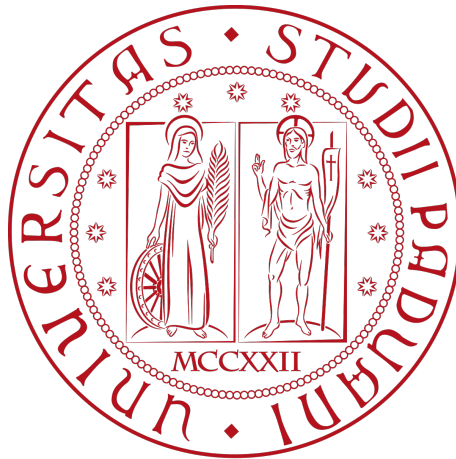


Università degli studi di Padova
Dipartimento di Scienze Statistiche
Corso di Laurea Triennale in
Statistica per le Tecnologie e le Scienze



**L'EFFETTO DELLE MIGRAZIONI SULLE DINAMICHE
DEMOGRAFICHE: APPLICAZIONE AL CASO DELLE
PROVINCE ITALIANE**

Relatore Prof. Stefano Mazzuco
Dipartimento di Scienze Statistiche

Romeo Silvestri
Matricola N. 1169828

Anno Accademico 2019/2020

Indice

1	Introduzione	3
2	Effetto delle migrazioni sui tassi e indici di struttura	4
2.1	Tassi e misure delle nascite	4
2.1.1	Tasso di fecondità totale	4
2.1.2	Età media al parto	5
2.1.3	Tasso di natalità	5
2.2	Indici di struttura	6
2.2.1	Età media	6
2.2.2	Indice di vecchiaia	6
2.2.3	Indice di dipendenza strutturale	7
2.2.4	Indice di ricambio della popolazione attiva	7
2.3	Modelli per l'analisi	8
2.3.1	Modello ols	8
2.3.2	Modello glm gamma	9
2.3.3	Modello "first difference"	10
2.4	Adattamento del modello	11
3	Effetto delle migrazioni sulla struttura per età	13
3.1	Estensione del modello dei tassi di crescita specifici per età	14
3.2	Effetto diretto e indiretto delle migrazioni	15
4	Analisi delle province italiane	18
4.1	Contesto storico dell'emigrazione italiana	19
4.2	Contesto storico dell'immigrazione in Italia	20
4.3	Dataset e dati mancanti	21
4.4	Analisi Esplorativa	22
4.4.1	Analisi univariata	22
4.4.2	Serie storiche	28
4.4.3	Analisi bivariata	30
4.5	Applicazione dei modelli	34
4.6	Adattamento e confronto tra modelli	43
5	Conclusione	44
5.1	Critiche e suggerimenti	44
5.2	Appendice	45
5.3	Bibliografia	48

Introduzione

Per comprendere il fenomeno dei movimenti migratori risulta fondamentale definire in primo luogo cosa si intende per migrazione, in quanto ci possono essere varie specificazioni, ciascuna con le proprie caratteristiche ed effetti associati.

Secondo l'Istat le migrazioni di maggior rilevanza a livello demografico risultano quegli eventi, trattati come dati di flusso, che coincidono con le iscrizioni o cancellazioni all'anagrafe per trasferimento di residenza. Le migrazioni hanno dunque un impatto sull'ammontare della popolazione residente e sulla relativa struttura per età e per genere; inoltre hanno implicazioni di natura economica, sociale e culturale per i territori interessati.

I movimenti migratori avvengono tra aree territoriali politiche o amministrative, i cui confini devono essere ben definiti. In genere si distinguono le migrazioni interne ai confini nazionali (ad esempio tra comuni, province o regioni) da quelle esterne, le cosiddette migrazioni internazionali.

Se da una parte l'immigrazione e l'emigrazione possono essere considerati dei fenomeni associabili rispettivamente a natalità e mortalità per quanto riguarda l'introduzione e l'esclusione di un soggetto da una popolazione, esistono delle sostanziali differenze. Natalità e mortalità sono infatti le determinanti endogene dei cambiamenti della struttura per età di una popolazione, mentre le migrazioni sono eventi non necessari e si inseriscono in un contesto più complesso che interessa talvolta popolazioni distinte con peculiarità e attitudini proprie.

Nella quasi totalità dei casi, immigrazioni ed emigrazioni hanno un ruolo più marginale in demografia rispetto alle loro controparti naturali, ma non per questo si può tralasciare la loro influenza, spesso significativa. In questo testo si andranno ad analizzare in particolare gli effetti delle migrazioni internazionali per cambiamento di residenza sulle principali dinamiche demografiche, quali fecondità, natalità e struttura per età di una popolazione. Per fare questo verranno utilizzati dei tassi e degli indicatori strutturali sintetici volti a racchiudere in sé degli aspetti specifici rilevanti per la descrizione di una popolazione. Si applicherà poi parte di quanto trattato a livello teorico, al caso delle province italiane, sfruttando l'utilizzo di alcuni semplici modelli di regressione lineare e lineare generalizzati. L'approccio che verrà utilizzato sarà sempre per contemporanei e non per coorte di nati. In conclusione si lasceranno delle critiche, riflessioni e possibili miglioramenti per coloro che andranno a studiare questo argomento in futuro.

Effetto delle migrazioni sui tassi e indici di struttura

Per trattare l'effetto delle migrazioni in demografia si è alla ricerca di alcune misure rappresentative che raccolgano in sé le caratteristiche quantitative relative alla composizione e allo sviluppo di una popolazione. Siamo quindi interessati a trovare degli indicatori relativi alla fecondità, natalità e all'età di un insieme circoscritto di individui. L'approccio che sta per essere presentato non si riferisce mai alla mortalità, nonostante ci possano essere dei risvolti in questo senso a causa delle migrazioni. Tale scelta è stata fatta per agevolare la trattazione e non dover considerare un periodo di riferimento troppo ampio. In seguito verranno indicati i tassi e gli indici di struttura che verranno studiati sia a livello teorico che applicativo. Dopodiché presenteremo dei possibili modelli per cercare di esplorare effettivamente i rapporti tra flussi in entrata e in uscita con i suddetti indicatori e trarre delle conclusioni.

2.1 Tassi e misure delle nascite

In questa sezione vengono presentati dei tassi e un indice relativi al fenomeno delle nascite. In particolare, per un'analisi per contemporanei, i prossimi indicatori vanno a cogliere gli effetti della situazione sociale ed economica di riferimento.

2.1.1 Tasso di fecondità totale

Il tasso di fecondità totale (TFT) è una misura dell'intensità del fenomeno di fecondità che non risente della struttura per età della popolazione. La fecondità è intesa come la manifestazione concreta e statisticamente misurabile della capacità di procreare. Essa si distingue dalla fertilità, in quanto quest'ultima è la capacità o attitudine di procreare e che può mantenersi di fatto soltanto in stato latente o potenziale per una parte della popolazione.

