Laudatio

pronunciata dal Professor Franco Peracchi in occasione della
Laurea Honoris Causa al Professor Charles F. Manski

Tor Vergata, 24 maggio 2006

Il conferimento della Laurea Honoris Causa al Professor Manski costituisce un riconoscimento dei suoi numerosi e fondamentali contributi alla teoria dell'econometria e della statistica, alla teoria delle aspettative, e alla teoria delle decisioni in condizioni di incertezza. Costituisce inoltre un riconoscimento della sua straordinaria capacità di applicare la teoria econometrica e statistica all'analisi di una varietà di problemi concreti in campi assai diversi, che vanno dall'istruzione, alle politiche sociali, ai trasporti.

In particolare, per quanto riguarda la teoria dell'econometria e della statistica, il Professor Manski ha dato contributi fondamentali alla teoria della stima adattiva di modelli di regressione nonlineari, ai metodi di stima basati sul principio dell'analogia, e alla teoria dell'identificazione e della stima di modelli per variabili risposta discrete. Altri contributi fondamentali riguardano il problema dell'identificazione parziale di parametri di una popolazione, il problema della stima di tali parametri in presenza di varie forme di incompletezza dei dati, e il problema della valutazione dell'effetto di un trattamento con dati sperimentali e non sperimentali.

In questo intervento intendo soffermarmi sui contributi fondamentali che il Professor Manski ha dato alla teoria dell'identificazione e della stima in presenza di dati incompleti. Ovviamente, l'incompletezza dei dati può prendere varie forme. Qui intendo considerare due forme particolari:

(i) osservabilità indiretta,

(ii) dati censurati.

In entrambi i casi, il problema è quello di fare inferenza circa la media condizionata $\mu(X) = \mathrm{E}(Y \mid X)$ di una variabile casuale Y, che descrive la variabilità di un fenomeno nella popolazione di interesse, data l'osservazione di X, un vettore causale di covariate. Come noto, questo problema è di importanza fondamentale perché $\mu(X)$ è, tra tutti i predittori di Y, quello con errore quadratico medio minimo.

Per quanto riguarda la prima forma di incompletezza dei dati, consideriamo un semplice problema di scelta tra due alternative (per esempio, prendere l'auto o il trasporto pubblico per raggiungere Tor Vergata). Sia Y la differenza percepita di utilità tra le due alternative, e si assuma che la prima alternativa sia scelta se $Y \leq 0$ e la seconda se $Y > 0$. Si assuma inoltre che Y abbia media finita e che $Y = \mu(X) - U$, dove $\mu(X)$ è la differenza media di utilità tra le due alternative (con X un insieme di caratteristiche osservabili delle alternative o di chi effettua la scelta) e U è un errore di regressione con media nulla e indipendente da X.

Se è disponibile un campione casuale $\{(X_i, Y_i), i = 1, \ldots, n\}$ da (X, Y), allora $\mu(X)$ può essere stimata in vari modi. In particolare, se $\mu(X) = \alpha + \beta X$, dove α e β sono parametri ignoti, allora uno stimatore semplice, ma con buone proprietà statistiche, è lo stimatore dei minimi quadrati ordinari che minimizza la somma degli scarti quadratici $\sum_{i=1}^n (Y_i - \alpha - \beta X_i)^2$. Si noti che questo stimatore richiede solamente che l'errore U abbia media nulla e sia incorrelato con X. Se l'errore U possiede queste due proprietà, lo stimatore dei minimi quadrati ordinari rimane valido qualunque sia la funzione di distribuzione di U.

