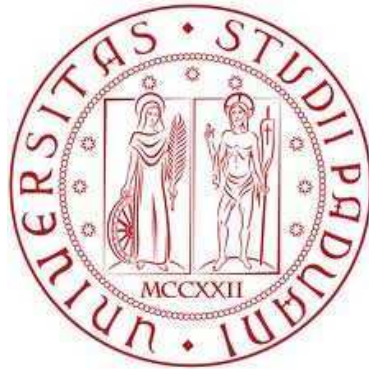


Università degli Studi di Padova
Corso di Laurea Magistrale in Scienze Statistiche



Effetti della data di nascita su istruzione e retribuzioni

Relatore: Prof. Adriano Paggiaro
Dipartimento di Scienze Statistiche

Laureanda:
Elena Ferraretto

Anno Accademico 2011/2012

Sommario

| | |
|--|----|
| 1. Introduzione | 5 |
| 1.1 Breve letteratura | 6 |
| 1.2 Analisi su dati italiani..... | 9 |
| 1.3 Approccio empirico | 10 |
| 1.4 I dati..... | 12 |
| 2. Regression Discontinuity Design | 15 |
| 2.1 Diploma di maturità..... | 16 |
| 2.2 Laurea triennale..... | 25 |
| 2.3 Titolo di studio elevato..... | 31 |
| 2.4 Retribuzione | 34 |
| 2.5 Verifica della validità del metodo per due coorti..... | 38 |
| 3. Analisi con più coorti | 43 |
| 3.1 Modello RDD con più anni..... | 45 |
| 3.1.1 Titolo di studio..... | 48 |
| 3.1.2 Retribuzione | 50 |
| 3.2 Modello completo | 51 |
| 3.2.1 Titolo di studio..... | 53 |
| 3.2.2 Retribuzione | 55 |
| 3.3 Verifica della validità del metodo per più coorti..... | 57 |
| 4. Conclusioni | 59 |
| Bibliografia..... | 63 |
| | |
| APPENDICE | 65 |
| A. Costruzione variabili di interesse per il titolo di studio..... | 65 |
| B. Costruzione variabili per l'anno scolastico | 67 |
| C. Costruzione variabili per famiglia | 68 |
| D. Analisi dati mancanti sulle retribuzioni | 69 |

1. Introduzione

In ogni paese le normative che disciplinano le iscrizioni al primo ciclo di istruzione per i bambini hanno come fine quello di ottimizzare l'apprendimento, e devono tenere conto di due aspetti fondamentali: quale sia l'età opportuna per iniziare la scuola, ad esempio se sia meglio a 5, 6 o 7 anni, e quale sia il più adatto criterio di suddivisione nei diversi anni scolastici, cioè la data entro la quale si dovranno compiere quel tot di anni per potersi iscrivere ad un preciso anno scolastico.

Queste decisioni fanno in modo che i bambini vengano 'selezionati' per la scuola in base alla loro data di nascita e ne consegue che nella stessa classe ci siano individui che hanno una differenza di quasi un anno di età, precisamente quelli nati subito dopo la data limite in relazione a quelli nati subito prima; e che invece coloro che sono nati attorno alla data prefissata, che hanno quasi la stessa età assoluta, inizino la scuola in due anni diversi.

I dubbi sorti a tal proposito hanno condotto a diversi studi, che rispondono fondamentalmente a tre domande diverse.

Relativamente alla decisione sull'età di inizio della formazione primaria ci si pone il dubbio sull'esistenza o meno del raggiungimento di una maturità tale per cui l'apprendimento scolastico venga massimizzato.

Per ciò che riguarda la decisione rispetto alla data limite ci si chiede invece se coloro che sono più vecchi all'interno di una classe ottengano risultati diversi da quelli più giovani, e se questa eventuale diversità sia dovuta al fatto che i primi sono più grandi in assoluto oppure sia conseguenza delle differenze causate dall'età confrontata a quella dei compagni.

In ogni caso, se gli individui nati subito dopo la data limite fossero migliori dei loro compagni, ci si chiede se sarebbe meglio per i più giovani aspettare un anno per

iscriversi a scuola in modo tale da essere i più grandi all'interno della classe o eventualmente più maturi in assoluto.

Se la data di nascita avesse un effetto sulle performance solo per i primi anni di scuola, per poi annullarsi col passare degli anni, non ci sarebbero ragioni per rivedere le leggi in merito.

Ma se, al contrario, i risultati di diversi outcome di interesse continuassero ad essere, in media, più alti per gli individui più grandi, allora potrebbero influenzare la carriera scolastica futura e quindi quella lavorativa di ciascuno.

Ci sono infatti diverse ragioni per cui questa disparità potrebbe riflettersi nel tempo: primo perché i bambini più avvantaggiati potrebbero progredire ad un ritmo più veloce, propagando così le differenze iniziali; secondo perché l'essere i più maturi potrebbe avere un'influenza sulla socialità, che porterebbe ad esempio ad aumentare l'autostima.

Inoltre, anche i risultati ottenuti nei primi anni dell'apprendimento potrebbero avere un effetto diretto sul percorso da effettuare: è molto probabile, infatti, che gli studenti con migliori 'performance' vengano indirizzati in una scuola superiore piuttosto che in un'altra, e che questo condizioni la futura iscrizione all'università e, in ultimo, il lavoro.

1.1 Breve letteratura

I molti studi empirici effettuati sull'argomento analizzano svariati aspetti del problema, e portano a conclusioni diverse, talvolta opposte.

Angrist e Krueger (1991) sono tra i primi ad accorgersi che le leggi per la scuola dell'obbligo possono creare delle disparità tra individui nati in diversi periodi dell'anno, dimostrando come negli Stati Uniti coloro che sono più piccoli all'interno di una classe risultino avvantaggiati perché ricevono una maggiore 'educazione'. Questo perché, oltre al fatto che i bambini incominciano il percorso

formativo con età differenti, le leggi impongono agli studenti di rimanere a scuola fino al compimento del loro sedicesimo o diciassettesimo anno.

Al contrario i dati norvegesi (Strom, 2004) evidenziano punteggi a scuola significativamente più alti per bambini più grandi all'interno della stessa coorte.

Allo stesso modo Bedard e Dhuey (2006), attraverso uno studio condotto su diversi paesi dell'OECD, dimostrano come, a causa della differenza di età all'interno delle classi, in media i bambini più giovani di ogni coorte ottengano, rispetto ai loro compagni, punteggi inferiori nei test svolti.

Anche in Svezia sembra che coloro che iniziano la scuola un anno più vecchi imparino di più (Fredriksson e Ockert, 2006) ma pare che ciò sia dovuto più ad una maturità assoluta che ad una relativa.

La particolarità del sistema tedesco, che indirizza gli studenti nella scelta di scuola secondaria, crea delle differenze nella probabilità di iscriversi ad una scuola di più alto profilo, che risulta maggiore per i più grandi della classe (Jurges e Schneider, 2007). Anche se, grazie alla flessibilità del sistema che prevede la revisione della scelta, le differenze scompaiono nel tempo (Puhani e Weber, 2007).

Elder e Lubotsky (2007) hanno dimostrato che l'entrare a scuola un anno più vecchi è relazionato positivamente ai risultati scolastici durante i primi anni di carriera, per poi decrescere notevolmente. Inoltre sostengono che l'età media dei compagni di classe influenza positivamente l'apprendimento.

Black, Devereux, Salvanes (2008) sottolineano però la difficoltà nel distinguere se gli studenti che iniziano quando sono più grandi riescano meglio per il solo fatto di essere più grandi quando fanno i test e quindi avvantaggiati oppure perché esistono benefici derivanti dall'iniziare la scuola quando si è più maturi. Analizzando i dati norvegesi e tenendo conto dell'età dell'individuo al momento del test, osservano pochi effetti dell'età in cui si inizia la scuola, ma dicono in ogni

caso di non riuscire ad individuare se le eventuali conseguenze dell'iniziare un anno prima o dopo siano causati dall'età assoluta o relativa.

Anche Crawford, Dearden, Meghir (2007) trovano uno significativo svantaggio nei punteggi dei test per gli studenti inglesi nati prima della data limite, e attribuiscono questo effetto quasi interamente alla differenza di età quando vengono effettuati i test, piuttosto che all'età relativa.

Altri studi concludono, invece, per la non significatività dell'effetto dell'età di inizio della scuola, dicendo che non ci sarebbe nessuna convenienza, per gli individui più piccoli, di posticipare la data di un anno (Fertig e Kluge, 2005).

Nemmeno Cascio e Schanzenbach (2007) che, nei dati del Tennessee, analizzano l'effetto dell'età relativa sull'apprendimento da parte degli studenti, trovano effetti significativi.

Sono invece meno coloro che si concentrano sui risultati post-scuola: Angrist e Krueger (1991) dimostrano che gli individui più piccoli, poiché ricevono negli Stati Uniti più educazione, abbiano una retribuzione più alta.

Fredriksson e Ockert (2006), argomentando sulla scelta di entrare a scuola un anno dopo, trovano due effetti contrastanti: 'da un lato iniziare più tardi aumenta l'apprendimento, dall'altro comporta un anno di esperienza lavorativa in meno', con un effetto negativo se si considera la retribuzione nell'arco della vita.

Per Dobkin e Ferreira (2007) essere il più giovane della classe non ha nessun effetto significativo sui risultati che riguardano la carriera lavorativa.

Dal più recente studio sui dati norvegesi (Solli, 2011) si osserva però che coloro che sono più piccoli hanno in media stipendi più bassi rispetto ai più grandi della stessa coorte, questo vale specialmente per gli uomini e per i ragazzini meno avvantaggiati.

1.2 Analisi su dati italiani

In Italia il Ministero dell'Istruzione stabilisce che 'i genitori debbano iscrivere alla classe prima della scuola primaria i bambini che compiono sei anni di età entro il 31 dicembre dell'anno di riferimento'.

Vista la distribuzione delle nascite in tutto l'arco di un anno, viene a crearsi una disparità di età tra gli individui all'interno della stessa classe, nel senso che coloro che sono nati a gennaio hanno 11 mesi in più di quelli nati a dicembre. Questa differenza tra bambini della stessa coorte viene chiamata in letteratura 'relative age'. Facendo riferimento alla data limite (nel nostro caso il 31 dicembre), quelli nati subito dopo sono i più grandi (nati nel mese di gennaio) e quelli nati subito prima (nati nel mese di dicembre) sono i più piccoli, a parità di coorte di nascita.

Avendo i bambini di gennaio quasi un anno in più rispetto ai loro compagni, differenza notevole se rapportata all'età totale che è di 5 o 6 anni, è ragionevole supporre, come infatti dimostrano la maggior parte degli studi, che i loro risultati scolastici siano migliori, soprattutto se misurati durante i primi anni dell'apprendimento.

Vogliamo presentare in questo elaborato un semplice studio su dati italiani, poiché non sembrerebbero esserci precedenti approfondimenti a riguardo.

Condizionati alle caratteristiche dei dati a noi disponibili, abbiamo deciso di analizzare gli effetti della data di nascita sui risultati di medio - lungo periodo (carriera scolastica e lavorativa). Andremo a confrontare inizialmente due coorti consecutive per poi allargarci a una popolazione più grande.

Ci concentreremo essenzialmente su una domanda: a parità di età vissuta, è meglio nascere a dicembre o a gennaio? Detto in altri termini, viste le leggi che regolano l'iscrizione alla formazione primaria, se i futuri genitori potessero paradossalmente decidere se far nascere un bambino a dicembre dell'anno X e

quindi mandarlo a scuola come 'più piccolo' nell'anno Y oppure a gennaio dell'anno X+1 e quindi mandarlo a scuola come 'più grande' dell'anno Y+1, cosa gli converrebbe scegliere?

1.3 Approccio empirico

Nel prossimo capitolo prenderemo in analisi due coorti consecutive e confronteremo i risultati raggiunti nel medio-lungo termine per i diversi mesi di nascita (24 totali).

Più che la differenza fra i più giovani e i più vecchi dello stesso anno, ci interessano quelle fra gli individui che stanno attorno alla data limite e che quindi, pur avendo praticamente la stessa età, appartengono a coorti diverse.

Dal momento che gli individui nati il 31 dicembre e il 1° gennaio sono equivalenti, la differenza tra le loro medie, per la variabile in esame, individua l'effetto della politica attuata, nel nostro caso l'effetto di fissare come data limite il 31 dicembre piuttosto che un altro giorno.

In ambito delle scienze sociali infatti, ragionando secondo una logica 'controfattuale', la valutazione dell'effetto di un intervento si identifica come differenza tra quanto si osserva in presenza di esso e quanto si sarebbe osservato in sua assenza.

Poiché il disegno della politica definisce una popolazione obiettivo, si creano delle discontinuità che determinano così quali siano le unità trattate e quali quelle non trattate (Trivellato, 2009). Ci si chiede a questo punto che cosa sarebbe successo ai trattati se non fossero stati esposti al trattamento.

Se definiamo il trattamento tramite:

$$D = \begin{cases} 1 & \text{trattati} \\ 0 & \text{non trattati} \end{cases}$$

e Y come l'outcome di interesse per gli individui trattati e non trattati

$$\begin{aligned} Y &= Y_0 \text{ se } D = 0 \\ Y &= Y_1 \text{ se } D = 1 \end{aligned} \quad ,$$

l'effetto dell'intervento sarebbe appunto la differenza, per la popolazione trattata, tra l'outcome misurato nel caso di 'trattamento' e l'outcome che si sarebbe misurato nel caso di 'non trattamento'.

-

$$E(Y_1 - Y_0 | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 1)$$

Tra questi due termini, il primo può essere osservato, mentre il secondo è ipotetico e non osservabile per definizione.

Nel nostro caso possiamo intendere come trattamento il fatto di essere nati dopo ('trattati') o prima ('non trattati') del 31 dicembre, e quindi andare a scuola un anno dopo o un anno prima (rispettivamente) dipende dalla data di nascita x .

Attorno alla soglia ($x = 31 \text{ dicembre}$) si assume verificata l'equivalenza tra trattati e non trattati rispetto alle caratteristiche che possono influire sia sull'assegnazione del trattamento sia sull'outcome. Si può utilizzare, al posto del secondo termine, l'outcome osservato per la popolazione non trattata, ovvero:

$$\begin{aligned} E(Y_1 - Y_0 | D = 1) &= E(Y_1 | D = 1) - E(Y_0 | D = 1) \\ &= E(Y_1 | D = 1) - E(Y_0 | D = 0) \end{aligned}$$

Gli effetti verranno identificati come differenza nelle medie della variabile risultato per i trattati che stanno immediatamente al di sopra della soglia e quella dei non trattati che stanno immediatamente al di sotto di essa. Questa strategia di valutazione è nota come Regression Discontinuity Design (RDD) (Trivellato, 2009).

