

## Nota metodologica PLUS n° 2 – Stime a massima verosimiglianza

L'indagine PLUS 2005 prevede una procedura di estrazione probabilistica del campione effettuata in base alla pianificazione ex-ante di particolari domini di studio e strati della popolazione di riferimento. I pesi di riporto all'universo che ne derivano (anche a seguito di un'allocazione vincolata a livelli predefiniti di varianza e di una procedura di calibrazione) costituiscono uno strumento indispensabile ai fini di una "corretta" interpretazione dei risultati delle analisi condotte.

Per quanto riguarda le analisi di regressione (lineare, logistica, ...) possono sorgere dei dubbi sull'opportunità dell'utilizzo di tali pesi. Indicazioni "certe" sembrano non esistere in letteratura, ma ignorare le conoscenze a priori sulla procedura di estrazione delle unità, non appare certamente una scelta "ragionevole".

Ad incoraggiare la ricerca di soluzioni alternative al problema contribuiscono due importanti constatazioni: *i*) l'esistenza di una letteratura particolarmente estesa sull'argomento; *ii*) la possibilità di sfruttare le potenzialità di alcuni software (Stata o SPSS), i quali permettono di implementare nelle procedure di stima le informazioni relative al piano di campionamento implementato.

I comandi [SVY] *svyset*, *svydes*, *svylogit*, *svymlogit*, ... di Stata, consentono di migliorare le procedure di stima in molte analisi econometriche, correggendo le statistiche test in funzione del disegno di campionamento (numerosità e composizione degli strati).

Il principale ambito di applicazione di tali procedure riguarda tutte quelle analisi in cui è previsto l'utilizzo di una funzione di massima verosimiglianza nella stima dei parametri di regressione (e delle statistiche di riferimento per i relativi test di significatività). In particolare, accettato di utilizzare i pesi campionari in tali analisi, sorge il problema di calcolare: *a*) la "giusta" funzione di massima verosimiglianza *b*) implementare le relative procedure di stima (diverse dal caso standard).

L'utilizzo di "procedure pesate", infatti, comporta la non ammissibilità degli stimatori ottenuti mediante l'ordinaria massima verosimiglianza pesata, sostituita dalla "quasi-verosimiglianza" (o "pseudo-verosimiglianza"). Da questa segue la non validità degli ordinari test di significatività: in generale, utilizzando le normali procedure di stima si ottengono degli standard error impropriamente superiori a quanto ottenibile attraverso le procedure *svy*.

**Survey Data Reference Manual:** vedi comandi Stata (v8-v9):

*svyset* e *svydes*; *svylogit*, *svyprobit*, *svymlogit*, *svyologit*, [...]; *lincom* → *nlcom*, *predict* → *predictnl*, *test* → *testnl*;