Cosa succede invece se non osserviamo le differenze di utilità ma solo le scelte, cioè non osserviamo Y ma solo un indicatore D, pari a 0 o 1 a seconda che sia scelta la prima o la seconda alternativa? Poiché $Y = \mu(X) - U$, dove U è indipendente da X con funzione di ripartizione G, è facile mostrare che

$$\mathrm{E}(D \mid X) = \mathrm{E}(1\{Y > 0\} \mid X) = \Pr\{Y > 0 \mid X\} = G(\alpha + \beta X),$$

dove $1\{Y > 0\}$ è un indicatore uguale a 0 se $Y \leq 0$ e a 1 se $Y > 0$ (più in generale, $1\{A\}$ è uguale a 1 se A è vero ed è uguale a 0 altrimenti). Se G è nota, possiamo allora stimare α e β tramite lo stimatore dei minimi quadrati nonlineari che minimizza la somma degli scarti quadratici $\sum_{i=1}^n [Y_i - G(\alpha + \beta X_i)]^2$.

Sebbene si possa sempre assumere una forma particolare per G (per esempio, la gaussiana o la logistica), la credibilità di questa assunzione può essere un problema. Si dimostra inoltre che piccole modifiche nella forma assunta per G possono modificare in modo sostanziale l'inferenza circa $\mu(X)$. In altre parole, a differenza di quando Y è osservabile, l'inferenza ottenuta con una scelta particolare di G non è "robusta". Cosa si può fare allora per stimare α e β se vi è incertezza circa G? Uno dei contributi fondamentali del Professor Manski ̀di avere per primo fornito una risposta a questa domanda. Cercherò ora di presentare in modo intuitivo tale risposta.

Supponiamo che U si distribuisca simmetricamente intorno a zero. In questo caso, la media e la mediana di U sono entrambe uguali a zero, e quindi la media condizionata e la mediana condizionata di Y sono entrambe uguali ad $\alpha + \beta X$. Poiché $D = 1\{Y > 0\}$ è una trasformazione monotona di Y, si dimostra che la mediana condizionata di D è uguale a $1\{\alpha + \beta X > 0 \mid X\}$, è cioè uguale a 1 se $\alpha + \beta X > 0$ e a 0 se $\alpha + \beta X \leq 0$. Si ricordi ora che la mediana ζ di una variabile causale Z necessariamente risolve il problema di minimizzare l'errore assoluto medio $\mathrm{E}\,|Y - c|$ rispetto a c. I parametri della mediana condizionata di Y devono perciò necessariamente risolvere il problema

$$\min_{\alpha,\beta} \mathrm{E}\,|D - 1\{\alpha + \beta X > 0\}|. \tag{1}$$

In base al principio dell'analogia, a cui il Professor Manski ha dedicato nel 1988 un importante volume, uno stimatore di α e β può essere ottenuto minimizzando il corrispondente empirico di (1), cioè risolvendo

$$\min_{\alpha,\beta} n^{-1} \sum_{i=1}^{n} |D_i - 1\{\alpha + \beta X_i > 0\}|. \tag{2}$$

Poiché $|D - 1\{A\}| = 1 - |D\, 1\{A\} + (1-D)\, 1\{A^c\}|$, dove A^c indica il complemento dell'insieme A, il problema (2) è equivalente al problema

$$\max_{\alpha,\beta}\; n^{-1} \sum_{i=1}^{n} |D_i\, 1\{\alpha + \beta X_i > 0\} + (1 - D_i)\, 1\{\alpha + \beta X_i \leq 0\}|,$$

che è il problema che definisce lo stimatore *maximum score* (MS), introdotto da Manski (1975), la cui consistenza per α e β è dimostrata da Manski (1985).

Lo stimatore MS vanta vari primati:

1. È il primo esempio di stimatore semiparametrico per modelli con variabile dipendente discreta.
2. È il primo esempio di stimatore per modelli nonlineari di regressione quantilica, precedendo addirittura lo stimatore introdotto nel 1978 da Roger Koenker e Gib Bassett per modelli lineari di regressione quantilica.

Lo stimatore MS ha anche peculiari caratteristiche di "non regolarità" che lo rendono interessante per uno studioso. Esso infatti non è $\sqrt{n}$-consistente e non è asintoticamente normale (Kim e Pollard 1990). Inoltre, per questo stimatore il *bootstrap* non sembra essere valido (Abrevaya e Huang 2005).

