]$$

è detta *varianza condizionale* di Y condizionata a X, $Var[Y|X]$.

Siano Y e X v.c. definite sulle stesso spazio probabilizzato.

1. Se $E|Y| < \infty$, allora

$$E_X[E_{Y|X}(Y|X)] = E[Y] \quad \textbf{La legge dei valori attesi iterati}$$

Infatti,

$$
\begin{aligned}
E[Y] &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{YX}(y,x)dxdy \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} y f_{Y|X}(y|x) f_X(x)dxdy \\
&= \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} y f_{Y|X}(y|x)dy\right] f_X(x)dx \\
&= \int_{-\infty}^{\infty} E_{Y|X}[Y|X] f_X(x)dx \\
&= E_X[E_{Y|X}[Y|X]]
\end{aligned}
$$

2. Se $h(X)$ è limitata e $E|g(Y)| < \infty$, allora

$$E[h(X)g(Y)|X] = h(X)E[g(Y)|X]$$

$$E[h(X)g(Y)] = E\{h(X)E[g(Y)|X]\}$$

3. La legge dei valori attesi iterati nella sua formulazione generale dice che:

$$E_{Y,X}[h(Y,X)] = E_X\{E_{Y|X}[h(Y,X)]\}$$

Se $E_{Y,X}\{[Y-g(X)]^2\} < \infty$, allora

$$\begin{aligned} E_{Y,X}\{[Y-g(X)]^2\} &= E_X\{E_{Y|X}[Y-g(X)]^2\} \\ &= E_X\{E_{Y|X}[Y-E(Y|X)+E(Y|X)-g(X)]^2\} \\ &= E_X[Var(Y|X)] + E_X\{[g(X)-E(Y|X)]^2\} \end{aligned}$$

Se $Var(Y)$ esiste e si pone $g(x) = E(Y), \forall x$ allora

$$Var(Y) = E_X[Var(Y|X)] + Var_X[E(Y|X)]$$

dove $Var_X[E(Y|X)] = E_X[E(Y|X)-E(Y)]^2$

Siano X, Y, W v.c. definite sullo stesso spazio probabilizzato e si indichi con $f_{W|X}(w|x)$ la densità di W condizionata a $X = x$. Allora

$$\begin{aligned} E[E(Y|X,W)|X = x] &= \int E(Y|X = x, W = w) f_{W|X}(w|x) dw \\ &= E(Y|X = x) \end{aligned}$$

e quindi $E[E(Y|X,W)|X] = E(Y|X)$.

La media condizionale, $E[Y|X]$, è anche denominata **funzione di regressione**, descrive come varia la media di Y al variare di X. E' quindi funzione delle realizzazioni di X.

Se la funzione di regressione è lineare in X si dice **retta di regressione**:

$$E[Y|X] = \beta_1 + \beta_2 X$$

$$E[Y] = E_X[E[Y|X]] = E[\beta_1 + \beta_2 X] = \beta_1 + \beta_2 E_X[X]$$

da cui

$$\beta_1 = \mu_y - \beta_2 \mu_x$$

dove

$$\mu_y \equiv E[Y]$$

$$\mu_x \equiv E[X]$$

$$
\begin{aligned}
Cov[Y,X] &= E[(Y-\mu_y)(X-\mu_x)] \\
&= E_X[E[(Y-\mu_y)(X-\mu_x)|X]] \\
&= E_X[(X-\mu_x)E[(Y-\mu_y)|X]] \\
&= E_X[(X-\mu_x)(\beta_1+\beta_2 X-\mu_y)] \\
&= \beta_1 E_X[(X-\mu_x)] + \beta_2 E_X[(X-\mu_x)X] - \mu_x E_X[(X-\mu_x)] \\
&= \beta_2 E_X[X^2] - \beta_2 \mu_x^2 \\
&= \beta_2 Var[X]
\end{aligned}
$$

quindi

$$\beta_2 = \frac{Cov[Y,X]}{Var[X]}$$

$$\beta_1 = \mu_y - \frac{Cov[Y,X]}{Var[X]}\mu_x$$

La variabile aleatoria Y è detta indipendente in media dalla X se e soltanto se la sua funzione di regressione non dipende da X

$$E[Y|X] = E[Y]$$

Se Y è indipendente in media da X allora Y e X hanno covarianza nulla.

