

## FATTORI PRENATALI, AMBIENTE SOCIALE E DISEGUAGLIANZA NEGLI ESITI REDDITUALI UTILIZZANDO I DATI DI (PSEUDO) GEMELLI<sup>1</sup>

Daniele Checchi<sup>2</sup>

La letteratura sulla trasmissione intergenerazionale delle diseguaglianze si è tipicamente concentrata su misure di correlazione degli esiti tra la generazione dei genitori e quella dei figli (tipicamente livelli di istruzione e redditi, ma anche propensione alla fertilità, orientamento politico o religioso).

In Italia esistono relativamente pochi studi in questa direzione, anche a causa della scarsità di banche dati che riportino informazioni su due generazioni contigue. Alcuni studi utilizzano metodi indiretti, basati sul ricordo degli intervistati e/o sulla imputazione di un reddito alla generazione dei genitori basato sull'occupazione (Mocetti 2007, Piraino 2007, Barbieri et al. 2020). Tutti questi studi confermano che l'Italia si caratterizza per essere un paese a bassa mobilità intergenerazionale.

Più recentemente alcuni studiosi hanno cercato di fare uso di dati amministrativi per evitare gli errori di misura associati alla capacità di ricordo degli intervistati. Acciari et al. (2019) utilizzano dati di fonte fiscale per misurare la correlazione dei redditi di padri e figli, identificati sulla base della co-residenza precedente. Tuttavia questo approccio si scontra con la ridotta copertura temporale degli archivi amministrativi informatizzati nel nostro paese e con la difficoltà di identificazione dei legami familiari all'interno degli archivi stessi, producendo stime di mobilità più elevate delle precedenti. Bloise e Raitano (2020) adottano una combinazione di dati di survey (SILC) e di fonte amministrativa (INPS) riconfermando che su un arco temporale più esteso la correlazione intergenerazionale dei redditi (misurati come medie pluriennali) si avvicina a 0,50, valore molto elevato per l'insieme dei paesi di area OCSE.

Non mancano ovviamente strategie di ricerca più ingegnose, basate sull'associazione nei redditi o nella ricchezza riconducibile alla condivisione dei cognomi. Barone e Mocetti (2016) mostrano come il *ranking* sociale basato sui redditi si preservi a distanza di sei secoli. Güell et al (2018) producono misure di associazione a livello provinciale basate sul potere predittivo dei cognomi, mostrando come tali misure di mobilità intergenerazionale siano associate al livello di sviluppo economico locale.

<sup>1</sup> Progetto non ancora completato, in attesa di convenzione con Istituto Superiore di Sanità per accesso al registro dei gemelli. Collaborazione con Sonia Brescianini (ISS)

<sup>2</sup> Checchi Daniele (Università di Milano e DCSR INPS)

In questo lavoro si fa uso della condivisione del cognome delle persone come identificativo dell'ambiente familiare di provenienza, restringendosi alle coppie di potenziali gemelli. L'archivio di lavoro è costituito dagli individui che condividono cognome, data e comune di nascita. I gemelli sono quindi un sottoinsieme di questa popolazione, che è tuttavia diluita dalla ricorrenza tipica di alcuni cognomi in alcuni territori. Nonostante alcuni tentativi di minimizzare la portata del problema, i risultati che seguono si riferiscono più correttamente ad una popolazione di pseudo-gemelli estratti dagli archivi degli Estratti Conto gestiti da Inps.

Perché è interessante studiare i gemelli? Perché la letteratura sulla mobilità intergenerazionale è incapace di identificare i canali attraverso cui si produce persistenza intergenerazionale, che potrebbe essere il mero riflesso di una trasmissibilità genetica delle capacità, oppure l'esito sociale di meccanismi di esclusione e stratificazione basati sul censo. Esiste un'ampia letteratura che mostra come i redditi dei fratelli siano tra di loro correlati, e interpreta questa evidenza dell'importanza della famiglia e della comunità locale nel periodo della formazione, fattori che i fratelli condividono tra loro (*Black and Devereux, 2011*). Tuttavia i gemelli, a differenza dei fratelli, condividono il patrimonio genetico in misura controllabile, specialmente laddove si possa distinguere tra gemelli monozigoti e gemelli dizigoti. Mentre i primi condividono il 100% del loro patrimonio, i secondi ne condividono tipicamente la metà, e la differenza di correlazione nei redditi tra i due gruppi di gemelli può essere informativa della rilevanza della componente genetica in senso lato (inclusiva quindi delle componenti pre-natali condivise). La componente ambientale viene quindi identificata residualmente.