In quest'ambito il TFT si presenta come una delle misure complessive della fecondità e deriva dai tassi di fecondità specifici per età. La sua espressione matematica si presenta come segue:

$$TFT(t) = \sum_{x=15}^{49} f_x(t)$$

dove $f_x(t)$ rappresenta il tasso di fecondità specifico per età esprimibile in permillesimi come:

$$f_x(t) = \frac{N_x(t)}{{}_f\bar{P}_x(t)} * 1000 \quad , \quad \text{con } 15 \leq x \leq 49$$

dove $N_x(t)$ è il numero di nati vivi da donne di età x al tempo t (tra t e $t+1$) e ${}_f\bar{P}_x(t)$ è il numero medio di donne di età x al tempo t (tra t e $t+1$).

Da notare che il TFT viene misurato per la popolazione femminile di età compresa tra i 15 e i 49 anni, periodo tradizionalmente considerato dell'età fertile (vi possono essere però alcune eccezioni, per questo si parla di un'approssimazione). Il TFT ha oltretutto un importante significato, ovvero va ad indicare numericamente il numero medio di figli per donna a tutte le età per un determinato anno t . In un'analisi per contemporanei, il valore assunto da questo tasso è di più difficile interpretazione perchè risente del contesto storico e va a rispecchiare la tendenza di una popolazione di rimandare o meno la genitorialità per l'anno di riferimento. In un approccio per coorti invece, il TFT indica la capacità di una popolazione di riprodursi mantenendo più o meno costante la propria struttura demografica a parità di mortalità (e di altri fattori di minore rilevanza). Ad esempio, un TFT generazionale pari a 2,1 segnala che una coorte di nati riesce ad avere un perfetto ricambio generazionale in condizioni di mortalità assente.

2.1.2 Età media al parto

Non è un vero e proprio indice di struttura per età, ma è strettamente legato ad essi. Come il TFT, l'età media al parto (EMP) è una misura di sintesi della fecondità. La differenza sta nel fatto che, invece di essere uno strumento adatto a quantificare l'intensità del fenomeno, indica invece la sua cadenza. La sua espressione è:

$$EMP(t) = \frac{\sum_{x=15}^{49} x f_x(t)}{\sum_{x=15}^{49} f_x(t)}$$

Nell'analisi per contemporanei l'EMP esprime la distanza media che intercorre tra la coorte fittizia delle madri e la coorte fittizia dei figli in assenza di mortalità.

2.1.3 Tasso di natalità

Il tasso (generico) di natalità è una misura di frequenza delle nascite in un preciso periodo di tempo t . Esso si presenta come il rapporto tra il numero di nati vivi per una popolazione e l'ammontare medio della popolazione stessa al tempo t in permillesimi:

$$n(t) = \frac{N(t)}{\bar{P}(t)} * 1000$$

A differenza del TFT, il tasso di natalità è influenzato dalla struttura per età e per

genere, per questo risulta essere uno strumento d'analisi meno rigoroso del comportamento riproduttivo di una popolazione. Infatti una popolazione strutturalmente giovane tenderà a presentare tassi più elevati rispetto ad una anziana; analogamente, se una popolazione è caratterizzata da un elevato numero di donne nelle età più fertili, il tasso di natalità dovrebbe risultare in linea di massima più alto. Questo indicatore torna utile quando non è possibile misurare il TFT, non essendo quest'ultimo sempre disponibile. Si può utilizzare quindi per verificare il livello di sviluppo di una popolazione poichè per esso incidono fattori economici, sociali e politici.

2.2 Indici di struttura

Esistono molteplici tipi di indici di struttura di una popolazione: per età, per genere, per stato civile e per cittadinanza. Quelli che verranno trattati rientrano nella macroclasse degli indici di struttura per età, ovvero indici sintetici che dipendono fortemente dalla struttura per età dell'insieme di individui in analisi.

Sono presenti moltissimi esempi di indici di struttura per età, ma è stato deciso di trattarne solo una parte per cogliere gli aspetti specifici considerati più significativi per una popolazione.

2.2.1 Età media

L'età media è un indice di centralità esprimibile come media delle età ponderata con l'ammontare della popolazione a ciascuna età:

$$x_m = \frac{\sum_{i=0}^{w-1} (x_i + 0,5) P_i}{\sum_{i=0}^{w-1} P_i}$$

dove w è l'età limite.

Questo indice basilare restituisce una rapida idea sul fatto che una popolazione sia più o meno anziana. Oltretutto aiuta a capire approssimativamente in che fase della transizione demografica si trova la popolazione. Con transizione demografica si intende la teoria secondo la quale ogni popolazione si evolve allo stesso modo seguendo precise tappe evolutive. In particolare, i paesi non sviluppati avranno i tassi di natalità e mortalità elevati, mentre quelli sviluppati li avranno bassi. Semplicisticamente la transizione demografica prevede che, man mano che un paese si sviluppi economicamente, culturalmente e socialmente, tenderà sempre più verso un abbassamento di natalità e mortalità con un conseguente innalzamento dell'età media.

2.2.2 Indice di vecchiaia

L'indice di vecchiaia rientra nella classe degli indici di sostenibilità demografica ed economica. Si misura in percentuale come:

$$I_v = \frac{P_{65+}(t)}{P_{0-14}(t)} * 100$$

Sostanzialmente questo indice permette di determinare, anche se in modo molto generico, l'invecchiamento di una popolazione. Poichè viene espresso come rapporto tra la classe di età più anziana e quella più giovane, più il valore dell'indice è alto, più la popolazione risulterà anziana.

2.2.3 Indice di dipendenza strutturale

L'indice di dipendenza strutturale (o di carico sociale) appartiene alla stessa classe dell'indice di vecchiaia ed è espresso in percentuale come:

$$I_d = \frac{P_{0-14}(t) + P_{65+}(t)}{P_{15-64}(t)} * 100$$

Esso rappresenta il rapporto tra la popolazione di età non attiva e quella attiva. Da notare che questo indice può essere espresso come somma tra l'indice di dipendenza dei giovani I_{dg} e l'indice di dipendenza degli anziani I_{da} , con:

$$I_{dg} = \frac{P_{0-14}(t)}{P_{15-64}(t)} * 100 \quad e \quad I_{da} = \frac{P_{65+}(t)}{P_{15-64}(t)} * 100$$

Questa scomposizione della popolazione non attiva permette di separare l'ammontare dei giovani da quello degli anziani.

2.2.4 Indice di ricambio della popolazione attiva

L'indice di ricambio della popolazione attiva è, assieme all'indice di ricambio generazionale, un indice di ricambio. In percentuale si presenta come:

$$I_r = \frac{P_{60-64}(t)}{P_{15-19}(t)} * 100$$

Tale indice va ad indicare se la popolazione che sta per entrare a far parte dell'insieme degli individui attivi in ambito lavorativo riuscirà a soppiantare quella popolazione che sta per uscire dall'età lavorativa (e potrebbe ricevere la pensione). Un valore elevato indica che la popolazione che sta per entrare non riuscirà a sostituire demograficamente quella che sta per uscire.