La principale difficoltà a cui si va incontro volendo affrontare questo tipo di analisi è l'individuazione di gruppi effettivamente confrontabili, ovvero che si differenzino per il solo fatto di fare o meno parte del trattamento, condizione fondamentale per la corretta identificazione dell'effetto di interesse.

Una volta individuati, si tratta solo di confrontare i valori medi delle variabili risposta nei diversi gruppi, e lo faremo utilizzando grafici e semplici modelli OLS, che illustreremo di volta in volta.

Questi ultimi ci permettono infatti di identificare facilmente, *ceteris paribus*, le differenze medie negli outcome dovute al periodo di nascita e valutarne la significatività.

Sempre con la stessa logica, proporremo poi dei metodi che permettano di estendere l'analisi a più coorti contemporaneamente (Cap. 3).

Con lo scopo di verificare la validità delle ipotesi, presenteremo, alla fine di ogni capitolo, brevi analisi per controllare che non vi siano variabili che possono influenzare la data di nascita di un individuo, ma che essa sia casualmente distribuita.

1.4 I dati

Il dataset di cui disponiamo viene elaborato dall'Istat attraverso la 'Rilevazione sulle Forze di Lavoro', che rappresenta la principale fonte di informazione statistica sul mercato del lavoro italiano. Oltre ai dati riguardanti la condizione occupazionale, il tipo di lavoro, le esperienze, e la retribuzione, contiene anche numerose variabili individuali e familiari (sesso, luogo di nascita, relazioni di parentela, titolo di studio, etc).

I dati vengono forniti trimestralmente ma le interviste si svolgono in maniera continuativa sulle 13 settimane del trimestre. Per ogni trimestre si raccolgono

informazioni su circa 70.000 famiglie in 1.246 comuni italiani, per un totale di 175.000 individui (pari all'1,2% della popolazione complessiva italiana) (istat.it). Per questo studio ci serviamo dei dati trimestrali relativi agli anni 2005-10, per un totale di 24 rilevazioni.

Gli outcome che analizzeremo, compatibilmente coi nostri dati, sono principalmente quattro: il tempo necessario per il raggiungere l'eventuale diploma di maturità, il tempo necessario per il raggiungere l'eventuale laurea triennale, il raggiungimento o meno di un titolo di studio elevato e la retribuzione.

Riteniamo infatti interessante valutare se le eventuali differenze di risultati tra compagni di classe più grandi e più piccoli, più verosimili se relative ai primi anni di istruzione, si riflettano nei risultati futuri, così da condizionare anche la carriera lavorativa dell'individuo.

Vincolati al numero limitato di anni di rilevazione, utilizzeremo diversi gruppi di coorti, a seconda della variabile risultato che vogliamo analizzare:

- Coorte 1987-88 per l'analisi sul raggiungimento del diploma di maturità;
- Coorte 1984-85 per l'analisi sul raggiungimento della laurea triennale;
- Coorte 1972-73 per l'analisi sul raggiungimento di un titolo di studio elevato e per la retribuzione;
- Coorte 2003-04 per l'analisi sulle caratteristiche della famiglia.

Nell'estensione a più coorti considereremo gli anni di nascita 1968-'74 per i risultati di medio - lungo periodo, e 2000-'04 per quelli sulle variabili familiari.

2. Regression Discontinuity Design

Come già annunciato, l'ottica con cui andiamo ad analizzare i nostri dati è quella di Regression Discontinuity. Ci interessa pertanto il confronto, per le nostre variabili risposta, tra i soggetti nati alla fine di un anno e all'inizio di quello dopo, essendo essi uguali in tutto tranne per il fatto che, come conseguenza delle normative vigenti, i primi hanno iniziato la scuola un anno prima dei secondi.

La particolarità di questo approccio sta nel fatto che 'non ambisce ad individuare tutte le cause della variabile risultato ma a stimare l'effetto di una (sola) causa' (Trivellato, 2009).

Per ogni analisi di questo capitolo suddivideremo e confronteremo i risultati di interesse condizionatamente al mese (o trimestre) di nascita per due anni consecutivi, e valuteremo la significatività delle differenze nei mesi sia all'interno dello stesso anno sia tra anni diversi, per verificare se effettivamente a scuola si vengano a creare delle disparità con conseguenze di lungo termine.

Il modello che stimiamo è di regressione lineare semplice con il metodo dei minimi quadrati (OLS). Ad esempio per l'analisi sui trimestri si ottiene:

$$Y_i = \beta_0 + \beta_1 * td1_i + \beta_2 * td2_i + \beta_3 * td3_i + \beta_4 * td4_i + \beta_6 * td6_i + \beta_7 * td7_i + \beta_8 * td8_i + \varepsilon_i \quad (1)$$

dove Y_i è la variabile risposta (indicatrice di possesso diploma di maturità, indicatrice di possesso di laurea triennale, indicatrice di possesso di titolo di studio elevato, retribuzione);
 $td1_i, td2_i, td3_i, td4_i, td6_i, td7_i, td8_i$ sono le dummy per i trimestri di nascita, esclusa quella per il 5°, usata come baseline ;
 ε_i è l'errore.

I coefficienti β_i misurano gli effetti differenziali tra l'essere nati nel trimestre i -esimo in relazione a quello omissivo, in questo caso il 5°.

Questa scelta è stata fatta per rendere facilmente confrontabili il valore medio per il 4° trimestre di un anno con quello del 1° dell'anno successivo. In questo modo infatti il test e il p value presentati negli output in corrispondenza del coefficiente β_4 ci indicheranno se l'ipotesi di uguaglianza tra le medie (tra quest'ultimo e il trimestre successivo) viene accettata o rifiutata.

Precisiamo inoltre che utilizziamo un modello lineare anche con variabili dicotomiche perché ha il vantaggio di essere di immediata interpretazione. Analisi preliminari con specificazioni alternative portano comunque alle stesse considerazioni.

2.1 Diploma di maturità

In questo paragrafo vogliamo valutare se vi sia o meno una differenza tra nati nei diversi mesi dell'anno per velocità di raggiungimento del diploma di maturità. Infatti, nel caso in cui gli individui più grandi si diplomassero prima di quelli più piccoli, o nel caso in cui ci fossero anomalie per gli individui attorno alla soglia (1° gennaio), significherebbe che vi sono degli effetti legati all'età.

Vogliamo seguire le due coorti di individui negli anni in cui iniziano a diplomarsi e, poiché possediamo un numero limitato di rilevazioni, abbiamo selezionato i nati nel 1987 e nel 1988.

La variabile risposta in questo caso è una dicotomica del tipo:

$$Y_i = \begin{cases} 1, & \text{se titolo è pari o superiore al diploma} \\ 0, & \text{altrimenti} \end{cases}$$

creata a partire dalla variabile Istat ' titolo di studio ' (vedi Appendice A).

Per interpretare più facilmente i dati creiamo inoltre una variabile 'anno scolastico' che comprende le rilevazioni effettuate entro dei limiti precisi, tali per cui risulterà $Y_i = 1$ se l'individuo si è diplomato nell'a.s. di riferimento, 0 altrimenti (vedi Appendice B).

Come primo esempio la Figura 1 presenta l'analisi condotta a livello 'mese di nascita', indicato con i valori da 1 a 12 per gennaio-dicembre 1987, e da 13 a 24 per gennaio-dicembre 1988.

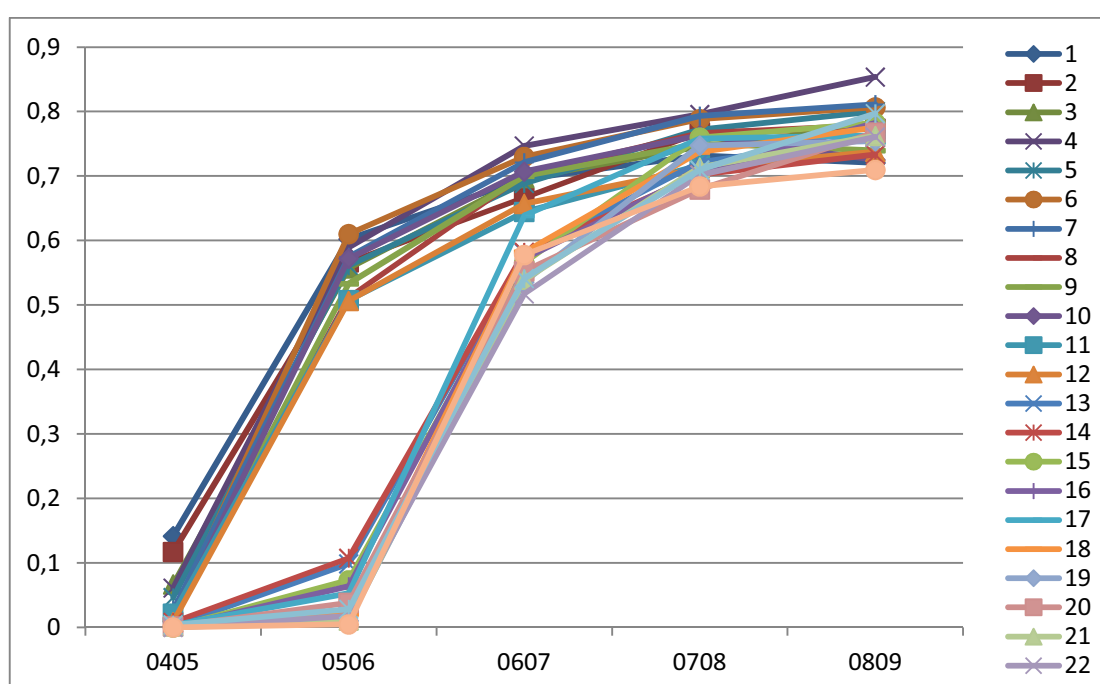


Figura 1. Percentuale diplomati per anno scolastico e mese di nascita, coorti 1987-88.

Poiché la variabile in esame è dicotomica 0- 1, i valori sul grafico sono da leggersi come percentuali dei diplomati in quell'anno scolastico sul totale degli intervistati in quel periodo.

Nonostante sia difficile distinguere precisamente le diverse curve, è immediato vedere come vi sia un'evidente separazione tra le diverse coorti di nascita e che il loro andamento nel tempo sia simile ma traslato di un anno: vi è una netta differenza tra i ragazzi del 12esimo(dicembre 1987) e 13esimo(gennaio 1988)

mese, nel senso che, nonostante una quasi parità di età, i nati in dicembre si diplomano in media un anno prima, come prevedibile.

Nell'osservare la differenza invece tra i nati nei primi e negli ultimi mesi dello stesso anno si deve tenere conto del fatto che circa il 10% dei più grandi all'interno della coorte hanno fatto la primina e hanno ottenuto il titolo prima del previsto (i diplomati nel 2004-05, nati nel 1987 o i diplomati nel 2005-06, nati nel 1988). Nell'analisi successiva puntualizzeremo questa considerazione.

Inoltre, alla fine delle rilevazioni (anno 2008-'09) non sembrerebbero esserci differenze sostanziali tra i 24 mesi.

La bassa numerosità per ogni gruppo considerato (Tabella 1) ci induce però a proseguire l'analisi considerando i trimestri di nascita invece dei mesi.

Tabella 1. Numerosità per mese di nascita e anno scolastico, coorti 1987-88.

| | | Trimestri di nascita | | | | | | | | | | | | Tot. |
|-----------------|------|----------------------|------|------|------|------|------|------|------|------|------|------|------|-------|
| | | 1987 | | | | | | | | | | | | |
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | |
| anno scolastico | 0405 | 459 | 427 | 478 | 474 | 539 | 541 | 638 | 496 | 489 | 521 | 469 | 488 | 6019 |
| | 0506 | 420 | 362 | 474 | 443 | 460 | 518 | 555 | 530 | 497 | 451 | 398 | 470 | 5578 |
| | 0607 | 410 | 417 | 429 | 407 | 456 | 470 | 488 | 534 | 520 | 406 | 360 | 414 | 5311 |
| | 0708 | 403 | 409 | 429 | 416 | 461 | 482 | 479 | 518 | 524 | 401 | 381 | 431 | 5334 |
| | 0809 | 365 | 349 | 364 | 355 | 426 | 424 | 466 | 430 | 457 | 393 | 365 | 423 | 4817 |
| | Tot. | 2057 | 1964 | 2174 | 2095 | 2342 | 2435 | 2626 | 2508 | 2487 | 2172 | 1973 | 2226 | 27059 |

| | | 1988 | | | | | | | | | | | | Tot. |
|------|------|-----------------|------|------|------|------|------|------|------|------|------|------|------|-------|
| | | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | |
| | | anno scolastico | 0405 | 548 | 498 | 497 | 505 | 555 | 541 | 522 | 581 | 599 | 495 | |
| 0506 | 517 | | 456 | 465 | 502 | 529 | 481 | 576 | 575 | 549 | 502 | 459 | 419 | 6030 |
| 0607 | 508 | | 402 | 481 | 390 | 518 | 457 | 506 | 503 | 519 | 464 | 438 | 457 | 5643 |
| 0708 | 395 | | 393 | 483 | 377 | 471 | 434 | 487 | 474 | 510 | 446 | 451 | 440 | 5361 |
| 0809 | 354 | | 345 | 432 | 404 | 422 | 410 | 403 | 426 | 526 | 393 | 384 | 334 | 4833 |
| Tot. | 2322 | | 2094 | 2358 | 2178 | 2495 | 2323 | 2494 | 2559 | 2703 | 2300 | 2211 | 2116 | 28153 |

La Figura 2 presenta l'analisi per trimestri (da 1 a 4 per la coorte del 1987, da 5 a 8 per quella del 1988), più esaustiva grazie all'aumento della numerosità per gruppo considerato.