Björklund et al (2005) presentano una rassegna della letteratura al riguardo, mostrando che la correlazione dei redditi permanenti tra due gemelli varia tra 0,2 e 0,5, dove le componenti pre-natali tendono ad essere dominanti nel rendere conto di tale associazione. Per contro usando dati sulle adozioni (dove figli naturali e figli adottivi condividono l'ambiente ma non il patrimonio genetico), tali misurazioni tendono a rovesciarsi (*Björklund et al. 2006*). Tuttavia la distinzione tra fattori genetici e fattori ambientali è ovviamente puramente convenzionale, dal momento che, per esempio, i tratti caratteriali vengono trasmessi in misura analoga tra fratelli pur non essendo necessariamente correlati all'ambiente (*Cesarini et al. 2009*).

Il modello di base di questa letteratura è il seguente

$$y_i = a_{f(i)} + c_{f(i)} + e_i \quad (1)$$

dove  $y_i$  è un risultato di interesse (per esempio il reddito permanente) dell'individuo  $i$  che appartiene alla famiglia  $f$ ;  $a_{f(i)}$  è quella che per comodità chiamiamo componente genetica, che introduciamo per via (log)additiva;  $c_{f(i)}$  è la componente ambientale, condivisa da tutti i membri della famiglia  $f$ , anch'essa introdotta per via (log)additiva. Infine  $e_i$  è una componente idiosincratca individuale.

Per ottenere risultati interpretabili, occorre fare alcune assunzioni minimali:

- a) che non vi sia preselezione (*assortative mating*) nel formarsi della coppia genitoriale: in questo caso i gemelli dizigoti condividono solo 50% del loro patrimonio genetico;
- b) che i gemelli condividano l'ambiente indipendentemente dalla loro zigosità;
- c) che non vi siano interazioni tra effetti genetici ed effetti ambientali (cioè la covarianza tra i due sia nulla).

In questo caso, re-esprimendo la variabile di risultato  $y_i$  come somma di deviazione dalla media, essa sarà distribuita con media zero e varianze date  $\sigma_a^2$ ,  $\sigma_c^2$ ,  $\sigma_e^2$ . Quindi la varianza dell'*outcome* sarà pari a

$$\text{var}(y_i) = \sigma_a^2 + \sigma_c^2 + \sigma_e^2 \quad (2)$$

Distinguendo tra gemelli monozigoti, la covarianza degli esiti tra loro sarà pari a

$$\text{cov}(y_i, y_i)_{MZ} = \sigma_a^2 + \sigma_c^2 \quad (3)$$

mentre la covarianza degli esiti tra gemelli dizigoti sarà invece pari a

$$\text{cov}(y_i, y_i)_{DZ} = \frac{1}{2} \sigma_a^2 + \sigma_c^2 \quad (4)$$

Grazie alla piena condivisione del patrimonio pre-natale, i risultati di gemelli monozigoti saranno maggiormente correlate di quelli dei gemelli dizigoti, ovvero  $\text{cov}(y_i, y_i')_{MZ} > \text{cov}(y_i, y_i')_{DZ}$ . Possiamo quindi usare le espressioni precedenti per misurare l'effetto della componente pre-natale e della componente ambientale sulla disuguaglianza degli esiti nei risultati.

$$\begin{aligned} \sigma_a^2 &= 2 [\text{cov}(y_i, y_i')_{MZ} - \text{cov}(y_i, y_i')_{DZ}] \\ \sigma_c^2 &= \text{cov}(y_i, y_i')_{MZ} - \sigma_a^2 \end{aligned}$$

Passando all'analisi empirica sul caso italiano, in attesa di poter identificare con più precisione le coppie reali di gemelli, sono state estratte dall'archivio degli Estratti Conto tutte le n-tuple relative agli individui con identico cognome, diverso nome di nascita (per evitare i casi di omonimia) nati nello stesso giorno nello stesso comune. Per focalizzarci sull'età lavorativa ci restringiamo all'età 23-54 anni, corrispondenti agli individui nati tra il 1964 e il 1994, osservati lavorativamente su più anni fino al 2018.