2.3 Modelli per l'analisi

Per trattare gli effetti delle migrazioni sui tassi e indici di struttura ci possono essere una moltitudine di approcci. Si è deciso di procedere introducendo prima dei modelli di regressione lineare multipla (ols) e di regressione lineare generalizzati (glm) basati sull'effetto puntuale dei flussi; in secondo luogo viene introdotto un modello ols di tipo "first difference" che va a studiare l'effetto sulle differenze degli indicatori tra anni successivi. Infine è proposto un metodo per valutare la bontà di adattamento ai dati dei modelli in questione tramite R^2 e $R^2_{aggiustato}$ e un modo per confrontare i modelli tramite il criterio AIC.

2.3.1 Modello ols

Il primo modello che viene presentato è il modello di regressione lineare multipla normale con stime OLS, ovvero in cui si minimizza la somma dei quadrati dei residui. Si introducono due variabili confondenti, ovvero che limitano la capacità di determinare i tassi in base ad emigrazione e immigrazione. Ad esempio infatti, stabilire un nesso causale tra l'emigrazione e la fecondità è una sfida, perché i fattori che innescano l'emigrazione (ad esempio la disoccupazione) potrebbero anche causare cambiamenti simultanei nei modelli di fecondità, rendendo difficile identificare separatamente l'effetto dell'emigrazione dall'effetto delle condizioni economiche (Massimo Anelli, Nicoletta Balbo, 2018). La prima variabile introdotta è il tasso di disoccupazione provinciale, la quale si impone come fattore di repulsione per i migranti in entrata e favorisce l'emigrazione. L'altra variabile è invece il prodotto interno lordo (PIL) ai prezzi di mercato per abitante della regione di appartenenza della provincia. E' definito come il risultato finale dell'attività di produzione delle unità produttrici residenti. Corrisponde alla produzione totale di beni e servizi dell'economia, diminuita dei consumi intermedi ed aumentata dell'Iva gravante e delle imposte indirette sulle importazioni. È altresì, pari alla somma dei valori aggiunti ai prezzi di mercato delle varie branche di attività economica, aumentata dell'Iva e delle imposte indirette sulle importazioni, al netto dei servizi di intermediazione finanziaria indirettamente misurati (Glossario Istat, 2016). Chiaramente in termini pro capite deve essere tutto diviso per il numero di abitanti (in questo caso i residenti in regione). Questo risulta un fattore di attrazione per l'immigrazione e di resistenza per l'emigrazione.

Il modello si presenta dunque come segue:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \dots + \beta_{111} X_{111} + \epsilon_i, i = 1, \dots, n$$

con:

- Y_i : variabile risposta quantitativa continua che coincide con il tasso o indice di struttura calcolato tra il tempo t e $t+1$ (o ad inizio del tempo t per gli indici di struttura in quanto dati di stock) che si vuole ottenere tramite le

variabili esplicative X_j con $j=1,...,111$

- X_1 : emigrazioni internazionali per la provincia tra t-1 e t (possono essere espresse sia in termini assoluti che in termini di tasso di emigratorietà tra t-1 e t)
- X_2 : immigrazioni internazionali per la provincia tra t-1 e t (esprimibili sia in termini assoluti che come tasso di immigratorietà tra t-1 e t)
- X_3 : tempo t-1 espresso in anni
- X_4 : tasso di disoccupazione a livello provinciale all'inizio del tempo t-1 espresso in percentuale
- X_5 : pil pro capite per regione di appartenenza della provincia all'inizio del tempo t-1 espresso in euro
- $X_6, ..., X_{111}$: variabili dicotomiche che vanno ad esprimere il fattore provincia; X_j con $j=6,...,111$ vale 1 se viene osservata la j-esima provincia, 0 altrimenti, se valgono tutte 0 ci si riferirà alla provincia di "Agrigento" in quanto prima provincia per ordine alfabetico
- $\epsilon_i \sim N_n(0, \sigma^2 I_n)$ con $i=1,...,n$: errori indipendenti e identicamente distribuiti come una normale multivariata n-dimensionale con medie nulle e matrice di varianze e covarianze diagonale con valori pari a $\sigma^2 > 0$ (omoschedasticità)

Nel modello così espresso si valuterà anche la presenza o meno di un'interazione tra il tasso di disoccupazione e il pil pro capite regionale. Inoltre, si andrà a vedere se è possibile fare una selezione per le variabili non d'interesse (ad esclusione delle variabili relative al fattore provincia). Da notare che l'assunzione di base è che la variabile risposta segua una distribuzione normale del tipo $Y \sim N(\mu, \sigma^2)$, ipotesi che sarà valutata graficamente in fase di diagnostica dei modelli.

L'obiettivo del modello è quello di valutare l'effetto delle emigrazioni e immigrazioni per provincia sulla determinazione puntuale dei tassi e indici di struttura. Questo si può fare analizzando i valori di β_1 e β_2 separatamente al netto delle altre variabili esplicative.

2.3.2 Modello glm gamma

Questo modello è costruito sulla falsariga del precedente, ma vi sono importanti differenze che riguardano la teoria dei modelli lineari generalizzati. Si assume infatti che la distribuzione della variabile risposta appartenga ad un'ampia classe di funzioni. In particolare si ha che la densità di Y_i , $i=1,...,n$ appartiene a una famiglia di dispersione esponenziale univariata se è esprimibile nella forma:

$$p(y_i; \theta_i, \phi) = \exp\left\{\frac{\theta_i y_i - b(\theta_i)}{a_i(\phi)} + c(y_i, \phi)\right\}$$

con $y_i \in S \subseteq R, \theta_i \in \Theta \subseteq R, a_i(\phi) > 0$. Il parametro θ_i è detto parametro naturale, mentre ϕ è detto parametro di dispersione (Alessandra Salvan, Nicola Sartori, Luigi Pace, 2019).

Il caso che verrà trattato per questo problema prevede che la risposta segua una distribuzione gamma $Y \sim Ga(\alpha, \lambda_i)$, dove si può esprimere in termini di famiglia esponenziale come:

$$p(y_i; \lambda_i, \alpha) = \exp\{-\lambda_i y_i + \alpha \ln \lambda_i\} \frac{y_i^{\alpha-1}}{\Gamma(\alpha)}, \quad y_i, \lambda_i, \alpha > 0$$

in particolare risulta che il valore atteso è $E(Y_i) = \mu_i = \frac{\alpha}{\lambda_i}$, la varianza $Var(Y_i) = \frac{\mu_i^2}{\alpha}$, il parametro naturale $\theta_i = -\frac{1}{\mu_i} = -\frac{\lambda_i}{\alpha}$ e il parametro di dispersione $\phi = \frac{1}{\alpha}$.

Il modello che verrà quindi utilizzato espresso in termini di valore atteso sarà del tipo:

$$g(E[Y_i]) = g(\mu_i) = \eta_i = X_i \beta$$

dove $g(\mu_i)$ è la funzione legame e η_i il predittore lineare.