- **Analisi di Correlazione**: valutare il grado di associazione lineare fra due variabili
- **Analisi di Regressione**: studio della dipendenza di una variabile (la variabile dipendente) da una o più variabili esplicative per stimare o predire il valore medio della dipendente in funzione di valori noti o fissi (in campioni ripetuti) delle variabili indipendenti

Survey di redditi familiari: Un campione di famiglie è estratto casualmente ed i loro redditi e spese per varie categorie di beni e servizi sono registrati.
Un econometrico può, per esempio, regredire le spese per l'alimentazione sul reddito disponibile, la composizione familiare, il numero dei bambini, il livello d'istruzione, ecc.

In questo modello i dati non sono sperimentali ed i regressori non possono essere descritti come fissi in campioni ripetuti. Se estraiamo un nuovo campione di famiglie, una nuova $\mathbf{y}$ ed una nuova $\mathbf{X}$ sono selezionate casualmente ogni volta.

Perchè la media condizionale di $\mathbf{Y}$ dato $\mathbf{X}$?

La media condizionale è una funzione ottimale per la previsione.

Una misura di accuratezza previsiva è il *Mean Squared Error* (**MSE**) (Errore qudratico medio):

$$E[(y_t - m(\mathbf{X}))^2]$$

è la media (o aspettativa) dell'errore di previsione al quadrato.

La media condizionale è una funzione di previsione ottimale (nel senso del MSE) relativamente a tutte le altre funzioni delle variabili condizionanti.

Supponiamo che i primi due momenti condizionali di y_t dato $\mathbf{X}$ esistano. La media condizionale della variabile casuale y_t date le variabili casuali in $\mathbf{X}$, $E[y_t|\mathbf{X}]$, è una funzione di previsione di y_t dato $\mathbf{X}$ con il minimo MSE.

Prova. Indichiamo con

$$\mu_t(\mathbf{X}) \equiv E[y_t|\mathbf{X}]$$

$$\begin{aligned} E[(y_t - m_t(\mathbf{X}))^2|\mathbf{X}] &= E[(y_t - \mu_t(\mathbf{X}) + \mu_t(\mathbf{X}) - m_t(\mathbf{X}))^2|\mathbf{X}] \\ &= Var[y_t|\mathbf{X}] + (\mu_t(\mathbf{X}) - m_t(\mathbf{X}))^2 \end{aligned}$$

perchè

$$\begin{aligned} E[(\mu_t(\mathbf{X}) - m_t(\mathbf{X}))(y_t - \mu_t(\mathbf{X}))|\mathbf{X}] &= (\mu_t(\mathbf{X}) - m_t(\mathbf{X}))E[(y_t - \mu_t(\mathbf{X}))|\mathbf{X}] \\ &= (\mu_t(\mathbf{X}) - m_t(\mathbf{X})) \cdot 0 \end{aligned}$$

Il MSE è quindi pari a:

$$E[(y_t - m_t(\mathbf{X}))^2] = E[Var[y_t|\mathbf{X}]] + E[(\mu_t(\mathbf{X}) - m_t(\mathbf{X}))^2]$$

minimizzato quando

$$E[(y_t - m_t(\mathbf{X}))^2] = E[Var[y_t|\mathbf{X}]]$$

cioè

$$E[(\mu_t(\mathbf{X}) - m_t(\mathbf{X}))^2] = 0$$

$$m_t(\mathbf{X}) = \mu_t(\mathbf{X}) \equiv E[y_t|\mathbf{X}]$$

Attenzione: La media condizionale non è necessariamente una funzione lineare, né lo sono le funzioni di previsione che la dominano in termini di MSE.