In questo modo otteniamo dalla prima estrazione 133.829 n-tuple che riguardano 267.658 individui distinti. Dal momento che alcune di queste n-tuple sono costituite da più di due individui (potendo quindi trattarsi di parti trigemellari ma anche di presenza di omonimia), ci restringiamo alle coppie dove ciascun individuo compaia una (e una soltanto) sola volta. Lavoriamo quindi con 248.484 pseudo-gemelli, espressione di 124.242 coppie. Di questi individui conosciamo genere, età e luogo di nascita. La loro distribuzione per anno di nascita è abbastanza uniforme per i nati nel periodo 1964-81 (circa 10.000 individui per coorte di nascita), mentre si affievolisce successivamente per le coorti più giovani.<sup>3</sup>

Dal momento che non disponiamo di informazioni sulla zigosità, definiamo come dizigoti le coppie di pseudogemelli che presentano due generi diversi nella coppia. Questo porta chiaramente ad una sottostima della loro presenza, in quanto esistono coppie dizigote dello stesso genere, come si vede chiaramente nella tavola seguente. Dal momento che l'aspettativa teorica è quella di avere un terzo di monozigoti in una popolazione di gemelli, è palese che stiamo osservando imperfettamente la zigosità nella nostra popolazione.

Tavola I

Popolazione di riferimento		
	n. individui	%
Dizigoti	89.424	35,99%
Monozigoti	159.060	64,01%
<b>Totali</b>	<b>248.484</b>	<b>100,00%</b>

<sup>3</sup> Si ricordi che per entrare nel database analizzato occorre che entrambi i pseudogemelli abbiano svolto qualche attività lavorativa formale nel periodo compreso tra 36 anni (nati nel 1982) e 24 anni (nati nel 1994)

Dopo un ulteriore round di pulizia dei dati<sup>4</sup>, gli individui rimanenti (246.474 individui in 123.237 coppie pseudogemellari) sono stati associati a 4.679.541 episodi lavorativi dell'archivio Estratti Conto di Inps, che includono anche eventi collaterali quali congedi parentali, disoccupazione o malattia. Noi definiamo come *reddito permanente* la media aritmetica dei redditi ottenuti sugli anni lavorati, dopo aver azzerato i valori negativi e aver convertito i singoli valori a prezzi costanti (€ 2010). In alternativa restringiamo alla media dei redditi percepiti negli anni lavorativi iniziali (30-45) come miglior proxy del reddito permanente (condizionatamente all'aver lavorato per almeno 10 anni).

In aggiunta a queste due misure di reddito (reddito medio in arco di vita e reddito medio nel periodo 30-45 anni) aggiungiamo anche l'età del primo ingresso nel mercato del lavoro, come misura (imprecisa) della lunghezza degli studi intrapresi. Le correlazioni per vari gruppi di gemelli sono riportate nella tavola 2, mentre la classica decomposizione di cui prima è riportata nella tavola 3.

---

<sup>4</sup> Escludiamo coppie in cui uno solo degli pseudogemelli risulta nato all'estero (670 individui, individui con record lavorativi antecedenti l'anno di nascita (474 individui) o precedenti il compimento dei 10 anni di età (728 individui), individui precocemente ritirati dal mercato del lavoro prima del 2000 (18 individui). L'esclusione di un individuo comporta ovviamente l'uscita della coppia dal dataset (ulteriori 597 individui)

Tavola 2

Correlazione tra pseudogemelli				
Coefficienti di correlazione tra variabili standardizzate (media 0 e varianza 1 – standard errors ottenuti tramite bootstrapping con 50 replicazioni)	Coppie di diverso sesso (“dizigoti”)	Coppie dello stesso sesso (“monozigoti”)	Coppie di soli uomini (“monozigoti”)	Coppie di sole donne (“monozigoti”)
Età al primo lavoro	0,276 (0,00)	0,508 (0,00)	0,484 (0,00)	0,517 (0,00)
Reddito da lavoro in arco di vita	0,161 (0,01)	0,268 (0,08)	0,235 (0,09)	0,394 (0,01)
Reddito da lavoro 30-45 anni	0,067 (0,00)	0,367 (0,01)	0,344 (0,01)	0,339 (0,01)

Tavola 3

Decomposizione tra fattori esplicativi			
	fattori pre-natali ( $\sigma_a^2$ )	fattori ambientali ( $\sigma_c^2$ )	componente individuale ( $\sigma_u^2$ )
Età al primo lavoro	0,464	0,044	0,492
Reddito da lavoro in arco di vita	0,214	0,054	0,732
Reddito da lavoro 30-45 anni	0,600	-0,233	0,633

Da questi risultati emergerebbe che i fattori pre-natali esercitano un condizionamento ben più rilevante dei fattori ambientali nel determinare la carriera scolastica e successivamente reddituale degli pseudogemelli qui analizzati.