Tabella 2. Numerosità per trimestre di nascita e anno scolastico, coorti 1987-88.

| | | 1987 | | | | 1988 | | | | Tot. |
|-----------------|------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| anno scolastico | 0405 | 1.364 | 1.554 | 1.623 | 1.478 | 1.543 | 1.601 | 1.702 | 1.440 | 12.305 |
| | 0506 | 1.256 | 1.421 | 1.582 | 1.319 | 1.438 | 1.512 | 1.700 | 1.380 | 11.608 |
| | 0607 | 1.256 | 1.333 | 1.542 | 1.180 | 1.391 | 1.365 | 1.528 | 1.359 | 10.954 |
| | 0708 | 1.241 | 1.359 | 1.521 | 1.213 | 1.271 | 1.282 | 1.471 | 1.337 | 10.695 |
| | 0809 | 1.078 | 1.205 | 1.353 | 1.181 | 1.131 | 1.236 | 1.355 | 1.111 | 9.650 |
| | Tot. | 6.195 | 6.872 | 7.621 | 6.371 | 6.774 | 6.996 | 7.756 | 6.627 | 55.212 |

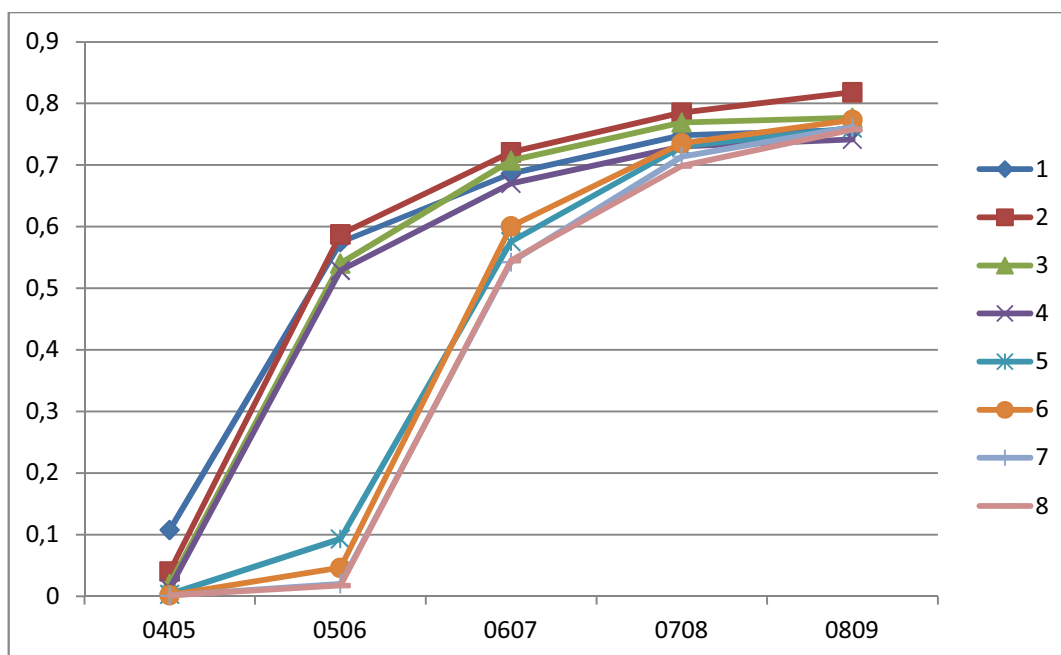


Figura 2. Percentuale diplomati per anno scolastico e trimestre di nascita, coorti 1987-88.

Il grafico relativo ai trimestri è molto più chiaro rispetto al precedente: soffermandosi ad un anno scolastico e confrontando i punti in colonna si possono osservare le differenze nelle percentuali dei diplomati tra i diversi 8 periodi di

nascita, osservando invece i punti allineati in riga si può vedere se il comportamento delle coorti, ovviamente traslato di un anno scolastico, è uguale. Possiamo notare come la differenza tra ultimo trimestre del 1987 e il primo del 1988 sia netta, mentre invece non vi siano grandi differenze tra i trimestri all'interno degli stessi anni di nascita.

Parrebbe opportuno dire che gli individui più giovani di una coorte stanno comunque 'meglio' dei loro quasi coetanei, nel senso che si diplomano un anno prima. Per di più, non sembrano essere diversi dai loro compagni più grandi.

Come già accennato, è interessante vedere che una parte di coloro che sono nati nei primi mesi dell'anno si sono iscritti a scuola un anno prima, unico motivo che li potrebbe portare a diplomarsi in anticipo rispetto al previsto.

E' quindi evidente che attorno alla soglia non vi sia esattamente un salto 0 – 1 tra 'non trattati' e 'trattati' (vedi paragrafo 1.4), perché, mentre tutti coloro che sono nati prima del 31 dicembre vanno a scuola nell'anno stabilito, non tutti coloro che sono nati dopo lo fanno. Può essere cioè che gli individui di gennaio si iscrivano a scuola nello stesso anno di quelli di dicembre, negli anni 1987-88 sembrerebbero circa il 10%.

In questo caso, poiché il valore della discontinuità è minore di uno e l'assegnazione del trattamento non è una funzione deterministica del mese di nascita, si parlerebbe di Fuzzy Regression Discontinuity Design.

L'eventuale effetto che misuriamo analizzando la differenza tra medie non rappresenta esattamente il vero effetto (in questo caso rappresenta una sottostima) che si misurerebbe tenendo conto che una parte di nati dopo la soglia non si possono considerare 'trattati'. Si parla in questo caso di individui "non compliers", che scelgono un trattamento diverso da quello assegnato. Si veda fra gli altri Hahn, Todd e Van der Klaauw (2001) per una descrizione dei metodi per l'identificazione in questo contesto.

Tramite una semplice regressione (modello (1)), andiamo a stimare le differenze fra le medie. Le suddividiamo in anni scolastici, per vedere cosa cambia al passare del tempo (Regressione 1-5).

Nella 'Regressione 1' (a.s. 2004-05) il coefficiente per il primo trimestre, positivo e pari a 0.10, sta ad indicare la percentuali di alunni che, iscritti a scuola un anno prima del previsto, si sono diplomati in anticipo. Risultano significativi, anche se con valori meno rilevanti, anche il secondo e terzo trimestre. In particolare quest'ultimo non è giustificato dalle leggi italiane.

Nel secondo anno (Regressione 2) vi è invece una netta differenza tra le due coorti, nel senso che i diplomati nati nei quattro trimestri del 1987 sono, come era logico aspettarsi, circa il 50% in più dei nati nel 1988.

La differenza significativa tra il primo trimestre dell'88 e i compagni di coorte è invece di nuovo dovuta alla percentuale di studenti che si iscrivono e completano la scuola un anno prima ed ha un valore ancora attorno a 0.10.

Nel terzo anno, in cui si diplomano gli studenti dell'88 che sono 'in regola', i diplomati dell'87 ritornano ad essere solo il 10 % in più. Poiché, come appariva anche dal grafico, i comportamenti delle due coorti si possono definire equivalenti ma traslati di un anno, questo valore corrisponde ad una stima dei cosiddetti 'ripetenti'.

Nel quarto e nel quinto anno le differenze tendono ad annullarsi e diventano non significative, ad indicare che, salvo la presenza di un picco di apparente stagionalità nel secondo trimestre, la percentuale di individui che prima o dopo raggiungono il diploma è indipendente dalla coorte e dall'anno di nascita. Ritorneremo nel seguito su questo argomento.

Regressione 1. Possesso del diploma di maturità, A.S. 2004-05, coorti 1987-88

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 1° trimestre di nascita | 0,1045 | 0,0059 | 0,000 |
| 2° trimestre di nascita | 0,0373 | 0,0052 | 0,000 |
| 3° trimestre di nascita | 0,0171 | 0,0051 | 0,001 |
| 4° trimestre di nascita | 0,0096 | 0,0052 | 0,066 |
| 6° trimestre di nascita | -0,0014 | 0,0052 | 0,790 |
| 7° trimestre di nascita | -0,0021 | 0,0051 | 0,683 |
| 8° trimestre di nascita | -0,0019 | 0,0053 | 0,725 |
| costante | 0,0032 | 0,0037 | 0,376 |
| N° osservazioni | 12305 | | |

Regressione 2. Possesso del diploma di maturità, A.S. 2005-06, coorti 1987-88

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 1° trimestre di nascita | 0,4817 | 0,0144 | 0,000 |
| 2° trimestre di nascita | 0,4944 | 0,0140 | 0,000 |
| 3° trimestre di nascita | 0,4473 | 0,0136 | 0,000 |
| 4° trimestre di nascita | 0,4360 | 0,0142 | 0,000 |
| 6° trimestre di nascita | -0,0469 | 0,0138 | 0,001 |
| 7° trimestre di nascita | -0,0732 | 0,0134 | 0,000 |
| 8° trimestre di nascita | -0,0758 | 0,0141 | 0,000 |
| costante | 0,0932 | 0,0098 | 0,000 |
| N° osservazioni | 10954 | | |

Regressione 3. Possesso del diploma di maturità, A.S. 2006-07, coorti 1987-88

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 1° trimestre di nascita | 0,1105 | 0,0186 | 0,000 |
| 2° trimestre di nascita | 0,1451 | 0,0183 | 0,000 |
| 3° trimestre di nascita | 0,1317 | 0,0177 | 0,000 |
| 4° trimestre di nascita | 0,0945 | 0,0189 | 0,000 |
| 6° trimestre di nascita | 0,0249 | 0,0182 | 0,172 |
| 7° trimestre di nascita | -0,0333 | 0,0177 | 0,060 |
| 8° trimestre di nascita | -0,0306 | 0,0182 | 0,093 |
| costante | 0,5758 | 0,0128 | 0,000 |
| N° osservazioni | 10951 | | |

Regressione 4. Possesso del diploma di maturità, A.S. 2007-08, coorti 1987-88

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 1° trimestre di nascita | 0,0185 | 0,0175 | 0,291 |
| 2° trimestre di nascita | 0,0550 | 0,0171 | 0,001 |
| 3° trimestre di nascita | 0,0391 | 0,0167 | 0,019 |
| 4° trimestre di nascita | 0,0003 | 0,0176 | 0,987 |
| 6° trimestre di nascita | 0,0054 | 0,0174 | 0,754 |
| 7° trimestre di nascita | -0,0163 | 0,0168 | 0,331 |
| 8° trimestre di nascita | -0,0316 | 0,0172 | 0,066 |
| costante | 0,7301 | 0,0123 | 0,000 |
| N° osservazioni | 10695 | | |

Regressione 5. Possesso del diploma di maturità, A.S. 2008-09, coorti 1987-88

| | Coefficiente |
|-------------------------|-----------------------------------|
| | St. Error |
| | P value |
| 1° trimestre di nascita | -0,0034 <i>0,0179</i> 0,848 |
| 2° trimestre di nascita | 0,0579 <i>0,0174</i> 0,001 |
| 3° trimestre di nascita | 0,0164 <i>0,0170</i> 0,334 |
| 4° trimestre di nascita | -0,0186 <i>0,0173</i> 0,287 |
| 6° trimestre di nascita | 0,0131 <i>0,0173</i> 0,451 |
| 7° trimestre di nascita | 0,0012 <i>0,0170</i> 0,942 |
| 8° trimestre di nascita | -0,0025 <i>0,0178</i> 0,888 |
| costante | 0,7604 <i>0,0125</i> 0,000 |
| N° osservazioni | 9650 |

Concludiamo quindi che, quanto a velocità di raggiungimento del diploma, non sembrano esservi differenze tra trimestri all'interno delle coorti, soprattutto considerando il fatto che una parte dei nati nei primi mesi si è iscritta a scuola un anno prima.

Si osserva inoltre che, pur essendo i più piccoli della loro coorte, i nati nel 4° trimestre '87 si diplomano in media un anno prima dei loro 'quasi' coetanei del 1° trimestre '88. E quindi ipoteticamente entreranno un anno prima nel mercato del lavoro.

2.2 Laurea triennale

Procediamo allo stesso modo della precedente analisi, cambiando però variabile di interesse e coorti.

La variabile risposta è come al solito dicotomica, ed indica il raggiungimento di un titolo almeno pari alla laurea triennale (vedi Appendice A).

$$Y_i = \begin{cases} 1, & \text{se titolo è pari o superiore alla laurea triennale} \\ 0, & \text{altrimenti} \end{cases}$$

Le coorti osservate sono in questo caso 1984 e 1985, poiché si presuppone che conseguano il titolo negli anni di rilevazione disponibili.

La variabile per la suddivisione cambia rispetto a prima e raggruppa tutti gli intervistati in un certo anno scolastico (vedi Appendice B).

Di seguito, la Tabella 3 presenta le numerosità per i dati analizzati.

Tabella 3. Numerosità per trimestre di nascita e anno scolastico, coorti 1984-85.

| | | 1984 | | | | 1985 | | | | Tot. |
|-----------------|------|-------|-------|-------|-------|-------|-------|-------|-------|--------|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| anno scolastico | 0506 | 1.628 | 1.698 | 1.895 | 1.711 | 1.600 | 1.837 | 1.735 | 1.575 | 13.679 |
| | 0607 | 1.126 | 1.227 | 1.316 | 1.156 | 1.161 | 1.364 | 1.221 | 1.106 | 9.677 |
| | 0708 | 1.393 | 1.533 | 1.598 | 1.443 | 1.386 | 1.585 | 1.531 | 1.348 | 11.817 |
| | 0809 | 1.297 | 1.420 | 1.541 | 1.371 | 1.284 | 1.402 | 1.389 | 1.289 | 10.993 |
| | 0910 | 1.310 | 1.396 | 1.373 | 1.306 | 1.254 | 1.327 | 1.356 | 1.229 | 10.551 |
| | Tot. | 6.754 | 7.274 | 7.723 | 6.987 | 6.685 | 7.515 | 7.232 | 6.547 | 56.717 |

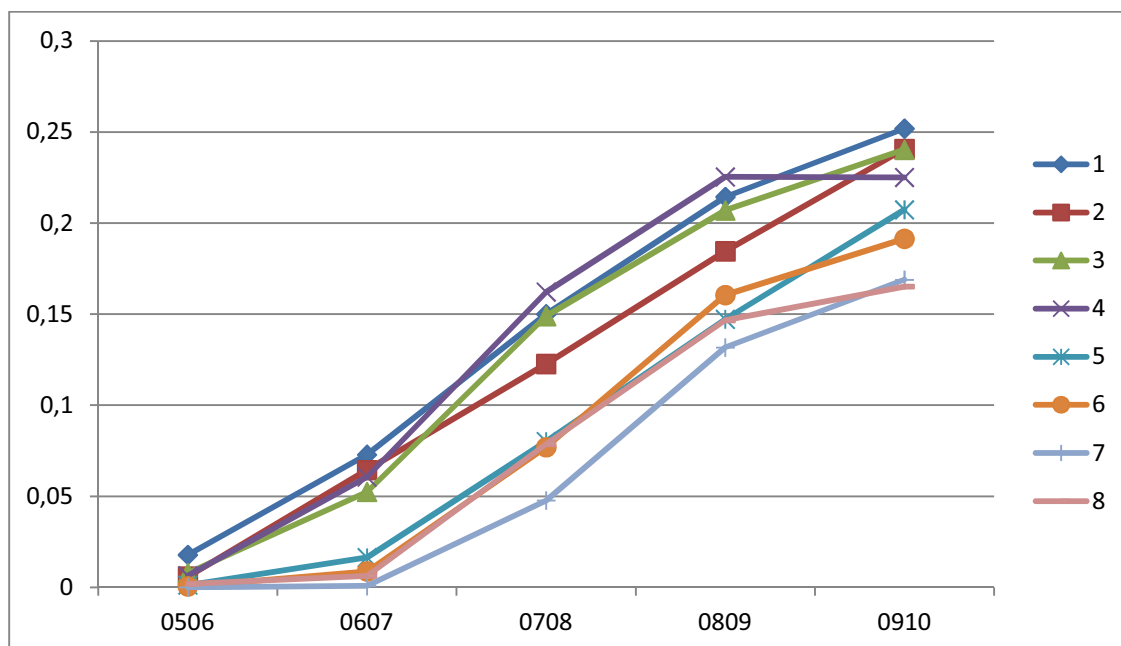


Figura 3. Percentuali laureati triennali (o laurea superiore) per anno scolastico e trimestre di nascita, coorti 1984-85.