Sia la variabile risposta che le esplicative rimangono inalterate rispetto al modello ols normale; a cambiare saranno invece i parametri β_i che verranno stimati attraverso le stime ai minimi quadrati pesati iterati. Infatti, oltre al fatto che ora la variabile risposta segue una distribuzione gamma e non normale, si può decidere di usare una diversa funzione legame. Nell'analisi trattata si partirà con la funzione legame canonica per la gamma, ovvero la funzione inversa, dopodichè si valuterà l'eventuale utilizzo di una funzione legame alternativa tra cui l'identità e la logaritmica. L'obiettivo del modello è invariato rispetto a quello del modello introdotto precedentemente.

2.3.3 Modello "first difference"

Questo tipo di modello è costruito a partire dalle esplicative espresse come differenza tra il valore assunto al tempo t (cioè nel periodo tra $t-1$ e t) e al tempo precedente $t-1$ (tra $t-1$ e $t-2$). L'unica eccezione è rappresentata dall'anno (variabile X_3) e dall'insieme delle variabili dicotomiche che esplicitano il fattore provincia, le quali difatti non cambiano di espressione né di significato. La variabile risposta è invece calcolata come differenza tra il valore assunto in $t+1$ e t , in ottica di causalità.

Il modello è:

$$Y_i = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \dots + \beta_{111} X_{111} + \epsilon_i, i = 1, \dots, n$$

con:

- Y_i : variabile risposta quantitativa continua che coincide con la differenza tra il tasso o indice di struttura calcolato tra il tempo t e $t+1$ (o a inizio di $t+1$) con quello calcolato tra $t-1$ e t (o a inizio di t)
- X_1 : differenza di emigrazioni internazionali per la provincia misurate tra $t-1$ e t con quelle tra $t-2$ e $t-1$
- X_2 : differenza di immigrazioni internazionali per la provincia misurate tra $t-1$ e t con quelle tra $t-2$ e $t-1$
- X_3 : tempo $t-1$ espresso in anni
- X_4 : differenza del tasso di disoccupazione a livello provinciale espresso in percentuale con quello dell'anno precedente (quindi tra $t-1$ e t)
- X_5 : differenza di pil pro capite per regione di appartenenza della provincia espresso in euro con quello dell'anno precedente (quindi tra $t-1$ e t)
- X_6, \dots, X_{111} : variabili dicotomiche che vanno ad esprimere il fattore provincia; X_j con $j=6, \dots, 111$ vale 1 se viene osservata la j -esima provincia, 0 altrimenti, se valgono tutte 0 ci si riferirà alla provincia di "Agrigento" in quanto prima provincia per ordine alfabetico
- $\epsilon_i \sim N_n(0, \sigma^2 I_n)$ con $i=1, \dots, n$: errori indipendenti e identicamente distribuiti come una normale multivariata n -dimensionale con medie nulle e matrice di varianze e covarianze diagonale con valori pari a $\sigma^2 > 0$ (omoschedasticità)

Le assunzioni sottostanti il modello e il metodo di stima sono uguali a quelle del modello ols normale.

L'obiettivo del modello consiste nel cercare di esprimere gli effetti dei flussi migratori sulla variazione del valore del tasso/indice di struttura per due anni consecutivi.

2.4 Adattamento del modello

Per il modello di regressione lineare multipla normale e per quello normale di tipo first difference, una prima misura della bontà di adattamento ai dati osservati è rappresentata dall'indice di determinazione R^2 , così costruito:

$$R^2 = \frac{SQR_p}{SQT} = 1 - \frac{SQE_p}{SQT}$$

dove SQR_p rappresenta la devianza spiegata dal modello, SQE_p la devianza residua e SQT la devianza totale.

Questo indice si presenta come proporzione tra la variabilità dei dati e la correttezza del modello statistico. Una misura di bontà di adattamento che tenga conto anche del principio di parsimonia è invece l' $R^2_{aggiustato}$, definibile come:

$$R^2_{aggiustato} = 1 - \frac{\frac{SQE_p}{(n-p)}}{\frac{SQT}{(n-1)}} = R^2 - (1 - R^2) \frac{p-1}{n-p}$$

dove n è il numero di osservazioni e p il numero di parametri compresi nel modello. Si nota quindi che è presente una penalizzazione data dal numero di parametri, la quale permette, a parità di devianze, di preferire il modello più parsimonioso.

Un altro metodo per valutare la qualità di un modello è quello di verificare in modo specifico la significatività di ciascun parametro tramite la statistica t di student e, più generalmente, tramite una statistica F di Fisher-Snedecor che valuti la significatività complessiva del modello rispetto a quello con sola intercetta. Queste si possono scrivere come:

$$t = \frac{\hat{\beta}_j}{\sqrt{Var[\hat{\beta}_j]}}, j = 1, \dots, p \quad e \quad F = \frac{\frac{(SQR_p - SQR_{p_0})}{p-p_0}}{\frac{SQE_p}{n-p}}$$

dove SQR_{p_0} è la devianza spiegata dal modello con sola intercetta.

Alternativamente, un metodo grafico è invece quello basato sullo studio dei residui, residui standard ed eventuali outliers identificabili mediante distanze di Cook.

Infine, in caso di indecisione sulla scelta tra più modelli, si farà ricorso al criterio di informazione di Akaike (AIC) che si presenta come segue:

$$AIC(F_d) = 2d - 2l(\hat{\theta}^{(d)}; y)$$

dove d è il numero di parametri e $l(\hat{\theta}^{(d)}; y)$ è la funzione di log-verosimiglianza associata al modello.

Questo criterio per modelli annidati si basa su penalizzazioni della log-verosimiglianza. Si tiene quindi conto sia del principio di bontà di adattamento che del principio di parsimonia. Ad essere selezionato sarà il modello che presenta il minimo AIC.

Effetto delle migrazioni sulla struttura per età

La precedente trattazione inerente l'effetto delle migrazioni sugli indici di struttura discussi costituiva un modo sintetico per valutare le conseguenze sulla struttura per età di una popolazione. In questo capitolo si è interessati invece a valutare un possibile metodo per analizzare l'effetto dei flussi migratori internazionali sull'intera struttura per età, andando quindi a visualizzare un quadro più completo ed esaustivo sull'argomento in questione. Questo metodo non verrà applicato nella parte analitica e viene inteso come modo alternativo per affrontare il tema degli effetti migratori sulla struttura per età.

L'effetto delle migrazioni può essere sia di tipo diretto che indiretto. Il primo comporta dei risultati istantanei sull'ammontare e sulla struttura per età sia per quanto riguarda la popolazione dalla quale è partito il flusso sia per quella ricevente. L'effetto indiretto ha invece conseguenze di lungo periodo perchè ci sono riflessi sul numero complessivo di nascite future e, a differenza del precedente, dipende anche dalla propensione dei migranti ad avere figli. Ad esempio, se il migrante è una donna in età fertile, potrà avere figli nel paese di destinazione; in caso contrario, se ha superato l'età fertile, non andrà più a generare figli nel paese ospitante e gli eventuali figli pregressi verranno conteggiati soltanto nel paese d'origine (salvo anche il loro trasferimento col genitore).

