Modello SURE

Nel modello SURE, si esegue un approcio più realistico: degli r regressori si usano solo i regressori effettivamente legati alle diverse variabili dipendente, che potrebbero anche essere tutti, come nel caso classico del modello lineare multivariato, ma anche in caso di non significatività potrebbero portare ad una differenziazione delle diverse equazioni. Inoltre questo modello permette di risolvere anche il problema di una numerosità diversa delle osservazioni tra le diverse equazioni. In altre parola nel modello SURE abbiamo regressori diversi per ogni equazione all'interno dell'insieme complessivo dei regresori per l'insieme delle equazioni del modello:

$$\sum_{j=1}^{n} rj$$

Si dà il caso di differenti variabili esplicative per ogni equazione e correlazione tra residui casuali di disturbo associati a diverse equazioni:

$$y_1 = \beta_{10} + \beta_{11}z_1 + \beta_{1m}z_m + \varepsilon_1;$$

....

$$y_j = \beta_{j0} + \beta_{j2}z_2 + \ldots + \beta_{jr}z_r + \varepsilon_j;$$

....

$$y_m = \beta_{m0} + \beta_{mm} z_m + \ldots + \beta_{mr} z_r + \varepsilon_m \ (*)$$

Data n_j come la numerosità delle osservazioni per l'equazione j-esima allora il complesso delle numerosità è dato da

$$\sum_{j=1}^{m} nj$$

La soluzione dei minimi quadrati per la stima dei coefficienti sembra simile a quella dei minimi quadrati generalizzati ma solo in apparenza:

Il modello è caratterizzato dalla presenza delle variabili esplicative diverse da equazione ed equazione.

Gli errori sono:

- Omoschedastici e incorrelati nella stessa equazione
- Eteroschedastici fra diverse equazioni
- Correlati per lo stesso individuo e incorrelati tra individui diversi fra diverse equazioni.

Vi sono due casi importanti in cui le stime SURE risultano equivalenti all'OLS equazione per equazione, in modo che non vi sia alcun guadagno nella stima congiunta del sistema. Questi casi sono:

- Quando la matrice Σ è nota per essere diagonale, cioè non ci sono correlazioni di equazione incrociata tra i termini di errore. In questo caso il sistema diventa apparentemente ma non realmente correlato.
- Quando ogni equazione contiene lo stesso insieme di regressori, ovvero X1 = X2 = …. Xm

Considerando B* come stimatore dei coefficienti del modello OLS e B* come lo stimatore del modello SURE, possiamo affermare che i due risultano identici, ma per la costruzione dello SURE, B* possiede dei parametri uguali a zero proprio per permettere di ottenere equazioni differenziate nel numero delle variabili esplicative. La possibilità di differenziare il numero di covariate risiede infatti nel porre uguale a zero il valore della variabile non presente nell'equazione x così che questa, moltiplicata per il rispettivo coefficiente B generi un'influenza nulla sulla stima di Y.

Quando si omettono delle variabili, essendo i coefficienti di regressione parziali, influenzati, cioè dalle altre variabili esplicative poiché stimati sulla totalità di esse, cambiano tutti i coefficienti di regressione in base all'equazione di riferimento. Infatti, ipotizzando di avere una variabile X_j presente in 2 equazioni distinte il coefficiente corrispondente sarà diverso proprio in virtù del fatto che le stime sono influenzate le une dalle altre.

La scelta del modello tra diverse possibilità avviene in diversi passaggi:

1. Innanzitutto per quanto riguarda gli errori si è richiamato al tema di verifica di omoschedasticità e correlazione all'interno delle equazioni.
2. Riguardo alla scelta tra modello lineare generalizzato e Sure occorre osservare la significatività univariata delle variabili esplicative: si possono scartare le variabili non significative.
3. Valuto poi l'adattamento ai dati insieme alla parsimonia equazione per equazione. Se migliora l'R^2 corretto dal numero di regressori o più semplicemente, come quando costruivo modelli non lineari, eliminando una variabile la diminuzione di R^2 e F è minima, elimino quella variabile dalla specifica equazione.