Ancora una volta, tramite il grafico in Figura 3, possiamo confrontare quanti, tra i nati in un certo trimestre, si siano laureati nei diversi anni di nascita (confronto verticale) e osservare le differenze tra i trimestri delle due coorti (in orizzontale).

Si noti che, se ad esempio la percentuale dei laureati sul totale di questa popolazione arriverà a raggiungere circa il 20-25 %, un valore nel grafico pari al 10% corrisponde al 40-50% di coloro che prima o poi si laureeranno.

Osserviamo un andamento quasi parallelo delle due coorti di nascita in esame (1984-85), con un distacco, di circa un anno scolastico, un po' meno marcato rispetto all'analisi precedente perché si tratta di un percorso universitario e quindi di durata più variabile. Si vede comunque che i nati nei quattro trimestri dell'84 si laureano in media un anno prima dei nati nell'85, senza differenze sostanziali tra periodi di nascita all'interno della stessa coorte.

Procediamo presentando le regressioni (modello (1)) suddivise per anni scolastici. Si ripropone una situazione molto simile a quella del diploma, anche se gli effetti sono di minore entità.

Nell'anno 2005-06 si sono laureati solo alcuni nati nel primo trimestre dell'84 (circa 1.6%), facenti parte molto probabilmente dei 'primini' dell'analisi precedente, e qualche persona nata negli altri trimestri della stessa coorte.

Il secondo anno di osservazione (Regressione 2) corrisponde esattamente alla tempistica per la laurea di tre anni per i nati nell'84, le cui percentuali risultano significativamente più alte rispetto a tutti quelli della coorte successiva.

Il segno negativo dei coefficienti per il secondo, terzo e quarto trimestre dell'85 è da attribuirsi ancora una volta alla presenza di individui all'interno del trimestre di riferimento che si laureano in anticipo, come prima pari a 1.6% (vedi costante).

Le differenze tra le coorti persistono negli anni 2007-08 e 2008-09 (coefficienti positivi e significativi per i primi quattro trimestri e negativi e quasi sempre non significativi per gli ultimi tre). Questo indica chiaramente che, tra coloro che raggiungono il titolo di laurea, gli studenti appartenenti alla coorte 1984 si laureano, com'è ragionevole aspettarsi, prima nel tempo rispetto a quelli della coorte 1985 ma con la stessa velocità media, indipendentemente dal trimestre di nascita.

Se si osservano più lontani nel tempo (anno scolastico 2009-10) le differenze tendono a diminuire. Il divario presente, a parità di anno di nascita, tra primi e ultimo trimestre è causato dalla presenza di coloro che si iscrivono all'università un anno prima.

Regressione 6. Possesso della laurea triennale, A.S. 2005-06, coorti 1984-85

| | Coefficiente |
|-------------------------|-----------------------------------|
| | St. Error |
| | P value |
| 1° trimestre di nascita | 0,0166 <i>0,0025</i> 0,000 |
| 2° trimestre di nascita | 0,0046 <i>0,0025</i> 0,063 |
| 3° trimestre di nascita | 0,0067 <i>0,0024</i> 0,006 |
| 4° trimestre di nascita | 0,0052 <i>0,0025</i> 0,038 |
| 6° trimestre di nascita | -0,0007 <i>0,0025</i> 0,773 |
| 7° trimestre di nascita | -0,0013 <i>0,0025</i> 0,615 |
| 8° trimestre di nascita | 0,0007 <i>0,0025</i> 0,797 |
| costante | 0,0013 <i>0,0018</i> 0,485 |
| N° osservazioni | 13679 |

Regressione 7. Possesso della laurea triennale, A.S. 2006-07, coorti 1984-85

| | Coefficiente |
|-------------------------|-----------------------------------|
| | St. Error |
| | P value |
| 1° trimestre di nascita | 0,0565 <i>0,0076</i> 0,000 |
| 2° trimestre di nascita | 0,0480 <i>0,0074</i> 0,000 |
| 3° trimestre di nascita | 0,0361 <i>0,0073</i> 0,000 |
| 4° trimestre di nascita | 0,0442 <i>0,0076</i> 0,000 |
| 6° trimestre di nascita | -0,0076 <i>0,0073</i> 0,297 |
| 7° trimestre di nascita | -0,0156 <i>0,0075</i> 0,037 |
| 8° trimestre di nascita | -0,0100 <i>0,0076</i> 0,189 |
| costante | 0,0164 <i>0,0053</i> 0,002 |
| N° osservazioni | 9677 |

Regressione 8. Possesso della laurea triennale, A.S. 2007-08, coorti 1984-85

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 1° trimestre di nascita | 0,0699 | 0,0117 | 0,000 |
| 2° trimestre di nascita | 0,0425 | 0,0114 | 0,000 |
| 3° trimestre di nascita | 0,0689 | 0,0113 | 0,000 |
| 4° trimestre di nascita | 0,0821 | 0,0116 | 0,000 |
| 6° trimestre di nascita | -0,0031 | 0,0113 | 0,784 |
| 7° trimestre di nascita | -0,0324 | 0,0114 | 0,005 |
| 8° trimestre di nascita | -0,0015 | 0,0118 | 0,902 |
| costante | 0,0801 | 0,0083 | 0,000 |
| N° osservazioni | 11817 | | |

Regressione 9. Possesso della laurea triennale, A.S. 2008-09, coorti 1984-85

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 1° trimestre di nascita | 0,0671 | 0,0150 | 0,000 |
| 2° trimestre di nascita | 0,0373 | 0,0147 | 0,011 |
| 3° trimestre di nascita | 0,0598 | 0,0144 | 0,000 |
| 4° trimestre di nascita | 0,0782 | 0,0148 | 0,000 |
| 6° trimestre di nascita | 0,0133 | 0,0147 | 0,367 |
| 7° trimestre di nascita | -0,0155 | 0,0148 | 0,295 |
| 8° trimestre di nascita | -0,0006 | 0,0150 | 0,970 |
| costante | 0,1472 | 0,0106 | 0,000 |
| N° osservazioni | 10993 | | |

Regressione 10. Possesso della laurea triennale, A.S. 2009-10, coorti 1984-85

| | Coefficiente |
|-------------------------|-----------------------------------|
| | St. Error |
| | P value |
| 1° trimestre di nascita | 0,0446 <i>0,0160</i> 0,006 |
| 2° trimestre di nascita | 0,0334 <i>0,0159</i> 0,035 |
| 3° trimestre di nascita | 0,0330 <i>0,0159</i> 0,038 |
| 4° trimestre di nascita | 0,0178 <i>0,0161</i> 0,270 |
| 6° trimestre di nascita | -0,0159 <i>0,0161</i> 0,321 |
| 7° trimestre di nascita | -0,0385 <i>0,0160</i> 0,016 |
| 8° trimestre di nascita | -0,0422 <i>0,0164</i> 0,010 |
| costante | 0,2073 <i>0,0115</i> 0,000 |
| N° osservazioni | 10551 |

Quello che emerge dalle due analisi fino a qui presentate è che: non vi sono differenze, nella velocità di raggiungimento del titolo, tra individui della stesso anno ma di diversi trimestri di nascita, non vi sono differenze tra i comportamenti delle due coorti nel tempo se non per il fatto che, come ci si aspettava, quelli della prima coorte si iscrivono e quindi ottengono il titolo in media un anno prima di quelli della coorte successiva.

A dispetto della credenza che i più piccoli della classe si avvicinino o ottengano risultati peggiori rispetto ai più grandi dell'anno successivo, possiamo dire che, in ogni caso, i nati subito prima della soglia si diplomano in media un anno prima.

2.3 Titolo di studio elevato

Nelle precedenti analisi avevamo la possibilità di seguire le due coorti di interesse nei diversi momenti di rilevazione, cosa che ci permetteva di valutare i cambiamenti di stato (la variabile di interesse passava da 0 a 1 per molti individui) col passare del tempo. Dato l'anno scolastico, potevamo quindi capire quanti fossero coloro che possedevano già il titolo e fare dei confronti sulla velocità di raggiungimento di esso.

In questo paragrafo ci proponiamo invece di osservare quanti siano coloro che, alla 'fine' della loro carriera scolastica, abbiano raggiunto un titolo elevato come laurea o dottorato, per misurare l'impatto di lungo termine.

L'effetto permanente delle leggi scolastiche deve essere osservato per delle coorti che si presuppone abbiano già completato il loro percorso di studi, pertanto la scelta è ricaduta sui nati nel 1972-73, che hanno fra i 32 e i 38 anni nei periodi di osservazione.

Come prima, ci interessa confrontare le medie dei nati nei primi e negli ultimi trimestri dell'anno, ed eventualmente la presenza di andamenti particolari all'interno dell'anno stesso.

La variabile in esame, indicatrice del possesso di un titolo elevato, è la seguente (vedi appendice A):

$$Y_i = \begin{cases} 1, & \text{se l'individuo ha un titolo pari o superiore alla laurea} \\ 0, & \text{altrimenti} \end{cases}$$

La Tabella 4 presenta la numerosità per trimestre di nascita.

Tabella 4. Numerosità per trimestre di nascita, coorti 1972-73.

| | | Trimestre di nascita | Frequenze |
|------|--|----------------------|-----------|
| 1972 | | 1 | 11.701 |
| | | 2 | 12.310 |
| | | 3 | 12.552 |
| | | 4 | 11.718 |
| 1973 | | 5 | 11.654 |
| | | 6 | 12.072 |
| | | 7 | 12.410 |
| | | 8 | 10.732 |
| Tot. | | | 95.149 |

Il grafico in Figura 4 ci indica, per ognuno degli otto trimestri (vedi ascissa) la percentuale, tra gli intervistati, di coloro che possiedono almeno la laurea.

La suddivisione, a differenza dai casi precedenti, non viene più fatta anche per periodo di rilevazione, in quanto si ritiene ragionevole pensare che, per la coorte di interesse, la percentuale dei laureati non cambi se viene misurata quando questi hanno 31 o 32 anni (rilevazioni 2004) o quando ne hanno 37 o 38 (rilevazioni 2010).

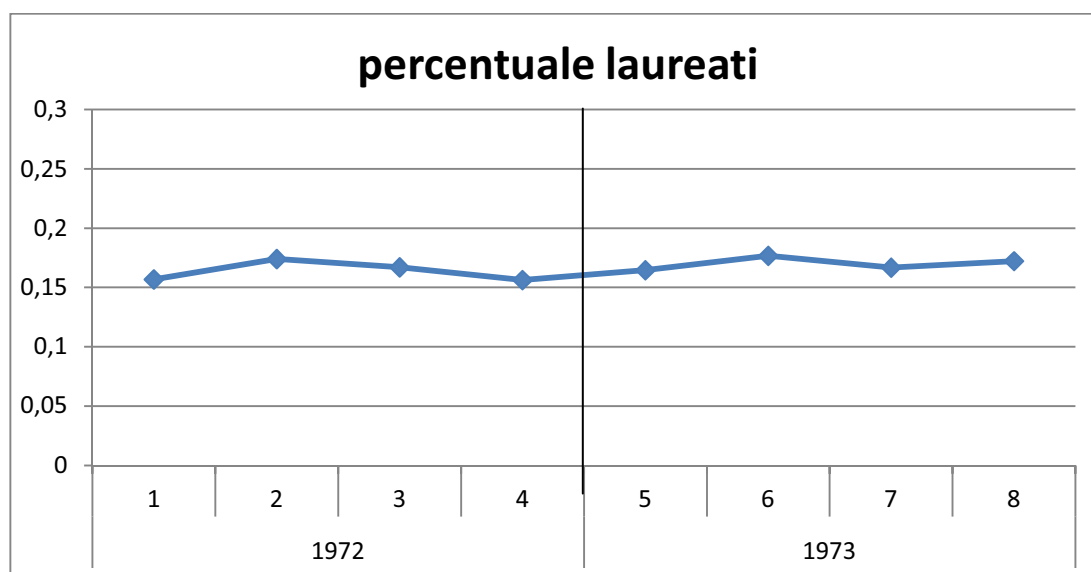


Figura 4. Percentuale individui con titolo di studio elevato per trimestre di nascita, coorti 1972-73.

Regredendo la variabile risposta sulle dummy di trimestre, possiamo valutare se le differenze osservate nel grafico sono o meno significative.

L'output sembra mostrarci una differenza negativa ma poco significativa (rifiuto dell'ipotesi di uguaglianza tra medie solo con un livello di significatività del 10%) tra 'quasi coetanei' (4° e 5° trimestre), e un'apparente stagionalità (media di laureati più alta) per il secondo trimestre dei due anni.

Regressione 11. Possesso di un titolo di studio elevato, coorti 1972-73

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 1° trimestre di nascita | -0,0078 | 0,0049 | 0,112 |
| 2° trimestre di nascita | 0,0091 | 0,0048 | 0,050 |
| 3° trimestre di nascita | 0,0024 | 0,0048 | 0,616 |
| 4° trimestre di nascita | -0,0082 | 0,0049 | 0,091 |
| 6° trimestre di nascita | 0,0122 | 0,0048 | 0,012 |
| 7° trimestre di nascita | 0,0021 | 0,0048 | 0,656 |
| 8° trimestre di nascita | 0,0077 | 0,0050 | 0,122 |
| costante | 0,1646 | 0,0035 | 0,000 |
| N° osservazioni | 95149 | | |

2.4 Retribuzione

La variabile più comunemente presa come riferimento per esaminare l'effetto della politica scolastica sul mercato del lavoro è la retribuzione, tramite la quale si valuta se le eventuali differenze dovute alla data di nascita si protraggano nel lungo termine.

Ci serviamo quindi della variabile 'retribuzione' fornitaci dai dati Istat per analizzare le medie per periodo di nascita, non prima di aver controllato le caratteristiche dei dati mancanti, che, per non invalidare la nostra analisi, devono essere indipendenti dalla data di nascita (vedi Appendice D).

La variabile viene rilevata solo a partire dall'anno 2008 ed è definita come 'Retribuzione netta del mese (precedente all'intervista) escluse altre mensilità (tredicesima, quattordicesima, ecc.) e voci accessorie non percepite regolarmente tutti i mesi (premi di produttività annuali, arretrati, indennità per missioni, straordinari non abituali, ecc.)'.