1. La media condizionale è lineare:

$$E[y_t|\mathbf{x}_t] = \mathbf{x}_t'\boldsymbol{\beta}$$

2. Campionamento casuale. Per ogni istante (unità) t un nuovo vettore (y_t, x_t') è estratto dalla popolazione in modo indipendente.

La conoscenza di x_s per ogni $s \neq t$ non può aiutare nella previsione di y_t.: $(y_t, \mathbf{x}_t)$ sono estratti in modo indipendente, questo significa:

$$E[y_t|\mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_N] = \mathbf{x}_t'\boldsymbol{\beta} \quad t = 1, \ldots, N$$

e

$$E[y_t|\mathbf{x}_1, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_N, y_1, \ldots, y_{t-1}, y_{t+1}, \ldots, y_N] = \mathbf{x}_t'\boldsymbol{\beta}$$

3. Il rango di $\mathbf{X}$ è K.

4. Il termine di disturbo

$$\begin{aligned} E[\varepsilon|\mathbf{X}] &= \mathbf{0} \\ E[\varepsilon\varepsilon'|\mathbf{X}] &= \sigma^2\mathbf{I}_N \end{aligned}$$

In modo non condizionale, per la legge delle aspettative iterate:

$$\begin{aligned} E\{E[\varepsilon|\mathbf{X}]\} = E[\varepsilon] &= \mathbf{0} \\ E\{E[\varepsilon\varepsilon'|\mathbf{X}]\} = E[\varepsilon\varepsilon'] &= \sigma^2\mathbf{I}_N \end{aligned}$$

Le proprietà statistiche dello stimatore OLS dipendono dalle assunzioni sul processo che genera i dati.
Questa dipendenza contrasta con le proprietà geometriche che non richiedono assunzioni.

Le assunzioni statistiche potrebbero venire meno. Si possono effettuare dei controlli diagnostici per valutare l'evidenza contro le assunzioni.

Non distorsione

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

$$E[\widehat{\boldsymbol{\beta}}|\mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\varepsilon|\mathbf{X}] = \boldsymbol{\beta} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' \cdot \mathbf{0} = \boldsymbol{\beta}$$

Lo stimatore OLS è condizionalmente non distorto, ma anche non condizionatamente (per la legge delle aspettative iterate):

$$E\{E[\widehat{\boldsymbol{\beta}}|\mathbf{X}]\} = E[\widehat{\boldsymbol{\beta}}] = \boldsymbol{\beta}$$

inoltre,

$$E\left[\mathbf{X}\widehat{\beta}\,|\mathbf{X}\right] = \mathbf{X}\beta$$

$$E\left[\mathbf{y} - \mathbf{X}\widehat{\beta}\,|\mathbf{X}\right] = \mathbf{0}$$

infatti

$$\begin{aligned} E\left[\mathbf{y} - \mathbf{X}\widehat{\beta}\,|\mathbf{X}\right] &= E\left[\mathbf{y}|\mathbf{X}\right] - E\left[\mathbf{X}\widehat{\beta}\,|\mathbf{X}\right] \\ &= \mathbf{X}\beta - \mathbf{X}E[\widehat{\beta}|\mathbf{X}] \\ &= \mathbf{X}\beta - \mathbf{X}\beta = \mathbf{0} \end{aligned}$$

Varianza dello stimatore OLS:

$$\begin{aligned} Var[\widehat{\boldsymbol{\beta}}|\mathbf{X}] &= E[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\mathbf{X}] \\ &= E[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\varepsilon\varepsilon'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}|\mathbf{X}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'E[\varepsilon\varepsilon'|\mathbf{X}]\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= \sigma^2(\mathbf{X}'\mathbf{X})^{-1} \end{aligned}$$

