Università di Pavia

Previsioni

Eduardo Rossi

Dato il MRLM:

$$\mathbf{y} = \mathbf{X}\beta + \epsilon$$

il valore

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

cioè la stima di

$$E[\mathbf{y}|\mathbf{X}]$$

rappresenta la previsione di $\mathbf{y}$.

Supponiamo che la stima del modello sia stata fatta con N osservazioni e che si vogliano calcolare le previsioni per n^0 osservazioni i cui valori sono supposti generati da

$$\mathbf{y}^0 = \mathbf{X}^0\beta + \epsilon^0$$

ovviamente non si osservano i valori in $\mathbf{y}^0$, ma solo quelli in $\mathbf{X}^0$. Costruiamo una regressione aumentata:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{X}^0 & -\mathbf{I}_{n^0} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} + \begin{bmatrix} \epsilon \\ \epsilon^0 \end{bmatrix}$$

$$\widetilde{\mathbf{y}} = \widetilde{\mathbf{X}}\delta + \mathbf{u}$$

sono state aggiunte n^0 osservazioni ed un insieme di n^0 nuove variabili.

Ogni colonna nel blocco

$$\begin{bmatrix} \mathbf{0} \\ -\mathbf{I}_{n^0} \end{bmatrix}$$

contiene una variabile dummy che assume valore -1. Il nuovo vettore γ corrisponde a $\mathbf{y}^0$. La regressione di $\widetilde{\mathbf{y}}$ su $\widetilde{\mathbf{X}}$ produce $\left[\widehat{\beta}', \widehat{\gamma}'\right]$

$$\begin{array}{ccc} \widehat{\beta} & \text{stima di} & \beta \\ \widehat{\gamma} & \text{previsione di} & \mathbf{y}^0 \end{array}$$

Infatti,

$$\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{X}^0 & -\mathbf{I}_{n^0} \end{bmatrix}$$

$$\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}} = \begin{bmatrix} \mathbf{X}'\mathbf{X} + \mathbf{X}^{0\prime}\mathbf{X}^0 & -\mathbf{X}^{0\prime} \\ -\mathbf{X}^0 & \mathbf{I}_{n^0} \end{bmatrix}$$

$$\widetilde{\mathbf{X}}'\widetilde{\mathbf{y}} = \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{0} \end{bmatrix}$$

$$\left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} = \begin{bmatrix} (\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{0\prime} \\ \mathbf{X}^0(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{I} + \mathbf{X}^0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{0\prime} \end{bmatrix}$$

$$\widehat{\delta} = \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1}\widetilde{\mathbf{X}}'\widetilde{\mathbf{y}} = \begin{bmatrix} (\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{0\prime} \\ \mathbf{X}^0(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{I} + \mathbf{X}^0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{0\prime} \end{bmatrix} \begin{bmatrix} \mathbf{X}'\mathbf{y} \\ \mathbf{0} \end{bmatrix}$$

$$\widehat{\delta} = \begin{bmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{bmatrix} = \begin{bmatrix} (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \\ \mathbf{X}^0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} \end{bmatrix}$$

I residui della regressione per le prime N osservazioni sono gli originali residui OLS e per i restanti n^0 sono uguali a zero:

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{0} \end{bmatrix} - \begin{bmatrix} \mathbf{X} & \mathbf{0} \\ \mathbf{X}^0 & -\mathbf{I}_{n^0} \end{bmatrix} \begin{bmatrix} \widehat{\beta} \\ \widehat{\gamma} \end{bmatrix} = \begin{bmatrix} \mathbf{y} - \mathbf{X}\widehat{\beta} \\ \widehat{\gamma} - \mathbf{X}^0\widehat{\beta} \end{bmatrix}$$

$$\widetilde{\mathbf{y}} - \widetilde{\mathbf{X}}\widehat{\delta} = \begin{bmatrix} \widehat{\epsilon} \\ \mathbf{0} \end{bmatrix}$$

La matrice di varianza e covarianza stimata per $\widehat{\gamma}$ contiene, nel blocco (1,1), la matrice di covarianza per le stime $\widehat{\beta}$ e nel blocco (2,2), la varianza-covarianza per gli errori di previsione

$$\widetilde{s}^2 = \frac{\widetilde{\mathbf{u}}'\widetilde{\mathbf{u}}}{N + n^0 - (K + n^0)} = \frac{\widehat{\epsilon}'\widehat{\epsilon}}{N - K} = s^2$$

$$\begin{aligned}
\widetilde{s}^2 \left(\widetilde{\mathbf{X}}'\widetilde{\mathbf{X}}\right)^{-1} &= \widetilde{s}^2 \begin{bmatrix} (\mathbf{X}'\mathbf{X})^{-1} & (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{0\prime} \\ \mathbf{X}^0(\mathbf{X}'\mathbf{X})^{-1} & \mathbf{I} + \mathbf{X}^0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{0\prime} \end{bmatrix} \\
&= \begin{bmatrix} \widehat{Var(\widehat{\beta})} & \widehat{Cov(\widehat{\beta},\widehat{\gamma})} \\ \cdot & \widehat{Var(\widehat{\gamma})} \end{bmatrix}
\end{aligned}$$

$$\begin{aligned}
\widehat{Var(\widehat{\gamma}_i)} &= \mathbf{e}_i'\widehat{Var(\widehat{\gamma})}\mathbf{e}_i \\
&= s^2\mathbf{e}_i'[\mathbf{I} + \mathbf{X}^0(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^{0\prime}]\mathbf{e}_i \\
&= s^2\left[1 + \mathbf{X}^0_{i\cdot}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}^0_{\cdot i}\right]
\end{aligned}$$