A differenza di quanto accade per le rilevazioni sul titolo di studio, l'informazione sullo stipendio percepito cambia a seconda di quando viene rilevata, indipendentemente da quale coorte si decide di esaminare.

Poichè disponiamo di rilevazioni trimestrali dal 2008 al 2010, può essere infatti che l'individuo del '72 venga intervistato quando ha 36 anni (quindi nel 2008) o quando ne ha 2 in più (2010). Ci vediamo però costretti, per avere una numerosità sufficiente a condurre un'analisi, ad utilizzare assieme tutti i 12 trimestri di rilevazione.

Per gli studi sulla retribuzione, in generale, sarà quindi necessario considerare anche l'età e/o il periodo di intervista per ogni individuo, oltre al trimestre di nascita, e prestare attenzione all'interpretazione delle stime.

Nel nostro caso particolare, utilizzando lo stesso modello visto finora (modello (1)) possiamo confrontare le medie per gli otto trimestri di nascita come lo si

farebbe nel caso in cui gli individui fossero rilevati nello stesso istante di tempo. Questo perché, se calcoliamo l'età media per ogni trimestre di nascita, vediamo che tra un gruppo e l'altro questa differisce circa di un trimestre (0.25), come se venissero intervistati nello stesso momento (vedi Tabella 5).

Tabella 5. Età media per trimestre di nascita, coorti 1972-73

| | Trimestre di nascita | Età media (in trimestri) | Differenze rispetto al trimestre precedente |
|------|----------------------|--------------------------|---|
| 1972 | 1 | 37,34744 | |
| | 2 | 37,14654 | -0,20 |
| | 3 | 36,85281 | -0,29 |
| | 4 | 36,62585 | -0,23 |
| 1973 | 5 | 36,37245 | -0,25 |
| | 6 | 36,11122 | -0,26 |
| | 7 | 35,83789 | -0,27 |
| | 8 | 35,63746 | -0,20 |

Di seguito, la Tabella 6 riporta le frequenze mentre il grafico (Figura 5) ci mostra la retribuzione media per trimestre.

Tabella 6. Numerosità dati retribuzione per trimestre di nascita, coorti 1972-73.

| | Trimestre di nascita | Frequenze |
|------|----------------------|-----------|
| 1972 | 1 | 3.159 |
| | 2 | 3.372 |
| | 3 | 3.272 |
| | 4 | 3.118 |
| 1973 | 5 | 3.321 |
| | 6 | 3.240 |
| | 7 | 3.524 |
| | 8 | 2.921 |
| | Tot. | 25.927 |

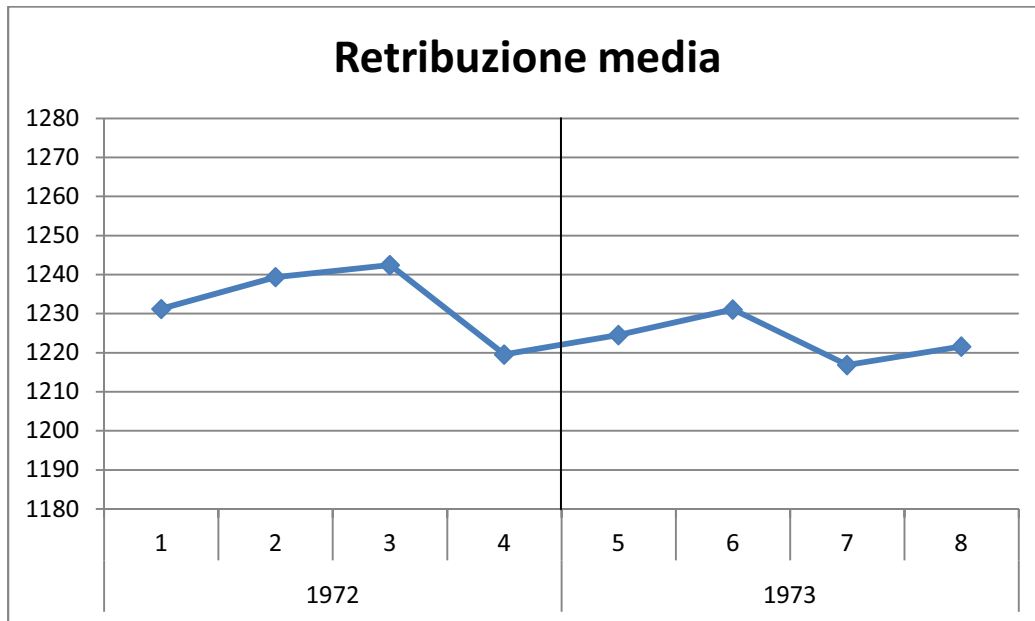


Figura 5. Retribuzione media per trimestre di nascita, coorti 1972-73.

Notiamo un andamento leggermente decrescente della retribuzione media, che sta ad indicare che questa cresce al crescere dell'età media.

Per verificare l'entità delle differenze tra i trimestri e la loro la significatività osserviamo la 'Regressione 12'.

Il fatto che l'ultimo trimestre abbia una media un po' più bassa del primo per entrambi gli anni risulta statisticamente irrilevante.

Quello che ci interessa maggiormente, però, è che nemmeno attorno alla soglia di nostro interesse le differenze sono significative. Quindi, alla luce di questi dati, si può dire che gli individui di ogni trimestre del 1972 e 1973 percepiscono, in media, la stessa retribuzione.

Regressione 12. Retribuzione,
coorti 1972-73

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 1° trimestre di nascita | 6,6510 | 11,6001 | 0,566 |
| 2° trimestre di nascita | 14,8185 | 11,4108 | 0,194 |
| 3° trimestre di nascita | 17,8649 | 11,4970 | 0,120 |
| 4° trimestre di nascita | -5,0081 | 11,6392 | 0,667 |
| 6° trimestre di nascita | 6,5029 | 11,5256 | 0,573 |
| 7° trimestre di nascita | -7,6725 | 11,2881 | 0,497 |
| 8° trimestre di nascita | -2,9848 | 11,8399 | 0,801 |
| costante | 1224,5710 | 8,0994 | 0,000 |
| N° osservazioni | 25927 | | |

2.5 Verifica della validità del metodo per due coorti

Buckles e Hungerman (2008) accusano la maggior parte degli studi di non aver tenuto conto delle caratteristiche della famiglia, che, a giudicare dalla loro analisi, avrebbero molta influenza sulla data di nascita dei figli.

Potrebbe essere ad esempio che, attraverso la programmazione delle nascite, i figli della famiglie più abbienti nascano con più probabilità in certi periodi piuttosto che in altri.

Nel caso in cui la data di nascita degli individui cambiasse a seconda delle caratteristiche socioeconomiche dei genitori, sarebbe infatti necessario, per valutare il vero effetto del periodo di nascita, tenerne conto nella stima dei modelli.

Le analisi finora presentate si basano sul presupposto che non vi siano variabili che possano influenzare la data di nascita e contemporaneamente la carriera scolastica di un individuo.

Utilizzando come dati coorti di bambini nati tra il 2003 e il 2004 stimiamo ora il solito modello a otto trimestri (modello (1)) utilizzando come variabile risposta una proxy della condizione sociale ed economica della famiglia a cui essi appartengono.

In questo modo potremmo verificare se vi sia o meno una differenza tra le medie per periodo di nascita.

La variabile creata indica il possesso di un titolo pari o superiore alla laurea da parte di almeno uno dei genitori della famiglia (vedi Appendice C).

$$Y_i = \begin{cases} 1, & \text{se almeno un genitore ha un titolo pari o superiore alla laurea} \\ 0, & \text{altrimenti} \end{cases}$$

Di seguito, al solito, la tabella delle numerosità e il grafico delle percentuali.

Tabella 7. Numerosità per trimestre di nascita, coorti 2003-04

| | Trimestre di nascita | Frequenze |
|--------|----------------------|-----------|
| 2003 | 1 | 1.946 |
| | 2 | 2.069 |
| | 3 | 2.149 |
| | 4 | 1.973 |
| 2004 | 5 | 1.838 |
| | 6 | 1.903 |
| | 7 | 2.111 |
| | 8 | 2.110 |
| Totale | | 16.099 |

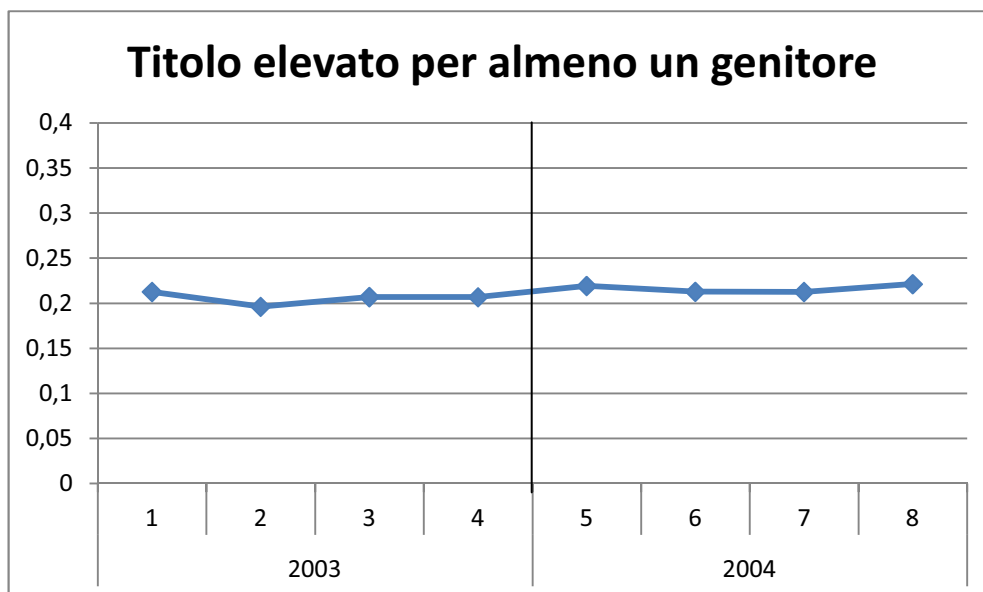


Figura 6. Percentuale figli con almeno un genitore che possiede un titolo di studio elevato, coorti 2003-04

Dalla Figura 6 non sembrerebbe apparire una relazione tra il periodo di nascita e la probabilità di avere un laureato in famiglia, cosa che ci viene confermata anche dalla non significatività dei coefficienti nel modello (Regressione 19).

Regressione 19. Possesso di titolo studio
 elevato all'interno della famiglia,
 coorti 2003-04

| | Coefficiente |
|-------------------------|----------------------------|
| | St. Error |
| | P value |
| 1° trimestre di nascita | -0,0065 0,0133 0,623 |
| 2° trimestre di nascita | -0,0230 0,0131 0,078 |
| 3° trimestre di nascita | -0,0122 0,0130 0,347 |
| 4° trimestre di nascita | -0,0125 0,0132 0,346 |
| 6° trimestre di nascita | -0,0064 0,0133 0,630 |
| 7° trimestre di nascita | -0,0066 0,0130 0,614 |
| 8° trimestre di nascita | 0,0021 0,0130 0,874 |
| costante | 0,2193 0,0095 0,000 |
| N° osservazioni | 16099 |

Le considerazioni fatte sugli outcome visti finora, dopo la verifica della validità delle ipotesi alla base, ci porterebbero a dire che: tra i nati nell'ultimo trimestre di un anno e il primo dell'anno successivo non vi siano differenze sostanziali. Sarebbe quindi indifferente, per i genitori, che il figlio nasca appena prima o appena dopo la data limite.

Non sembrerebbero esistere nemmeno andamenti particolari all'interno delle stesse coorti, se non una leggera stagionalità per il raggiungimento di un titolo di studio elevato.

C'è inoltre da sottolineare che, per ciò che riguarda l'eventuale conseguimento del diploma e della laurea, c'è una distinzione per gli appartenenti alla prima e seconda coorte per l'anno di conseguimento del titolo (un anno di differenza in media) ma non per la velocità con la quale i diversi gruppi di nascita lo raggiungono. Quindi, in ogni caso, un bambino di dicembre rispetto al suo coetaneo di gennaio sarà avvantaggiato per il fatto di essere andato a scuola un anno prima e potersi diplomare ed eventualmente finire l'università un anno prima.

Nonostante la validità del metodo finora adottato, ci siamo scontrati con il problema di una bassa numerosità, che non ci permette di generalizzare le nostre conclusioni, e di cogliere gli eventuali effetti che, come dimostrato in letteratura, potrebbero essere anche molto piccoli.

Nel prossimo capitolo presenteremo quindi alcune valide alternative che ci permettano di utilizzare un campione più ampio.

3. Analisi con più coorti

Proponiamo, in questo capitolo, alcuni metodi alternativi di analisi che siano coerenti con la logica finora seguita, ma con la possibilità di accorpate più coorti di nascita ed avere stime più precise a fronte dell'aumento delle osservazioni disponibili.

Ci concentriamo sugli ultimi due outcome: titolo di studio elevato e retribuzione.

La natura stessa delle variabili per la velocità del raggiungimento del diploma e della laurea triennale, la cui osservazione è vincolata agli anni di rilevazione, non ci permette invece di ripetere l'analisi unendo diverse coorti.

Si noti che con il fine di accrescere la numerosità si sono aggregati sia diversi anni di rilevazione sia diversi anni di nascita, e che, poiché le interviste sono fatte in momenti anche molto lontani, non vi è corrispondenza biunivoca tra età e data di nascita.

Abbiamo ritenuto opportuno estendere lo studio a sette anni totali, comprendendo le coorti che vanno dal 1968 al 1974. La scelta è stata fatta tenendo conto che si voleva aumentare la numerosità, ma allo stesso tempo unire individui con caratteristiche più simili possibile, e quindi non nati in anni troppo lontani. Inoltre, si sono selezionati proprio questi anni perché corrispondono alla popolazione più 'giovane' tra coloro la cui carriera scolastica si presuppone conclusa.

Conduciamo in primis delle analisi preliminari, per avere un'idea del comportamento delle variabili risultato per i trimestri delle diverse coorti.

Il grafico sottostante ci mostra un andamento crescente della percentuale di laureati: com'è ragionevole aspettarsi all'aumentare dell'anno di nascita aumenta il numero di coloro che raggiungono, nella loro carriera scolastica, almeno la laurea. Nel senso che se si osservano coorti più giovani c'è una probabilità più alta

di trovare laureati. Non ci sono invece chiare evidenze di un andamento legato al trimestre di nascita.