La matrice di covarianza misura quanto informatico è il campione per i parametri. La varianza non condizionale

$$Var[\widehat{\boldsymbol{\beta}}] = E\{Var[\widehat{\boldsymbol{\beta}}|\mathbf{X}]\} = \sigma^2 E[(\mathbf{X}'\mathbf{X})^{-1}]$$

Se viene ripetuto l'esperimento casuale con estrazioni casuali di $\mathbf{X}$, la distribuzione di $\widehat{\beta}$ è descritta da $Var[\widehat{\boldsymbol{\beta}}]$.

$$\widehat{\boldsymbol{\epsilon}} = \mathbf{y} - \mathbf{X}\widehat{\boldsymbol{\beta}} = (\mathbf{I}_N - \mathbf{P}_X)\mathbf{y}$$

$$\begin{aligned}
\widehat{\boldsymbol{\epsilon}} &= \mathbf{M}_X\mathbf{y} \\
&= \mathbf{M}_X(\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}) \\
&= \mathbf{M}_X\mathbf{X}\boldsymbol{\beta} + \mathbf{M}_X\boldsymbol{\epsilon} \\
&= \mathbf{M}_X\boldsymbol{\epsilon}
\end{aligned}$$

Sebbene i residui siano stime di variabili non correlate per assunzione risultano correlati

$$E[\widehat{\boldsymbol{\epsilon}\boldsymbol{\epsilon}'}|\mathbf{X}] = E[\mathbf{M}_X\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\mathbf{M}_X'|\mathbf{X}] = \sigma^2\mathbf{M}_X$$

la distribuzione è singolare, la matrice di varianza-covarianza è singolare con rango $N - K$. Questa è la conseguenza dell'ortogonalità con $\mathbf{X}$.

$$E[\varepsilon_t^2|\mathbf{X}] = \sigma^2$$

Per la legge delle aspettative iterate:

$$E\{E[\varepsilon_t^2|\mathbf{X}]\} = E[\varepsilon_t^2] = \sigma^2$$

Stimatore non distorto:

$$s^2 = \frac{\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}}{N-K}$$

Per dimostrare la correttezza usiamo le seguenti proprietà della traccia

$$\begin{aligned} a &= tr(a) \quad a \in \mathbb{R} \\ tr(\mathbf{AB}) &= tr(\mathbf{BA}) \end{aligned}$$

$$
\begin{aligned}
E[s^2|\mathbf{X}] &= \frac{E[\boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon}|\mathbf{X}]}{N-k} \\
&= \frac{E[tr(\boldsymbol{\epsilon}'\mathbf{M}\boldsymbol{\epsilon})|\mathbf{X}]}{N-K} \\
&= \frac{E[tr(\mathbf{M}\boldsymbol{\epsilon}\boldsymbol{\epsilon}')|\mathbf{X}]}{N-K} = \frac{tr[E(\mathbf{M}\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X})]}{N-K} \\
&= \frac{tr[\mathbf{M}E(\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X})]}{N-K} \\
&= \frac{tr(\mathbf{M}\sigma^2\mathbf{I}_N)}{N-K} = \frac{tr(\sigma^2\mathbf{M})}{N-K} = \sigma^2\frac{tr(\mathbf{M})}{N-K} = \sigma^2\frac{N-K}{N-K} \\
&= \sigma^2
\end{aligned}
$$

Non condizionatamente:

$$E\left[\frac{\widehat{\boldsymbol{\epsilon}}'\widehat{\boldsymbol{\epsilon}}}{N-K}\right]=\sigma^2$$

s^2 è corretto solo nel caso di disturbi omoschedastici $(E[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'] = \sigma^2\mathbf{I}_N)$.

Teorema di Gauss-Markov

$\widehat{\boldsymbol{\beta}}$ è fra gli stimatori corretti quello con la varianza condizionale più piccola (nel senso delle matrici semidefinite positive).