La componente migratoria risulta di conseguenza una parte fondamentale da tenere in considerazione per determinare l'invecchiamento o il ringiovanimento di una popolazione in un dato periodo storico. Difatti coloro che migrano possono appartenere a fasce di età più o meno differenti e omogenee e questo porta a determinate conseguenze specifiche. Se ad esempio a trasferirsi sono per lo più i giovani, allora questo indurrà ad un invecchiamento della popolazione di partenza dei migranti e, al contrario, ad un ringiovanimento della popolazione di destinazione; in termini di struttura per età questo si traduce, per la popolazione di partenza, in una maggiore concentrazione di individui nelle classi di età più elevate, a discapito delle classi più basse, viceversa avviene per la popolazione di destinazione.

3.1 Estensione del modello dei tassi di crescita specifici per età

Il modello che viene utilizzato si basa sulla decomposizione del tasso di crescita istantaneo per l'ammontare del numero di individui di età x tra il tempo t e $t+n$ del tipo:

$$r_x(t, t+n) = \frac{1}{n} \ln \left(\frac{P_x(t+n)}{P_x(t)} \right)$$

dove $P_x(t)$ è la popolazione di età x al tempo t .

La scelta di questo tipo di tasso prevede che i nuovi individui influenzano l'ammontare della popolazione a partire dall'istante immediatamente successivo al loro ingresso, dato dal fatto che l'intervallo di tempo in questione viene suddiviso in sotto-intervalli infinitesimi.

Le componenti che vanno a descrivere il nuovo tasso di crescita così trovato sono legate ai quattro fattori fondamentali che fanno variare la dimensione numerica di una popolazione in un intervallo di tempo: natalità, mortalità, immigrazione ed emigrazione. Basandoci quindi sul modello dei tassi specifici per età con l'introduzione della doppia direzione dei flussi migratori (Agnieszka Fihel, Anna Janicka, Weronika Kloc-Nowak, 2018), si generalizza il modello al caso tra i tempi t e $t+n$:

$$r_x(t, t+n) = \frac{1}{n} \left(\ln \frac{N(t-x+n)}{N(t-x)} + \ln \frac{S_x(t+n)}{S_x(t)} + \ln \frac{E_x(t+n)}{E_x(t)} + \ln \frac{I_x(t+n)}{I_x(t)} \right)$$

dove abbiamo in ordine le seguenti componenti:

- 1° componente: riferita alla natalità, con $N(t-x)$ che rappresenta il numero dei nati per l'anno che termina a $t-x$ (quindi durante $t-x-1$), ovvero i nati della coorte con età x
- 2° componente: riferita alla mortalità, con $S_x(t)$ che indica la frazione degli individui della coorte di nati di età x che sono sopravvissuti fino a t
- 3° componente: riferita all'emigratorietà, con $E_x(t)$ che indica il fattore degli individui della coorte di nati di età x per cui è cambiata la dimensione della popolazione per emigrazione fino a t tra quelle sopravvissute fino a t
- 4° componente: riferita all'immigratorietà, con $I_x(t)$ che indica il fattore degli individui della coorte di nati di età x per cui è cambiata la dimensione della popolazione per immigrazione fino a t tra quelle sopravvissute e non emigrate fino a t

E' da notare il fatto che la terza componente sia misurata al netto della seconda, mentre l'ultima componente è calcolata sia al netto della terza che conseguentemente della seconda.

Numericamente $E_x(t)$ è sempre minore o uguale a 1, con ad esempio un valore di 0,8 che va a significare che l'80% dei sopravvissuti fino all'anno t non sono emigrati (il 20% è invece emigrato all'estero), mentre un valore pari a 1 implica nessuna emigrazione. Nel caso invece di $I_x(t)$, tale fattore risulta sempre maggiore o uguale a 1, con un valore di ad esempio 1,2 che denota un aumento per immigrazione del contingente della popolazione del 20% al netto della mortalità e dell'emigrazione, mentre il valore estremo 1 indica l'assenza di immigrazione.

A questo punto si può decidere di trattare l'emigrazione di ritorno come un fenomeno separato oppure come annullamento dell'effetto migratorio. Nel primo caso si avrà che le componenti relative all'emigratorietà e all'immigratorietà (quindi sia il fattore $E_x(t)$ che $I_x(t)$) rifletteranno un aumento dell'emigrazione e dell'immigrazione, in quanto il ritorno nel luogo d'origine verrà considerato una vera e propria migrazione. In caso contrario, l'emigrazione di ritorno farà in modo che lo spostamento dei migranti non risulti né nella componente di emigratorietà né in quella di immigratorietà. Un altro aspetto importante è simboleggiato dalla componente immigratoria, dal momento che essa va ad incorporare anche la componente di mortalità riferita agli immigrati (in questo caso $S_x(t)$ non viene dunque influenzato).

Poichè la terza e la quarta componente esplicitano soltanto l'effetto diretto sul tasso di crescita, si vuole successivamente andare a comprendere come può essere trattato l'effetto indiretto.

3.2 Effetto diretto e indiretto delle migrazioni

Si vuole comprendere come e in che misura le emigrazioni e le immigrazioni possano influire sulle nascite future. Le emigrazioni portano infatti, non solo ad una diminuzione dell'ammontare della popolazione al momento dello spostamento, ma anche ad una ulteriore riduzione data dai possibili nascituri. Al contrario le immigrazioni vanno ad aumentare la popolazione residente anche per i futuri nati che si andranno ad aggiungere a causa del trasferimento della coorte immigrata. Per fare questo, si introduce innanzitutto il tasso lordo di riproduttività esprimibile approssimativamente anche in termini di tasso di fecondità totale:

$$R(t) = \sum_{x=15}^{49} \frac{{}_fN(t)}{{}_f\bar{P}_x(t)} \simeq \frac{{}_fN(t)}{{}_fN(t) + {}_mN(t)} TFT(t) \simeq 0,485TFT(t)$$

dove ${}_fN(t)$ esprime il numero di nati femmine al tempo t e ${}_mN(t)$ quello dei maschi al tempo t, invece ${}_f\bar{P}_x(t)$ rappresenta l'ammontare medio della popolazione femminile di età x durante l'anno tra t e t+1.

Questo tasso va ad indicare il numero medio di figlie femmine nate da una donna al tempo t in assenza di mortalità; esprime anche il rapporto tra il totale delle figlie ad età 0 e quello delle madri all'età 0. L'utilizzo di $R(t)$ invece del $TFT(t)$ ha il vantaggio di non dovere assumere che i tassi di mortalità e di migratorietà siano

uguali per genere e che il numero di nati femmine sia uguale al numero di nati maschi. Le relazioni successivamente presentate faranno riferimento alla sola popolazione femminile; se volessimo invece riferirci alla popolazione maschile dovremmo costruire un tasso come somma a tutte le età fertili femminili (tradizionalmente 15-49) dei rapporti tra il numero di nati maschi sul totale dei nati.