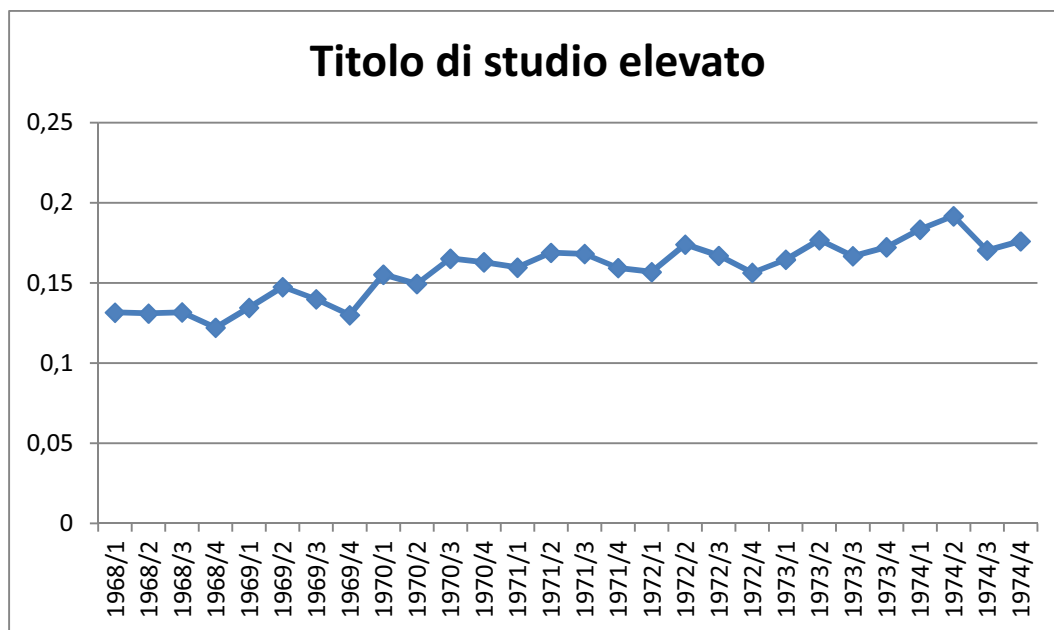


Figura 7. Percentuale individui con titolo di studio elevato per trimestre di nascita, coorti 1968-74.

Il grafico della retribuzione (Figura 8) mostra come, man mano che l'anno di nascita cresce, la retribuzione diminuisce. Detto in altri termini, all'aumentare dell'età, e quindi dell'esperienza lavorativa, la retribuzione aumenta.

Gli individui nati dopo nel tempo (quindi mediamente più giovani) percepiscono uno stipendio più basso rispetto a coloro che sono nati 6 anni prima, circa 100 euro in meno. Nemmeno in questo caso sembrerebbero esserci segnali di un legame con il trimestre di nascita.

Di seguito presentiamo due nuovi modelli che ci permettano di osservare, congiuntamente per gli anni 1968-74, la discontinuità attorno alla soglia.

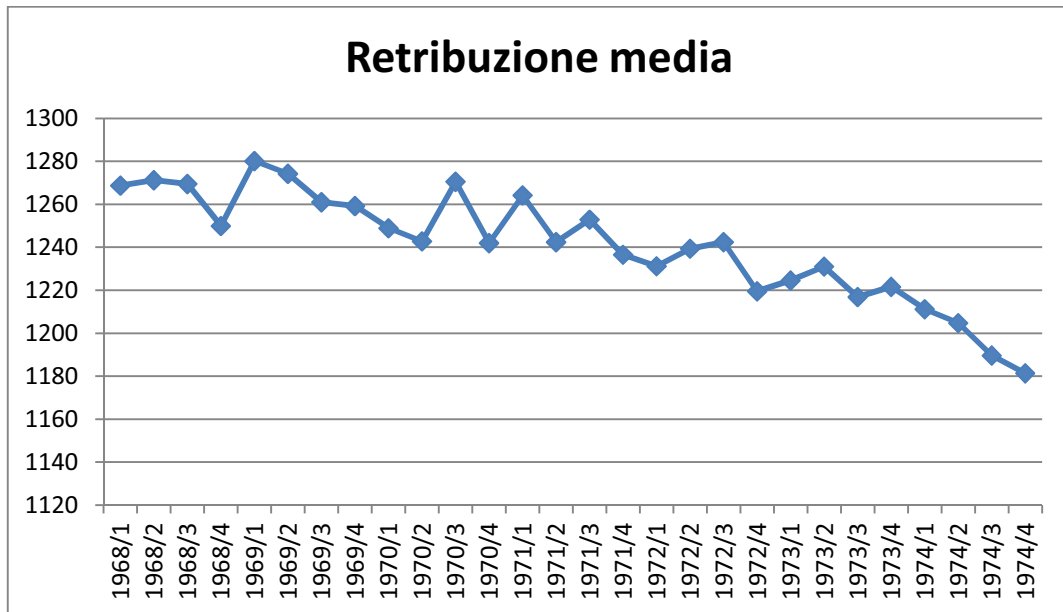


Figura 8. Retribuzione media per trimestre di nascita, coorti 1968-74.

3.1 Modello RDD con più anni

Il primo modello che stimeremo per entrambi gli outcome è una semplice applicazione del classico RDD, nel senso che le variabili risposta saranno osservate prendendo in considerazione solo i trimestri attorno alla soglia di discontinuità (31 dicembre).

Definiamo quindi il nuovo modello :

$$Y_i = \beta_0 + \beta_1 * gfm_i + \varepsilon_i \quad (2)$$

dove gfm_i è la variabile per il 'trattamento', che indica l'appartenenza al periodo subito prima o subito dopo la soglia.

E' definita come:

$$gfm_i = \begin{cases} 1, & \text{se l'individuo è nato a gennaio, febbraio o marzo} \\ 0, & \text{se l'individuo è nato a ottobre, novembre o dicembre} \end{cases}$$

ε_i è l'errore.

Il modello ((2)) viene quindi stimato su diversi anni ma solo sugli individui la cui data di nascita cade nell'ultimo o nel primo trimestre.

β_1 valuta il vantaggio (o lo svantaggio) dato dall'essere nati subito dopo la soglia rispetto all'essere nati subito prima.

Contrariamente a quanto avveniva nel caso degli otto trimestri, non è immediato identificare la risposta alla nostra iniziale domanda, ovvero se ci siano differenze, a parità di età, dovute alla diversa data di nascita.

Utilizzando un campione allargato si rischia infatti di confondere l'effetto all'interno di un anno di nascita (gennaio e dicembre dell'anno X) con l'effetto tra quasi coetanei appartenenti a due anni diversi (dicembre dell'anno X e gennaio dell'anno X+1).

La nostra proposta è quella di identificarli tramite una diversa selezione del campione.

Comprendendo tra i nostri dati il primo e l'ultimo trimestre di ogni anno di nascita a partire dal 68 fino al 74 (Tabella 8) si valuta la differenza media tra gli individui nati nei primi o negli ultimi tre mesi di uno stesso anno.

Tabella 8. Schema di selezione A.

| 1968 | | | | 1969 | | | | 1970 | | | | 1971 | | | | 1972 | | | | 1973 | | | | 1974 | | | | | | | |
|------|---|---|---|------|---|---|---|------|---|---|---|------|---|---|---|------|---|---|---|------|---|---|---|------|---|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |

Possiamo vedere infatti, in Tabella 9, che l'anno di nascita medio è lo stesso per i due gruppi, mentre invece le età differiscono di circa tre trimestri.

**Tabella 9. Media di età e anno di nascita per trattati e non,
schema di selezione A.**

| | Età media (in trimestri) | Anno di nascita medio |
|---------------------------------|-----------------------------|--------------------------|
| Ottobre Novembre Dicembre | 36,16 | 1970,90 |
| Gennaio Febbraio Marzo | 36,96 | 1970,86 |

Selezionando invece il campione dall'ultimo trimestre del 68 al primo del 74 (escludendo cioè gli individui del primo e dell'ultimo gruppo) si osserva la differenza media tra i trimestri a cavallo della soglia (Tabella 10).

Tabella 10. Schema di selezione B.

| | | | | | | | | | | | | | | | |
|------|---|------|---|------|---|------|---|------|---|------|---|------|---|---|---|
| 1968 | | 1969 | | 1970 | | 1971 | | 1972 | | 1973 | | 1974 | | | |
| 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |

Così facendo riproduciamo il caso di nostro interesse, ovvero confrontiamo individui nati in anni successivi e che differiscono di un solo trimestre per quanto riguarda l'età (Tabella 11).

**Tabella 11. Media di età e anno di nascita per trattati e non,
schema di selezione B.**

| | Età media (in trimestri) | Anno di nascita medio |
|---------------------------------|-----------------------------|--------------------------|
| Ottobre Novembre Dicembre | 36,64 | 1970,42 |
| Gennaio Febbraio Marzo | 36,43 | 1971,39 |

3.1.1 Titolo di studio

Come nel paragrafo 2.3, vogliamo verificare se ci siano differenze per il titolo di studio raggiunto, a fine carriera, da due individui quasi coetanei, eventualmente causate dalla presenza di una data prefissata che li separa in due anni scolastici diversi.

Di seguito (Tabella 12) osserviamo come auspicato un notevole aumento della numerosità per gruppo rispetto alle analisi viste finora.

**Tabella 12. Numerosità per trimestre di nascita
e schema di classificazione**

| Trimestri | Schema A | Schema B |
|-----------|----------|----------|
| 1° | 88002 | 74215 |
| 4° | 82194 | 71160 |
| Tot. | 170196 | 145375 |

Procediamo con la stima del modello (2), sulla base dei due metodi di selezione sopra descritti.

Regressione 13. Possesso di un titolo di studio elevato, solo 1° e 4° trimestre di nascita, schema A

| | Coefficiente St. Error P value |
|-------------------------|--------------------------------------|
| 1° trimestre di nascita | 0,0008 <i>0,0017</i> 0,630 |
| costante | 0,1532 <i>0,0013</i> 0,000 |
| N° osservazioni | 170196 |

Regressione 14. Possesso di un titolo di studio elevato, solo 1° e 4° trimestre di nascita, schema B

| | Coefficiente St. Error P value |
|-------------------------|--------------------------------------|
| 1° trimestre di nascita | 0,0086 <i>0,0019</i> 0,000 |
| costante | 0,1497 <i>0,0014</i> 0,000 |
| N° osservazioni | 145375 |

Se osserviamo la prima regressione (schema A) il coefficiente risulta chiaramente non significativo, vale a dire che non sembrerebbe esserci una differenza, nella percentuale di coloro che hanno raggiunto un titolo di studio elevato, tra i più giovani e i più vecchi della stessa coorte.

Selezionando lo stesso campione, ma escludendo gli estremi della prima e dell'ultima coorte (Schema B), osserviamo invece che il coefficiente per il confronto tra individui nati vicini ma in due coorti diverse risulta significativo e positivo, anche se piccolo.

Si potrebbe cioè dire che coloro che nascono subito dopo il 31 dicembre abbiano una probabilità più alta di quasi un punto percentuale di laurearsi rispetto a coloro che sono nati subito prima.

Questo risultato sembra riflettere più che altro il trend del grafico in Figura 7, dove si era visto che all'aumentare dell'anno di nascita aumenta il numero di laureati sul totale.

Ne risulta così che, per quanto riguarda il raggiungimento di un titolo di studio elevato, non vi è una differenza tra individui della stessa coorte, ma tra individui di coorti diverse.

3.1.2 Retribuzione

Nell'analizzare la retribuzione teniamo presente che, a differenza del titolo di studio, essa si modifica se viene misurata per la stessa persona in istanti di tempo diversi.

Poiché le informazioni sulla variabile in esame vengono rilevate solo dal 2008 in poi, e non per tutta la popolazione (vedi Appendice D), le numerosità (Tabella 13) sono più basse, ma comunque nettamente superiori al caso di analisi di due coorti.

Tabella 13. Numerosità per trimestre di nascita e schema di classificazione

| Trimestri | Schema A | Schema B |
|-----------|----------|----------|
| 1° | 23819 | 20118 |
| 4° | 21994 | 18952 |
| Tot. | 45813 | 39070 |

Di seguito riportiamo le stime del modello (2) per la retribuzione.

Regressione 15. Retribuzione, solo 1° e 4° trimestre di nascita, schema A

| | Coefficiente St. Error P value |
|-------------------------|--------------------------------------|
| 1° trimestre di nascita | 17,4101 4,5085 0,000 |
| costante | 1230,8390 3,2509 0,000 |
| N° osservazioni | 45813 |

Regressione 16. Retribuzione, solo 1° e 4° trimestre di nascita, schema B

| | Coefficiente St. Error P value |
|-------------------------|--------------------------------------|
| 1° trimestre di nascita | 5,7116 4,8684 0,241 |
| costante | 1238,7710 3,4935 0,000 |
| N° osservazioni | 39070 |

Al contrario del caso precedente, nella prima regressione il coefficiente di β_1 risulta positivo e significativo, e nella seconda risulta non significativo. Questo

vorrebbe dire che, all'interno di una stessa coorte, coloro che guadagnano di più in media sono i nati nei primi mesi dell'anno, rispetto ai loro compagni di coorte più piccoli. Al contrario, non sembrerebbe esserci un distacco tra i nati alla fine di un anno e all'inizio del successivo.

Si avrebbe quindi che a (quasi) parità di età la retribuzione non cambia in base alla posizione rispetto alla soglia, ma a parità di anno di nascita, coloro che nascono prima (più grandi in media) guadagnano di più.

3.2 Modello completo

La ricerca di un modello 'completo' e più generale per rappresentare i nostri dati è resa difficile proprio dalla natura dei dati stessi. Infatti, mentre molti studi si servono di censimenti o di informazioni rilevate in un momento preciso, noi disponiamo di rilevazioni effettuate nei quattro diversi trimestri e nell'arco di più anni.

Dovendole utilizzare tutte per avere una numerosità sufficiente e avendo come fine quello di analizzare le differenze tra trimestri di nascita, ci troviamo ad affrontare le tipiche difficoltà dell'identificazione età-periodo-coorte ('APC', si veda Mason e Fienberg (1985) per un'analisi completa del problema).

E' infatti problematico isolare gli effetti dell'appartenenza ad una data generazione o coorte (effetto coorte) dagli effetti che si associano da un lato ad una fase della vita che una persona sta attraversando (effetto età) e, dall'altro, alle condizioni della congiuntura (effetto periodo).

"Data di nascita, età e data di rilevazione sono infatti legate da una correlazione lineare perfetta, e, in generale, non è possibile identificare contemporaneamente tutti e tre gli effetti". (Schizzerotto, Trivellato, Sartor, 2011).

Pur non potendo distinguere le tre componenti, presentiamo il modello che a nostro parere rendere l'interpretazione delle stime più semplice e immediata possibile.

Diversamente da prima ci serviamo delle osservazioni di tutta la popolazione (non solo di quella attorno alla soglia) e inseriamo quindi le dummy per tutti i trimestri di nascita.