Vale infine la pena di interrogarci sulla validità esterna dei risultati ottenuti. La discussione delle potenziali distorsioni indotte dalla scelta campionaria e dalle regole di selezione rispetto ad un'analisi correttamente condotta sull'intera popolazione di veri gemelli è riassunta nella Tavola 4. Complessivamente si vede che le scelte (necessariamente) adottate ci portano ad una sottostima della covarianza dei redditi tra coppie di gemelli reali, portando quindi a sovrastimare la componente individuale.

Tavola 4

Discussione della validità esterna dell'analisi			
Regole di selezione	Impatto	Correlazione tra gemelli monozigoti reali	Correlazione tra gemelli dizigoti reali
Utilizzo degli archivi INPS	esclude coloro che non hanno mai avuto attività lavorativa formalizzata (casalinghe, disoccupati senza sussidi, disabili)	↓	↓
Restrizione ad individui che compaiono una sola volta	esclude n-tuple plurigemellari ma anche omonimie	↕	↕
Uso dello stesso sesso come proxy della monozigosità	ignora i gemelli dizigoti dello stesso sesso	↓	↓
Nello stesso campione utilizzare finestre di osservazione di lunghezza variabile	gli individui più vecchi presentano una stima più precisa del loro potenziale di reddito rispetto ai più giovani	↓	↓

## RIFERIMENTI

Acciari, Paolo, Alberto Polo and Giovanni Violante. 2019. “*And Yet, It Moves: Intergenerational Mobility in Italy*”. IZA Discussion Papers, No. 12273.

Barbieri, Teresa, Francesco Bloise and Michele Raitano. 2020. “*Intergenerational Earnings Inequality: New Evidence From Italy*”. *Review of Income and Wealth* 66(2), pp.418-433.

Barone, Guglielmo and Sauro Mocetti. 2016. “*Intergenerational mobility in the very long run: Florence 1427-2011*”. Temi di discussione (Economic working papers) 1060, Bank of Italy.

Björklund, A., Markus Jäntti and Gary Solon. 2005. “*Influences of Nature and Nurture on Earnings Variation: A Report on a Study of Various Sibling Types*”. In: Bowles, S., Gintis, H., Groves, M.O. (Eds.), “*Unequal Chances: Family Background and Economic Success*”. Princeton University Press, Princeton, pp. 145-164.

- Björklund, Ander, M. Lindahl and Eric Plug (2006). “*The Origins of Intergenerational Associations: Lessons from Swedish Adoption Data*”. *The Quarterly Journal of Economics*, 121(3), 999-1028.
- Black, Sandra and Paul Devereux. J. 2011. “*Recent developments in intergenerational mobility*”, volume 4B of *Handbook of Labor Economics*. Elsevier, Amsterdam pp. 1487-1541.
- Bloise, Francesco and Michele Raitano. 2020. “*Intergenerational Earnings Persistence in Italy between Actual Father–Son Pairs Accounting for Lifecycle and Attenuation Bias*”. *Oxford Bulletin of Economics and Statistics*, 0305–9049.
- Cesarini, D., M. Johannesson, P. Lichtenstein and B. Wallace (2009). “*Heritability of overconfidence*”. *Journal of the European Economic Association*, 7(2-3), 617-627.
- Güell, Maia, Michele Pellizzari, Giovanni Pica and José V. Rodríguez Mora. 2018. “*Correlating Social Mobility and Economic Outcomes*”. *Economic Journal*, vol. 128(612), pages 353-403.
- Mocetti, Sauro. 2007. “*Intergenerational earnings mobility in Italy*”. *BE Journal of Economic Analysis & Policy*. Vol. 7(2), pp. 1–25.
- Piraino, Patrizio. 2007. “*Comparable estimates of intergenerational income mobility in Italy*”. *BE Journal of Economic Analysis & Policy*, Vol. 7(2), pp. 1–27.