Si procede poi con la scomposizione del numero totale dei nati femmine da donne in età compresa tra $[\alpha, \beta]$ al tempo t nelle quattro componenti viste precedentemente:

$$N(t) = R(t) \sum_{x=\alpha}^{\beta} N(t-x) f_{0x}(t) S_x(t) E_x(t) I_x(t)$$

dove l'aggiunta di $f_{0x}(t)$ indica la distribuzione delle nascite per età delle madri al tempo t . In particolare si suppone che il tasso di fecondità specifico per età può essere espresso in permillesimi come:

$$f_x(t) = \frac{N_x(t)}{{}_f\bar{P}_x(t)} * 1000 = TFT(t) f_{0x}(t)$$

mentre per la sola popolazione di nate femmine si ha:

$$f_x^*(t) = \frac{{}_fN_x(t)}{{}_f\bar{P}_x(t)} * 1000 = R(t) f_{0x}(t)$$

Volendo ora distinguere i nati persi per emigrazione e i nati aggiunti per immigrazione, si procede a partire dalla scomposizione del numero totale di nati per scindere le due distinte componenti migratorie. Sotto l'ipotesi di assenza di immigrazione a tutte le classi di età e di uguali tassi di fecondità specifici per età degli emigrati rispetto ai non emigrati, si può esprimere il contingente di nati femmine persi da donne di età $[\alpha, \beta]$ nell'anno t per emigrazione, nel modo seguente:

$$N_E(t) = R(t) \sum_{x=\alpha}^{\beta} N(t-x) f_{0x}(t) S_x(t) (1 - E_x(t))$$

ovvero come differenza tra il caso di non emigrazione e quello di emigrazione.

Altrimenti, mantenendo solo l'ipotesi di uguali tassi di fecondità specifici per età degli emigrati rispetto ai non emigrati, si ha:

$$N_E(t) = R(t) \sum_{x=\alpha}^{\beta} E_x(t) f_{0x}(t)$$

Per quanto riguarda l'immigrazione, si ha che il numero di nati femmine da donne di età $[\alpha, \beta]$ all'anno t è pari a:

$$N_I(t) = R(t) \sum_{x=\alpha}^{\beta} I_x(t) f_{0xI}(t)$$

dove $f_{0xI}(t)$ indica la distribuzione delle nascite per età delle madri al tempo t degli immigrati.

Si vuole quindi separare gli effetti delle migrazioni diretti da quelli indiretti. Seguendo la logica precedente, in assenza di immigrazione e in presenza di uguali tassi di fecondità degli emigrati rispetto ai non emigrati, si ha che la differenza totale del numero di persone in età x al tempo t tra il caso di non emigrazione e di emigrazione è:

$$\Delta P_E(x, t) = N(t - x)S_x(t)(1 - E_x(t)) + N_E(t - x)S_x(t)E_x(t)$$

dove $N(t - x)S_x(t)(1 - E_x(t))$ rappresenta l'effetto diretto dell'emigrazione, mentre $N_E(t - x)S_x(t)E_x(t)$ l'effetto indiretto dovuto alle nascite perse.

Per valutare d'altro canto l'effetto dell'immigrazione, si assume soltanto che i nati da immigrati abbiano uguali tassi di emigratorietà e mortalità del resto della popolazione. La differenza totale del numero di persone in età x al tempo t tra il caso di immigrazione e di non immigrazione, si esprime allora come:

$$\Delta P_I(x, t) = I_x(t) + N_I(t - x)S_x(t)E_x(t)$$

dove $I_x(t)$ è l'effetto diretto e $N_I(t - x)S_x(t)E_x(t)$ quello indiretto per l'immigrazione.

Analisi delle province italiane

Si vuole ora studiare il fenomeno delle migrazioni applicato alle province italiane dal solo punto di vista degli effetti sui tassi e indici di struttura come teorizzato nel capitolo 2. Si introdurrà l'argomento partendo dal contesto storico dell'emigrazione e dell'immigrazione in Italia, con particolare riferimento alle dinamiche degli ultimi decenni. Lo studio in questione fa riferimento al periodo 2004-2018 e i dati utilizzati sono stati presi e rielaborati a partire dai dati pubblici dell'Istat.

Di seguito, quando si parlerà col generico termine di "province", ci si riferirà alle 107 divisioni territoriali al 2020 di:

- 80 province
- 14 città metropolitane
- 1 ex provincia della Valle d'Aosta
- 2 province autonome di Trento e Bolzano
- 4 enti di decentramento regionali del Friuli-Venezia Giulia
- 6 liberi consorzi comunali della Sicilia

L'unica eccezione è rappresentata dalla provincia del "Sud Sardegna", qui considerata come composta dalle sole ex province di "Medio Campidano" e "Carbonia-Iglesias", escludendo i territori dell'ex provincia di "Cagliari" che sono stati accorpati in essa. Di conseguenza al posto dell'ex città metropolitana di "Cagliari" verrà considerata l'omonima ex provincia. Questa scelta è stata fatta per creare una maggiore coerenza temporale per i vari cambiamenti territoriali avvenuti in quegli anni.

Per misurare i fenomeni migratori si è scelto di utilizzare i tassi di migratorietà, invece delle migrazioni in valori assoluti. Questo modo di procedere permette di tenere conto della popolazione residente ai vari anni per le province d'interesse. Di conseguenza, a parità di migrazioni in termini assoluti, avranno maggiore impatto le migrazioni per la provincia con meno abitanti. In particolare, troviamo il tasso di emigratorietà e immigratorietà in permillesimi così definiti:

$$e(t) = \frac{E(t)}{P(t)} * 1000 \quad , \quad i(t) = \frac{I(t)}{P(t)} * 1000$$

Come anticipato nella parte teorica, i due tassi di migratorietà utilizzati saranno calcolati tra $t-1$ e t in un'ottica di causalità e non tra t e $t+1$.

4.1 Contesto storico dell'emigrazione italiana

L'organo di riferimento che raccoglie i dati sulle emigrazioni per cambio di residenza è l'Anagrafe degli italiani residenti all'estero (AIRE), istituito nel 1988, ma che negli ultimi anni ha iniziato a confluire nell'Anagrafe nazionale della popolazione residente (ANPR) gestito direttamente dal Ministero dell'Interno. Questa scelta è stata fatta nell'intento di creare un'unica banca dati centralizzata che sostituisse le Anagrafi della popolazione residente dei comuni (APR) e l'AIRE. Infatti questo nuovo sistema permette di eliminare gli errori di comunicazione tra gli enti comunali, automatizzando il processo e introducendo degli standard validi a livello nazionale. Storicamente l'Italia ha avuto due grandi periodi di emigrazione. Il primo si colloca tra la nascita del Regno d'Italia e lo scoppio della Prima Guerra Mondiale, mentre il secondo coincide con il secondo dopoguerra. Successivamente, secondo le statistiche ufficiali, l'Italia ha avuto una migrazione netta positiva dal 1973, trasformandosi così in un "paese di immigrazione", dopo un lungo periodo in cui il suo status era di "paese di emigrazione". Questa svolta nel saldo migratorio italiano non deve però far pensare che, anche dopo gli anni settanta, non ci siano stati numeri relativamente significativi di persone che hanno lasciato il paese (Jean-Michel Laffleur, Mikolaj Stanek, 2017). Come si può osservare dalla tabella sottostante, nel periodo sotto studio c'è stato, salvo alcune eccezioni, un continuo aumento del fenomeno emigratorio:

anno	Nord-Ovest	Nord-Est	Centro	Sud	Isole	Italia
2004	11.504	8.982	6.897	14.065	8.462	49.910
2005	12.203	10.597	9.006	14.061	8.064	53.931
2006	12.881	11.939	11.069	14.265	8.253	58.407
2007	14.396	13.456	9.758	8.088	5.415	51.113
2008	17.352	16.845	11.242	10.265	5.967	61.671
2009	18.795	18.877	12.214	9.653	5.382	64.921
2010	20.892	20.021	13.050	8.755	4.783	67.501
2011	26.163	21.964	15.757	11.770	6.807	82.461
2012	34.115	27.885	19.534	15.693	8.989	106.216
2013	40.454	31.103	23.806	19.844	10.528	125.735
2014	42.927	32.494	25.940	22.127	12.840	136.328
2015	46.088	34.756	27.957	23.520	14.634	146.955
2016	47.954	36.983	30.500	26.444	15.184	157.065
2017	47.118	36.234	29.255	26.553	15.950	155.110
2018	47.323	37.365	30.090	26.094	16.088	156.960

Tab. 4.1: Emigrazioni dall'Italia dal 2004 al 2018, dati Istat

Si osserva quindi che i flussi in uscita dal 2004 al 2018 sono più che triplicati, passando da 49.910 a 156.960. Guardando da vicino i dati si può notare che l'aumento

del fenomeno è iniziato dal periodo successivo al 2007 con un repentino sviluppo a partire dal 2011 per assestarsi dal 2016 in poi. Questo è dovuto principalmente a causa della grave crisi economica mondiale avvenuta tra il 2007 e il 2013. Il grande aumento successivo al 2010 è da accostare agli effetti di ritardo dovuti alla crisi stessa. Negli ultimi anni le motivazioni che stanno spingendo parte della popolazione, soprattutto nelle fasce di età più giovani, ad emigrare dall'Italia, riguardano principalmente l'ambito lavorativo. Se si osserva la situazione per le cinque macroaree geografiche, si nota che l'andamento degli storici delle emigrazioni è simile. Il Nord-ovest risulta in ogni caso l'area di maggior diffusione del fenomeno, è da considerare però il fatto che sia la zona con più abitanti. D'altra parte, nonostante abbiano un ammontare di popolazione comparabile, il Nord-Est ha maggiori flussi emigratori rispetto al Centro e al Sud.

4.2 Contesto storico dell'immigrazione in Italia

Nella storia italiana il fenomeno immigratorio è stato poco rilevante fino agli anni settanta. Dopodichè da quegli anni in poi l'immigrazione in Italia ha iniziato ad aumentare drasticamente e ad assumere particolare importanza. Esistono varie cause di questo fenomeno. Il primo fattore, o insieme di fattori, è incentrato sulla facilità di ingresso in Italia e sulla sua funzione evolutiva di "destinazione sostitutiva" per i migranti esclusi dalla chiusura dei tradizionali paesi di immigrazione del Nord-Ovest dell'Europa nella seconda metà degli anni settanta (Russell King, 2014). Più in generale il motore trainante del fenomeno è stata la crescente prosperità del paese iniziata a partire da quegli anni.

Nella tabella qui sotto sono riportati i numeri dei flussi di immigrazione in Italia nel periodo più recente e che viene trattato nell'analisi:

anno	Nord-Ovest	Nord-Est	Centro	Sud	Isole	Italia
2004	151.204	98.119	96.808	52.848	15.901	414.880
2005	102.208	79.846	73.062	35.408	14.436	304.960
2006	92.339	74.357	65.940	32.998	14.080	279.714
2007	158.271	128.249	133.604	75.940	31.059	527.123
2008	145.834	130.565	127.476	62.810	27.709	494.394
2009	132.964	102.596	109.134	53.994	23.171	421.859
2010	145.910	106.573	107.198	62.534	25.529	447.744
2011	126.551	85.526	94.915	56.631	22.170	385.793
2012	107.944	73.572	95.555	52.630	21.071	350.772
2013	97.291	65.697	74.105	50.019	20.342	307.454
2014	83.958	57.484	68.106	47.338	20.745	277.631
2015	80.485	58.142	67.219	51.899	22.333	280.078
2016	89.536	64.850	70.218	54.034	22.185	300.823
2017	103.572	76.751	74.288	63.848	24.981	343.440
2018	99.674	78.277	70.005	59.415	24.953	332.324

Tab. 4.2: Immigrazioni in Italia dal 2004 al 2018, dati Istat

Si può subito notare che l’immigrazione dal 2004 al 2018 sia stato un fenomeno molto più consistente rispetto all’emigrazione, ma complessivamente i flussi sono diminuiti da 414.880 a 332.324. Inoltre risulta evidente un andamento decisamente più discontinuo in termini numerici. Al contrario dell’emigrazione, negli anni precedenti il 2007, i flussi immigratori erano in calo. Successivamente, dopo il grande boom di immigrazione del 2007, è tornato a calare fino al 2014, ma attestandosi comunque su livelli più elevati rispetto al 2006. Si constata dunque che la crisi economica sia stata motivo di attrazione verso il paese. Analizzando il comportamento per macroaree si nota anche in questo caso che il Nord-Ovest ha i maggiori flussi in entrata. A seguire il Nord-Est e il Centro che si sono invece alternati di posizione nel corso del tempo, mentre ora si assestano a livelli simili. Sud e Isole rimangono le aree meno caratterizzate dall’immigrazione.

4.3 Dataset e dati mancanti

Il dataset principale è composto da $n=1498$ osservazioni (107 province per 14 anni) relative alle province ai vari anni. Questo è dovuto al fatto che, dovendo mettere assieme dati di stock con dati di flusso in cui le esplicative e la risposta sono valutate ad anni l’uno successivo all’altro, un anno viene tagliato. Più precisamente i dati per le esplicative andranno dall’inizio del 2004 fino a fine 2017, mentre i dati per la risposta partiranno da inizio 2005 e termineranno a fine 2018. Per motivi analoghi, il dataset utilizzato per il modello first difference, dovendo considerare le differenze per anni consecutivi, è costituito da $n=1284$ osservazioni.

Volendo ottenere i dati per le province italiane dal 2004 al 2018 come descritto precedentemente, si deve tenere conto dei cambiamenti amministrativi e delle modalità di raccolta dei dati nel corso del periodo di riferimento. In quest’ottica, a partire dai dati dell’Istat per province, si vanno a ricostruire, e talvolta ad approssimare, i dati per le province attuali anche per gli anni in cui non erano ancora state istituite oppure quando i confini territoriali risultavano differenti.