Prendiamo tutti i nati tra il 1968 e il 1974 (dal 1° trimestre del '68 al 4° trimestre del '74) e, per non confondere l'effetto dicembre-gennaio dello stesso anno con dicembre-gennaio di anni successivi inseriamo dei controlli per la coorte di nascita.

Nell'analisi sulla retribuzione inseriamo inoltre le dummy per i 12 trimestri di rilevazione, poiché ci aspettiamo che vi sia un effetto medio crescente nel tempo.

Cosa che non accade invece per l'analisi sul titolo di studio che, poiché è misurata quando gli individui sono adulti (e quindi a carriera scolastica conclusa) rimane invariante nel tempo (si veda Par. 2.3).

$$Y_i = \beta_1 + \beta_2 * td2_i + \beta_3 * td3_i + \beta_4 * td4_i + \beta_5 * annonas_i + \\ + \beta_6 * tril2_i + \beta_7 * tril3_i + \dots \dots \dots + \beta_{12} * tril12_i + \varepsilon_i \quad (3)$$

dove Y_i è la variabile risposta;
 $td2_i, td3_i, td4_i$ sono le dummy per i trimestri di nascita, esclusa quella per il 1°, usata come baseline ;
 $annonas_i$ è l'anno di nascita dell'individuo, usando il 1968 come anno zero;
 $tril2_i, \dots, tril12_i$ sono le dummy per trimestre di rilevazione;
 ε_i è l'errore.

Attraverso questo modello possiamo osservare i diversi effetti di nostro interesse.

I coefficienti β_2 , β_3 , β_4 ci diranno quali siano le differenze negli outcome tra nati nel trimestre i-esimo e il 1° trimestre dello stesso anno.

Confrontando invece il coefficiente stimato per l'anno di nascita (β_5) e quello per l'ultimo trimestre (β_4) si potrà valutare cosa succede attorno alla soglia (illustreremo degli esempi per le due variabili in esame).

La costante ci indica il valore medio per i nati nel primo trimestre del 1968 e, se si analizza la retribuzione, osservati nella prima rilevazione.

3.2.1 Titolo di studio

Vediamo che, nell'analisi per il livello di studio, le numerosità per gruppo aumentano notevolmente comprendendo una popolazione così ampia (Tabella 14).

Tabella 14. Numerosità analisi titolo di studio per trimestre, coorti 1968-74

| Trimestre di nascita | Frequenze |
|----------------------|-----------|
| 1 | 88.002 |
| 2 | 88.091 |
| 3 | 90.820 |
| 4 | 82.194 |
| Tot. | 349.107 |

Di seguito vediamo le stime del modello (3) per il possesso di un titolo di studio elevato.

A prima vista notiamo che, a parità di anno di nascita, non vi sono differenze significative tra i nati nel primo e nell'ultimo mese dell'anno (come risultava dall'analisi nel paragrafo 3.1) ma appare un picco stagionale.

Sembra cioè che, in media, coloro che sono nati ad aprile, maggio o giugno abbiano una probabilità più alta di possedere una laurea (o un titolo superiore a essa), seppur molto piccola (0.7 punti percentuali in più).

Regressione 17. Possesso di un titolo di studio elevato, coorti 1968-74

| | Coefficiente |
|-------------------------|--------------|
| | St. Error |
| | P value |
| 2° trimestre di nascita | 0,0074 |
| | 0,0017 |
| | 0,000 |
| 3° trimestre di nascita | 0,0033 |
| | 0,0017 |
| | 0,052 |
| 4° trimestre di nascita | -0,0012 |
| | 0,0018 |
| | 0,501 |
| anno di nascita | 0,0080 |
| | 0,0003 |
| | 0,000 |
| costante | 0.1311 |
| | 0,0015 |
| | 0,000 |
| N° osservazioni | 349107 |

Anche al crescere della coorte di nascita aumentano coloro che possiedono un titolo di studio elevato, precisamente la percentuale cresce di 0.8 ogni anno, cosa che avevamo osservato anche nell'analisi RDD del precedente paragrafo.

Per vedere invece se vi è una differenza significativa tra l'essere nati a gennaio dell'anno X+1 e a dicembre dell'anno X dobbiamo sottrarre al coefficiente per l'anno di nascita (β_5) il coefficiente per il quarto trimestre (β_4) : $0.008 - (-0.0012) = 0.0092$.

La differenza è significativa poiché l'ipotesi di uguaglianza tra i due coefficienti ($H_0: \beta_4 = \beta_5$) viene rifiutata ($F(1,349102) = 26.39, p - value = 0.0000$).

Questo significa che coloro che nascono a gennaio dell'anno dopo hanno una probabilità più alta (quasi di un punto percentuale) di laurearsi, conseguentemente al fatto che il numero medio di laureati aumenta man mano che si analizzano coorti più giovani.

Una effetto di questo tipo si era riscontrato anche nel modello per le sole due coorti, dove il coefficiente risultava statisticamente poco significativo (vedi Par. 2.3).

3.2.2 Retribuzione

La Tabella 15 mostra un sostanziale aumento delle numerosità per trimestre di osservazione, anche per la stima del modello (3) per la retribuzione, nonostante il fatto di dover usare solo le rilevazioni a partire dal 2008.

Tabella 15. Numerosità dati retribuzione per trimestri di nascita, coorti 1968-74

| Trimestre di nascita | Frequenze |
|----------------------|-----------|
| 1 | 23.819 |
| 2 | 23.763 |
| 3 | 24.675 |
| 4 | 21.994 |
| Total | 94.251 |

Pare che, a parità di coorte, gli individui nati alla fine dell'anno abbiano una retribuzione significativamente più bassa (quasi di 17 euro) rispetto ai loro colleghi più grandi.

Si osserva inoltre che, all'aumentare dell'età media e quindi degli anni di esperienza (inversamente proporzionali all'anno di nascita) aumenta la retribuzione, di circa 11 euro.

Regressione 18. Retribuzione,
coorti 1968-74

| | Coefficiente | St. Error | P value |
|------------------------------|--------------|-----------|---------|
| 2° trimestre di nascita | -3,3673 | 4,4282 | 0,447 |
| 3° trimestre di nascita | -3,8642 | 4,3869 | 0,378 |
| 4° trimestre di nascita | -16,9335 | 4,5162 | 0,000 |
| anno di nascita | -11,2075 | 0,7828 | 0,000 |
| 2° trimestre di rilevazione | 6,919 | 7,4828 | 0,355 |
| 3° trimestre di rilevazione | 23,915 | 7,6321 | 0,002 |
| 4° trimestre di rilevazione | 26,716 | 7,6077 | 0,000 |
| 5° trimestre di rilevazione | 34,179 | 7,5419 | 0,000 |
| 6° trimestre di rilevazione | 35,384 | 7,5804 | 0,000 |
| 7° trimestre di rilevazione | 40,627 | 7,6990 | 0,000 |
| 8° trimestre di rilevazione | 56,048 | 7,5588 | 0,000 |
| 9° trimestre di rilevazione | 63,2491 | 7,5636 | 0,000 |
| 10° trimestre di rilevazione | 66,1879 | 7,5802 | 0,000 |
| 11° trimestre di rilevazione | 74,187 | 7,6525 | 0,000 |
| 12° trimestre di rilevazione | 76,8473 | 7,6055 | 0,000 |
| costante | 1239.063 | 6.3383 | 0,000 |
| N° osservazioni | 94251 | | |

Se si vogliono confrontare gli individui con un'età simile ma appartenenti a coorti diverse il calcolo, come prima, è dato da $\beta_5 - \beta_4 = 5.726$. Ma la differenza non è significativa (ipotesi $H_0: \beta_4 = \beta_5$ accettata con $F(1,349102) = 1.56$ e $p - value = 0.2120$).

I coefficienti dei trimestri di rilevazione mostrano infine una prevedibile tendenza all'aumento medio delle retribuzioni nei tre anni di osservazione.

3.3 Verifica della validità del metodo per più coorti

Poiché avevamo verificato l'ipotesi di indipendenza tra caratteristiche della famiglia e periodo di nascita solo per due coorti (paragrafo 2.5), vogliamo assicurarci che sia valida anche estendendo l'analisi a più coorti di nascita, come abbiamo proceduto nel corso del capitolo. Prendiamo in esame le coorti che vanno dal 2000 al 2004.

Utilizziamo il seguente modello:

$$Y_i = \beta_1 + \beta_2 * td2_i + \beta_3 * td3_i + \beta_4 * td4_i + \varepsilon_i \quad (4)$$

dove $Y_i = \begin{cases} 1, & \text{se almeno un genitore ha un titolo pari} \\ & \text{o superiore alla laurea} \\ 0, & \text{altrimenti} \end{cases}$

$td2_i, td3_i, td4_i$ sono le dummy per i trimestri di nascita, esclusa quella per il 1°, usata come baseline ;

ε_i è l'errore.

Osserviamo, in Tabella 16, un notevole aumento delle numerosità se si esaminano 5 coorti.

Tabella 16. Numerosità dati titolo di studio genitori per trimestri di nascita, coorti 2000-04

| Trimestre di nascita | Frequenze |
|----------------------|-----------|
| 1 | 9.726 |
| 2 | 10.111 |
| 3 | 10.813 |
| 4 | 10.249 |
| Total | 40.899 |

Regressione 16. Possesso di titolo studio elevato all'interno della famiglia, coorti 2000-04

| | Coefficiente | St. Error | P value |
|-------------------------|--------------|-----------|---------|
| 2° trimestre di nascita | -0,0003 | 0,0056 | 0,964 |
| 3° trimestre di nascita | -0,0073 | 0,0055 | 0,187 |
| 4° trimestre di nascita | 0,0037 | 0,0056 | 0,510 |
| costante | 0,1937 | 0,0040 | 0,000 |
| N° osservazioni | 40899 | | |

Le stime, anche aumentando la numerosità, continuano a non essere significative perciò possiamo concludere che non vi sia una relazione tra periodo di nascita e livello di istruzione della famiglia e confermare la validità dell'ipotesi sulla quale abbiamo condotto il nostro studio.

4. Conclusioni

In Italia il Ministero dell'Istruzione stabilisce che si debbano iscrivere alla scuola primaria i bambini che compiono sei anni entro il 31 dicembre dell'anno di riferimento. Questo comporta che nella stessa classe vi siano individui con età molto diverse (i nati in gennaio hanno quasi un anno in più dei nati in dicembre, a parità di coorte) e che invece individui con età molto simili (dicembre di un anno e gennaio del successivo) debbano andare a scuola in anni diversi.

Attraverso questo studio abbiamo voluto verificare se ci fossero eventuali differenze, sui risultati di medio-lungo termine, tra gli individui nati attorno alla soglia. In particolare ci interessava confrontare quelli che hanno quasi la stessa età ma sono nati in anni diversi, e che quindi, in funzione della soglia stabilita, si iscrivono a scuola in anni diversi.

Dalle analisi effettuate nella prima parte della tesi (Capitolo 2) è emerso che, quanto a velocità di raggiungimento del titolo di studio, i ragazzi di coorti successive si diplomano (o laureano) con un anno di distanza, ovvero ci mettono in media lo stesso tempo e che non vi sono differenze secondo il trimestre di nascita. Concentrandoci sul confronto attorno alla soglia, i nati alla fine dell'anno non sembrano penalizzati dal fatto di essere più giovani rispetto ai loro compagni e raggiungono comunque il titolo di studio un anno prima rispetto a coloro che sono nati vicini a loro nel tempo ma fanno parte dell'altra coorte.

Si nota inoltre che circa il 10% dei nati a gennaio febbraio o marzo si diploma prima del tempo, essendosi evidentemente iscritto a scuola un anno prima, possibile spiegazione del fatto che essi sono anche coloro che hanno una più alta probabilità di laurearsi presto.

Nel medio termine, invece, il titolo di studio più elevato viene raggiunto con la stessa probabilità dagli individui delle due coorti, e quindi non vi sono differenze

tra i nati negli ultimi mesi di un anno e dai primi di quello dopo. Anche la retribuzione sembra rimanere invariata per coloro che sono nati in periodi a cavallo tra le due coorti.

Se l'estensione dell'analisi a più coorti aumenta la numerosità e quindi la precisione delle stime, è vero anche che comporta una minore semplicità nell'interpretazione dei dati.

Essendo il titolo di studio misurato 'a fine carriera' non si possono fare confronti tra età al momento dell'osservazione. Possiamo dire però che non vi sia differenza, nella percentuale di coloro che possiedono un titolo di studio elevato, tra coloro che, al tempo della scuola, erano i più piccoli e i più grandi della classe, ma che vi sia un piccolo distacco tra coorti diverse. Questo perché vi è un evidente aumento della percentuale di laureati all'aumentare degli anni di nascita (in questo caso dal 1968 al 1974).

Si può dire invece che due individui (osservati nello stesso momento) nati vicini ma in anni diversi percepiscono in media la stessa retribuzione, mentre tra i compagni di coorte i più piccoli (nati del 4° trimestre) sembrerebbero ricevere uno stipendio significativamente più basso.

Alla luce dei risultati finora descritti possiamo concludere però che, per dei genitori in attesa del proprio figlio, non ci siano motivi per sperare che questo nasca prima o dopo del 31 dicembre, se non per il fatto che, forse per un effetto 'trascinamento', egli avrà una minima probabilità in più di iscriversi all'università se nascerà nella coorte successiva.

Dato che disponiamo di una numerosità non sempre sufficiente per trarre conclusioni generali, soprattutto per la parte che riteniamo di maggiore interesse (Cap. 2), preferiamo, più che cogliere i risultati che di volta in volta emergono, evidenziare il metodo e la logica sottostante.

Questo con l'intento di creare una base per studi futuri, che, disponendo magari di una maggiore quantità di informazioni, potranno sia estendere l'analisi sia portare a conclusioni più attendibili.

Una delle cose che ci sembra più importante, qualora i dati lo permettano, è ripetere l'analisi tenendo conto di quali siano i cosiddetti 'non compliers', ovvero quel 10% della popolazione che, pur appartenendo ad una certa coorte, si iscrive a scuola un anno prima del previsto, assieme agli individui della coorte precedente. Sarebbe interessante vedere come questi influenzano i risultati finora visti, soprattutto la retribuzione, e stimare il vero effetto del trattamento.

Bibliografia

Angrist, J. D. and Krueger, A. B. (1991) : “Does Compulsory School Attendance Affect Schooling and Earnings?”, *The Quarterly Journal of Economics* 106/4

Bedard, K. and Dhuey, E. (2006): “The persistence of early childhood maturity: International evidence of long-run age effects”, *The Quarterly Journal of Economics* 121/4.