Tra le varie generalizzazioni dei risultati ottenuti in Manski (1975, 1985), mi limito a ricordare l'estensione dello stimatore MS a dati di tipo longitudinale (Manski 1987) e la costruzione di versioni "quasi regolari" dello stimatore MS (Horowitz 1992, 2002).

Consideriamo ora una seconda forma di incompletezza dei dati, il caso cioè in cui Y è osservabile solo per una frazione della popolazione (per esempio, nel caso di un'indagine sul reddito delle famiglie, è osservabile solo il reddito di quelle famiglie che hanno accettato di partecipare all'indagine). Sia $Y = \mu(X) + U$ e sia ora D un indicatore uguale a 1 se Y è osservato e a 0 se Y non è osservato. Il problema posto da questa seconda forma di incompletezza è il seguente: possiamo stimare correttamente $\mu(X)$ usando solo i dati censurati, cioè le osservazioni per cui $D = 1$? Useremo il termine "problema di selezione" per indicare questo problema.

Il problema di selezione ha risposte diverse a seconda di come opera il processo di selezione. I dati censurati forniscono infatti informazione utile per stimare non $\mu(X)$ bensì

$$\mathrm{E}(Y \mid X, D = 1) = \mu(X) + \mathrm{E}(U \mid X, D = 1).$$

Se l'errore di regressione U è indipendente in media da D (magari condizionatamente a X), allora $\mathrm{E}(Y \mid X, D = 1) = \mu(X)$. In questo caso, che chiameremo di selezione esogena, il problema di selezione ha una risposta positiva, possiamo cioè stimare correttamente $\mu(X)$ usando solo i dati censurati.

Supponiamo però che, anche dopo il condizionamento su X, l'errore U non sia indipendente in media da D. In questo caso, che chiameremo di selezione endogena, $\mathrm{E}(Y \mid X, D = 1) \neq \mu(X)$ e quindi non possiamo stimare $\mu(X)$ usando solo i dati censurati. Cosa possiamo fare in questo caso?

Una strategia è quella di assumere uno specifico modello per D, per esempio $D = 1\{\delta + \gamma Z + V > 0\}$ (un modello cioè analogo a quello usato nel caso di osservabilità indiretta) dove Z è un vettore di covariate che aiutano a "spiegare" l'inosservabilità di Y e V è un errore stocastico, e derivare stimatori di $\mu(X)$ sfruttando opportunamente questa informazione. Il Premio Nobel a James Heckman è stato un riconoscimento del valore pratico di questa strategia.

Come nel caso di osservabilità indiretta, però, questa strategia incontra problemi seri di credibilità e di robustezza. Mostrare che una seconda stategia è possibile rappresenta un altro dei contributi fondamentali del Professor Manski.

Il nucleo centrale del ragionamento, delineato per la prima volta in Manski (1989) e poi sviluppato in una serie di lavori successivi, è semplice e intuitivo. Per la proprietà delle medie, abbiamo che

$$\mu(X) = \mu_0(X)\,\pi(X) + \mu_1(X)\,[1 - \pi(X)],$$

dove $\mu_0(X)$ è la media condizionata di Y quando Y è inosservata ($D = 0$), $\mu_1(X)$ è la media condizionata di Y quando Y è osservata ($D = 1$), e $\pi(X) = \Pr\{D = 0 \,|\, X\}$ è la probabilità condizionata di non osservare Y. La difficoltà di stimare $\mu(X)$ in modo credibile nasce dal fatto che, mentre è possibile stimare correttamente $\pi(X)$ e $\mu_1(X)$ semplicemente sulla base dei dati, ciò non è possibile invece per $\mu_0(X)$. Si noti che usare una stima di $\mu_1(X)$ per stimare $\mu(X)$ darebbe luogo a una distorsione pari a

$$\mu_1(X) - \mu(X) = [\mu_1(X) - \mu_0(X)]\,\pi(X).$$

Tale distorsione è nulla se $\mu_1(X) = \mu_0(X)$ (cioè Y è indipendente in media da D condizionatamente a X) oppure $\pi(X) = 0$ (assenza di selezione). Altrimenti, essa è tanto maggiore (in valore assoluto) quanto maggiore è la differenza $\mu_1(X) - \mu_0(X)$ oppure la probabilità di non osservare Y.