Black, S., Devereux, P. and Salvanes, K. (2008): “Too young to leave the nest? The effects of school starting age”, NBER working paper no. 13969.

Buckles, K. and Hungerman, D. (2008): “Season of birth: Old questions, new answers”, NBER working paper no. W14573.

Cascio, E. and Schanzenbach, D. (2007): “First in class? The education production function”, NBER working paper no. W13663

Crawford, C., Dearden, L. and Meghir, C.(2007): “When you are born matters: the impact of date of birth on child cognitive outcomes in England”, IFS working papers No. 10/06

Dobkin, C. and Ferreira, F. (2007): “Do school entry laws affect educational attainment and labor market outcomes?”, *Economics of education review* no. 29/1.

Elder, T. and Lubotsky, D. (2009): “Kindergarten entrance age and children’s achievement: Impacts of state policies, family background and peers”, *Journal of Human Resources* 44/3.

Fertig, M. and Kluve, J. (2005): “The effect of age at school entry on educational attainment in Germany”, IZA discussion paper no 1507.

Fredricksson, P. and Öckert, B. (2006): "Is Early Learning Really More Productive? The Effect of School Starting Age on School and Labor Market Performance", IFAU Working Paper 2006/12

Hahn, J., Todd, P. and Van der Klaauw, W. (2001): "Identification and estimation of treatment effects with a Regression-Discontinuity Design", *Econometrica*, Vol.69 No.1, 201-209

Jürges, H. and Schneider, K. (2007): "What can go wrong will go wrong: Birthday effects and early tracking in the German school system", CESifo Working Paper No. 2055.

Mason, W.W. and Fienberg, S.E. (1985): "Cohort analysis in social research: beyond the identification problem", Springer Verlag, New York

Puhani, P. A. and Weber, A.M. (2007): "Persistence of the school entry age effect in a system of flexible tracking", Institute for the Study of Labor Discussion Paper No. 2965.

Schizzerotto, A. , Trivellato, U. e Sartor, N.(2011): "Generazioni disuguali", ed. Il Mulino

Solli I.F. (2011): "Left behind by birth month", Uis Working Papers in Economics and Finance No 2012/8

Strøm, B. (2004): "Student achievement and birthday effects", working paper Department of Economics, NTNU.

Trivellato, U. (2009): "La valutazione degli effetti di politiche pubbliche: paradigma e pratiche", Irvapp Discussion Paper N. 2009-01

APPENDICE

A. Costruzione variabili di interesse per il titolo di studio

Le variabili utilizzate per le analisi sul raggiungimento di diploma di maturità, laurea triennale e di un titolo di studio elevato sono state costruite sulla base della risposta alla seguente domanda del questionario RFL .

-Titstud_ Qual è il titolo di studio più elevato che Lei ha conseguito?

- ❖ *Nessun titolo* 1
- ❖ *Licenza elementare* 2
- ❖ *Licenza media (o avviamento professionale)* 3
- ❖ *Diploma di scuola superiore di 2-3 anni che non permette l'iscrizione all'Università* 4
- ❖ *Diploma di scuola superiore di 4-5 anni che permette l'iscrizione all'Università* 5
- ❖ *Diploma di Accademia Belle Arti, Istituto Superiore Industrie Artistiche, Accademia di arte drammatica, Perfezionamento Accademia di Danza, Perfezionamento Conservatorio, Perfezionamento Istituto di Musica Pareggiato, Diploma accademico di alta formazione artistica e musicale* 6
- ❖ *Diploma universitario di due/tre anni, Scuola diretta a fini speciali, Scuola parauniversitaria* 7
- ❖ *Laurea di tre anni di primo livello (nuovo ordinamento)* 8
- ❖ *Laurea specialistica di due anni di secondo livello (nuovo ordinamento)* 9
- ❖ *Laurea di quattro anni o più (vecchio ordinamento o nuova laurea specialistica a ciclo unico)* 10

❖ *Specializzazione post-laurea (compresi master di 1° e 2° livello)*

11

❖ *Dottorato di ricerca*

12

Elenchiamo brevemente i criteri con i quali sono state create.

- Variabile indicatrice per ‘possesso del diploma di maturità’(par 2.1):

$$Y_i = \begin{cases} 1, & \text{se 'Titstud' è maggiore o uguale a 5} \\ 0, & \text{altrimenti} \end{cases}$$

Da notare che escludiamo dai ‘diplomati’ coloro che hanno completato non più di 2-3 anni di scuola superiore. Scelta dettata anche dal fatto che vogliamo osservare il cambiamento di stato (da 0, non diplomati a 1, diplomati) nel corso degli anni di rilevazione, e non sarebbe possibile se si mescolassero titoli con diversa durata. Precisiamo inoltre che essi rappresentano meno del 5% del campione.(coorti ‘87-‘88). Si considerano inoltre diplomati i pochi (meno dello 0.05 %) che dichiarano titoli più elevati incompatibili con l’età.

- Variabile indicatrice per ‘possesso almeno della laurea triennale’(par 2.2):

$$Y_i = \begin{cases} 1, & \text{se 'Titstud' è maggiore o uguale a 8} \\ 0, & \text{altrimenti} \end{cases}$$

Ovvero vengono inclusi sia quelli che raggiungono la laurea di 3 anni sia quelli che raggiungono titoli più elevati entro gli anni di rilevazione.

- Variabile indicatrice ‘possesso di titolo di studio elevato’ (par 2.3, 3.1):

$$Y_i = \begin{cases} 1, & \text{se 'Titstud' è maggiore o uguale a 7} \\ 0, & \text{altrimenti} \end{cases}$$

che indica se un individuo ha raggiunto la laurea o un titolo superiore ad essa. Si noti che per il possesso di un titolo elevato abbiamo compreso anche coloro i quali possiedono un diploma universitario o un titolo di scuola parauniversitaria (Titdtud=7) poiché lo riteniamo quasi al pari di una laurea. Non era invece possibile accorpere questa categoria ai possessori di una laurea triennale, dal momento che non riuscivamo a determinarne la durata prevista per il conseguimento.

B. Costruzione variabili per l'anno scolastico

- Per quanto riguarda l'analisi sui diplomati delle coorti 1987-88, la variabile creata raggruppa tutte le rilevazioni in modo tale da valutare, per ogni anno scolastico, quale sia la percentuale di diplomati alla conclusione di quell'anno o dei precedenti.

Si è deciso di escludere per ogni anno di rilevazione parte del mese di giugno e di luglio, poiché, come appare dal grafico (Figura 9), tra coloro che erano sul punto di diplomarsi alcuni rispondevano di esserlo già e alcuni no.

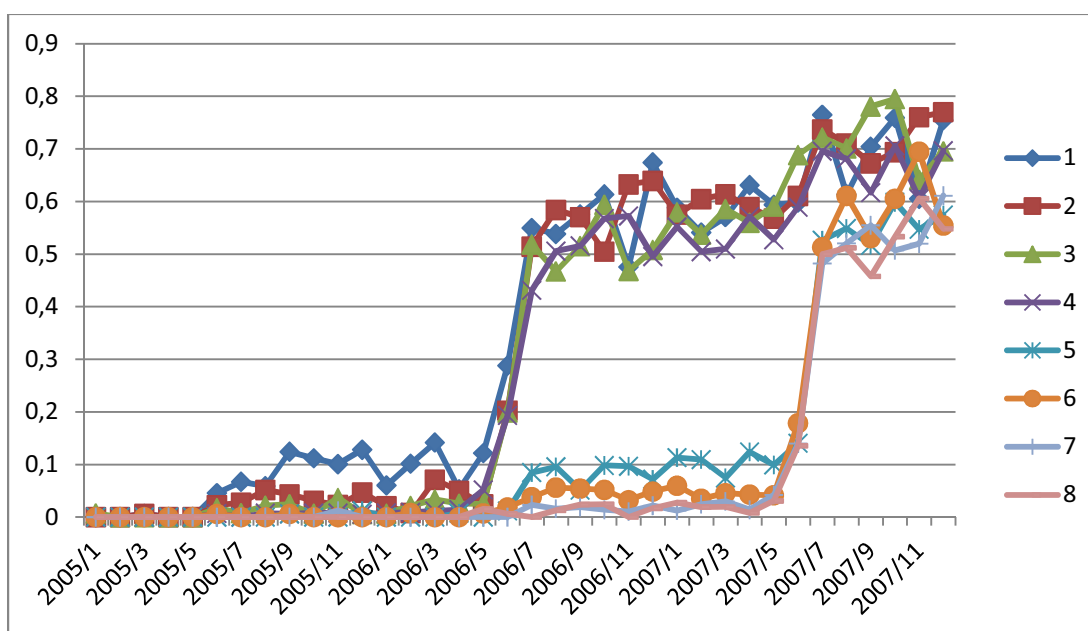


Figura 9. Percentuale diplomati per mese di rilevazione e trimestre di nascita, coorti 1987-88 (focus anni di rilevazione 2005-07).

La suddivisione è stata fatta considerando che tutti coloro che, prima di giugno di un anno specifico, rispondono di essere diplomati, significa che si sono diplomati alla fine dell'anno prima, poiché il titolo si può conseguire solo tra giugno e luglio.

$$\text{Anno scolastico diploma} = \begin{cases} 0405 & \text{per le interviste da metà luglio '05 a inizio giugno '06} \\ 0506 & \text{per le interviste da metà luglio '06 a inizio giugno '07} \\ 0607 & \text{per le interviste da metà luglio '07 a inizio giugno '08} \\ 0708 & \text{per le interviste da metà luglio '08 a inizio giugno '09} \\ 0809 & \text{per le interviste da metà luglio '09 a inizio giugno '10} \end{cases}$$

- La variabile anno scolastico per i laureati triennali, poiché vi sono diverse sessioni di laurea in un anno, considera tutti gli intervistati in quell'anno accademico, senza escludere nessun mese dell'anno.

$$\text{Anno scolastico laurea} = \begin{cases} 0506 & \text{per le interviste da ottobre '05 a settembre '06} \\ 0607 & \text{per le interviste da ottobre '06 a settembre '07} \\ 0708 & \text{per le interviste da ottobre '07 a settembre '08} \\ 0809 & \text{per le interviste da ottobre '08 a settembre '09} \\ 0910 & \text{per le interviste da ottobre '09 a settembre '10} \end{cases}$$

C. Costruzione variabili per famiglia

Per avere una proxy per il livello di istruzione della famiglia di un individuo, abbiamo calcolato il valore massimo del titolo di studio tra i genitori all'interno della stessa.

$$\text{Titolo di studio genitore} = \begin{cases} 1 & \text{se 'Titstud' è maggiore o uguale a 7} \\ & \text{e l'individuo è 'padre' o 'madre'} \\ 0 & \text{altrimenti} \end{cases}$$

$$\text{Almeno un genitore con titolo elevato} = \begin{cases} 1 & \text{se } \max(\text{titolo di studio genitore}) = 1 \\ 0 & \text{altrimenti} \end{cases}$$

D. Analisi dati mancanti sulle retribuzioni

All'interno del questionario Istat per la Rilevazione Forze Lavoro i dati relativi alla retribuzione vengono raccolti solo per coloro che si definiscono 'dipendenti' ma non per gli 'autonomi' o i 'collaboratori'. Risulta quindi indispensabile controllare le caratteristiche di questo gruppo, dal momento che possono interferire con le nostre considerazioni.

Dobbiamo pertanto verificare che i dipendenti si distribuiscano equamente all'interno dei trimestri delle nostre coorti. In caso contrario infatti, significherebbe che vi è una relazione tra l'essere dipendenti e la data di nascita, e non potremmo concludere circa la relazione tra i trimestri di nascita e la retribuzione.

Conduciamo l'analisi sugli anni di nascita analizzati nel corso della tesi, ovvero le coorti dal 1968 al 1974.

Per prima cosa ci assicuriamo che il dato relativo alla retribuzione sia presente solo per gli individui occupati, anche se per circa il 26% di loro il dato manca.

Tabella 17. Presenza dati retribuzione per condizione lavorativa, coorti 1968-74.

| | Dato presente | Dato mancante | Totale |
|-------------------------|---------------|----------------|----------------|
| Occupato | 94.251 | 33.431 | 127.682 |
| % | 73,82% | 26,18% | 100,00% |
| In cerca di occupazione | 0 | 8.345 | 8.345 |
| % | 0,00% | 100,00% | 100,00% |
| Inattivo | 0 | 34.278 | 34.278 |
| % | 0,00% | 100,00% | 100,00% |
| Totale | 94.251 | 76.054 | 170.305 |
| % | 55,34% | 44,66% | 100,00% |

Osserviamo poi che i dati mancanti corrispondono esattamente ai 'non dipendenti'.

Tabella 18. Presenza dati retribuzione per tipo di lavoro, coorti 1968-74.

| | Dato presente | Dato mancante | Totale |
|--------------------------|----------------|----------------|----------------|
| Autonomo - collaboratore | 0 | 33.431 | 33.431 |
| % | 0,00% | 100,00% | 100,00% |
| Dipendente | 94.251 | 0 | 94.251 |
| % | 100,00% | 0,00% | 100,00% |
| Total | 94.251 | 33.431 | 127.682 |
| % | 73,82% | 26,18% | 100,00% |

Verifichiamo quindi che non vi sono andamenti particolari all'interno dei trimestri e che i coefficienti di differenze fra medie non sono significativi.

Tabella 19. Tipo di lavoro per trimestre di nascita, coorti 1968-74

| | Autonomo - collaboratore | Dipendente | Totale |
|--------------|-----------------------------|---------------|----------------|
| 1° trimestre | 8.391 | 23.819 | 32.210 |
| % | 26,05% | 73,95% | 100,00% |
| 2° trimestre | 8.476 | 23.763 | 32.239 |
| % | 26,29% | 73,71% | 100,00% |
| 3° trimestre | 8.760 | 24.675 | 33.435 |
| % | 26,20% | 73,80% | 100,00% |
| 4° trimestre | 7.804 | 21.994 | 29.798 |
| % | 26,19% | 73,81% | 100,00% |
| Totale | 33.431 | 94.251 | 127.682 |
| % | 26,18% | 73,82% | 100,00% |

Regressione 20. Lavoratori dipendenti, coorti
1968-74

| | Coefficiente St. Error P value |
|-------------------------|--------------------------------------|
| 2° trimestre di nascita | -0,0024 0,0035 0,488 |
| 3° trimestre di nascita | -0,0015 0,0034 0,664 |
| 4° trimestre di nascita | -0,0014 0,0035 0,695 |
| costante | 0,7395 0,0024 0,000 |
| N° osservazioni | 127682 |

Simili risultati si ottengono anche per la condizione di occupato.

Possiamo quindi dire che non vi sia una relazione tra la data di nascita e la condizione lavorativa, e che i dipendenti si distribuiscano in modo equo tra i gruppi di nostro interesse.