Supponiamo ora di non conoscere (o non poter stimare correttamente) $\mu_0(X)$, la media condizionata di Y quando Y è inosservata, ma di sapere che essa appartiene a un intervallo chiuso $[a(X), b(X)]$, cioè di sapere che $a(X) \leq \mu_0(X) \leq b(X)$. Vale allora la seguente disuguaglianza

$$a(X)\,\pi(X) + \mu_1(X)\,[1 - \pi(X)] \leq \mu(X) \leq b(X)\,\pi(X) + \mu_1(X)\,[1 - \pi(X)], \tag{3}$$

dove il limite inferiore corrisponde allo "scenario peggiore" per $\mu_0(X)$ e quello superiore allo "scenario migliore". L'intervallo di valori definito dalla (3) rappresenta l'insieme dei valori plausibili per $\mu(X)$ alla luce dei dati e dell'informazione a priori circa $\mu_0(X)$. La sua ampiezza, uguale a $[b(X) - a(X)]\,\pi(X)$, rappresenta quindi una misura della nostra incertezza circa $\mu(X)$ alla luce dei dati e dell'informazione circa $\mu_0(X)$. Sebbene $\mu(X)$ rimanga ignota, la (3) ci consente di "identificarla parzialmente". Avere sottolineato il valore di questa informazione "intervallare", specialmente nei due volumi del 1995 e del 2003, rappresenta uno dei grandi meriti del Professor Manski.

Un caso interessante, perchè non richiede alcuna informazione a priori circa $\mu_0(X)$, si ha quando $Y = 1\{Z \leq z\}$, cioè Y è l'indicatore binario dell'evento che la variabile causale Z non assuma valori superiori a z. In questo caso, $\mu(X)$ corrisponde al valore $F(z \,|\, X) = \Pr\{Z \leq z \,|\, X\}$ della funzione di ripartizione condizionata di Z al punto z, mentre $\mu_0(X)$ e $\mu_1(X)$ corrispondono rispettivamente a $F_0(z \,|\, X) = \Pr\{Z \leq z \,|\, X, D = 0\}$ e a $F_1(z \,|\, X) = \Pr\{Z \leq z \,|\, X, D = 1\}$. Poiché una probabilità deve necessariamente prendere valori compresi tra 0 e 1, la disuguaglianza (3) diventa

$$F_1(z \,|\, X)\,[1 - \pi(X)] \leq F(z \,|\, X) \leq \pi(X) + F_1(z \,|\, X)\,[1 - \pi(X)]. \tag{4}$$

In questo caso, l'ampiezza dell'intervallo di valori plausibili per $F(z \,|\, X)$ è semplicemente $\pi(X)$, la probabilità condizionata di non osservare Y. Poiché la funzione di ripartizione condizionata $F(z \,|\, X)$ fornisce una caratterizzazione della distribuzione di probabilità di Z (condizionatamente a X), la disuguaglianza (4) permette di "identificare parzialmente", pur in presenza di censura, l'intero meccanismo probabilistico che descrive la variabilità di Z (condizionatamente a X).

Naturalmente, l'uso di informazione addizionale può consentire di restringere ulteriormente la (3) o la (4). Per esempio, Manski (1994) esamina il ruolo che può svolgere a questo riguardo una variabile strumentale, mentre Manski e Pepper (2000) esaminano il ruolo che può svolgere una variabile strumentale monotona.

Mi auguro che questa *laudatio* abbia contribuito a chiarire il ruolo fondamentale che il Professor Manski ha svolto, e continua a svolgere, nello sviluppo della teoria e della pratica econometrica. Per questa ragione, attribuirgli la Laurea Honoris Causa in Economia è per noi un grande onore.