Un primo cambiamento delle province italiane è avvenuto nel maggio del 2004 con la nascita delle province di "Monza e della Brianza", "Fermo" e "Barletta-Andria-Trani", scorporatesi rispettivamente dalle province di "Milano", "Ascoli Piceno" e "Bari". Tali province sono divenute effettivamente operative soltanto nel giugno del 2009 con l’elezione del primo consiglio provinciale (Unione Province d’Italia, 2004). L’altra grande variazione delle province è avvenuta con l’approvazione del referendum in Sardegna del maggio 2012, dove le vecchie province di "Olbia-Tempio" e "Ogliastra" sono state accorpate rispettivamente alle province di "Sassari" e "Nuoro" e l’istituzione della provincia del "Sud Sardegna" ha preso il posto delle province del "Medio Campidano" e di "Carbonia-Iglesias", oltre che accorpate i comuni della provincia di "Cagliari", quest’ultima divenuta città metropolitana. Anche in questo caso la concreta operatività delle nuove province è stata raggiunta solo molti anni dopo, nel febbraio del 2016 (Davide Batzella, 2017).

Il passaggio di rilevazione dei dati Istat dalle vecchie alle nuove province ricalca dal punto di vista territoriale e temporale le riforme appena descritte.

Per le province di "Monza e della Brianza", "Fermo" e "Barletta-Andria-Trani" non sono presenti i dati relativi alle migrazioni e al tasso di disoccupazione per gli anni

che vanno dal 2004 al 2009. Per il primo caso, ad ogni anno si vanno a ricavare i dati in maniera tale che le migrazioni siano assegnate alla provincia attuale e alla provincia da scorporare proporzionalmente al rispettivo numero di individui residenti per quell'anno. Questo modo di procedere si basa sull'assunzione che le immigrazioni e le emigrazioni avvengano con la stessa intensità su tutto il territorio interessato. Per il tasso di disoccupazione si assegna invece alla provincia attuale il tasso di quella a cui apparteneva prima dello scorporamento.

Per le province sarde, i dati mancanti sono relativi al periodo 2004-2016. Le migrazioni per la provincia di "Sassari" e "Nuoro" sono il risultato della somma delle migrazioni tra le province pre-riforma situate nelle aree delle province attuali. Per ogni tasso ed indice di struttura, si assegna a "Sassari" e a "Nuoro" la media ponderata dei tassi/indici delle province pre-riforma coinvolte con pesi pari ai corrispondenti ammontare della popolazione residente.

Per la provincia del "Sud Sardegna" si è deciso di procedere come appena descritto per le altre province sarde, ad eccezione del fatto che i comuni dell'ex provincia di "Cagliari" sono stati considerati come appartenenti alla nuova area metropolitana dell'omonima città. Per questo motivo i dati del 2017 e del 2018 per il "Sud Sardegna" e "Cagliari" sono stati ricalcolati coerentemente con questa scelta metodologica.

4.4 Analisi Esplorativa

L'analisi esplorativa è fondamentale per avere una prima idea sui dati e sulle variabili che si vanno a prendere in considerazione. Si tratta infatti di una fase preliminare all'analisi principale, tramite la quale si individua la distribuzione empirica dei dati e i rapporti che caratterizzano le variabili. Le analisi sono svolte tenendo conto di un livello di significatività fissato al 5%. Gli strumenti che vengono introdotti sono principalmente grafici e statistiche di sintesi, accompagnati da opportuni test. Per questo tipo di analisi vengono escluse le due variabili confondenti (tasso di disoccupazione e pil regionale), perchè non si hanno risvolti determinanti nell'applicazione dei modelli.

4.4.1 Analisi univariata

Nel dataset, ad esclusione delle variabili associate al fattore provincia, sono presenti solo variabili quantitative continue. Si condurranno quindi delle analisi grafiche tramite boxplot, istogrammi e q-q plot (grafici quantile quantile). Per verificare l'ipotesi di normalità si farà uso del test di Shapiro-Wilk. E' importante tenere conto del fatto che il test di normalità può essere influenzato dalla presenza di outliers, quindi non si esclude a priori un modello di regressione che preveda la gaussianità della variabile risposta.

Tasso di emigratorietà

Il tasso di emigratorietà varia da un minimo di 0,177 ad un massimo di 5,167, con frequenze assolute molto elevate da 0,5 a 1. La media è 1,612 e la mediana 1,433 con

varianza corretta stimata pari a 0,849. Dal boxplot e dall'istogramma si evince che sia presente una evidente asimmetria positiva, confermata dal fatto che la media sia superiore alla mediana. Il q-q plot suggerisce che non sia possibile assumere normalità, ipotesi verificata anche dal test di Shapiro-Wilk ($p\text{-value} < 2.2e-16$).

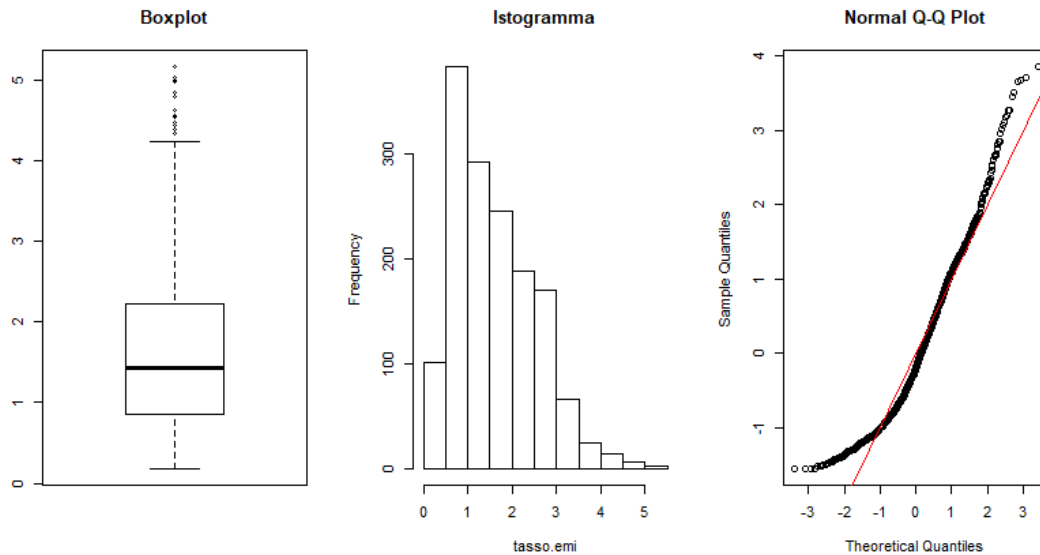


Fig. 4.1: Boxplot, istogramma e Q-Q plot per il tasso di emigratorietà

Tasso di immigratorietà

Il tasso di immigratorietà varia da un minimo di 1,047 ad un massimo di 16,398. La media è pari a 5,998 e la mediana 5,569 con varianza corretta stimata di 7,55 (molto superiore al tasso di emigratorietà). Come nel caso del tasso di emigratorietà vi è una forte asimmetria positiva, data anche dal fatto che sono presenti numeri outliers sulla coda destra. Il test di Shapiro-Wilk rifiuta ampiamente l'ipotesi di normalità ($p\text{-value} < 2.2e-16$).

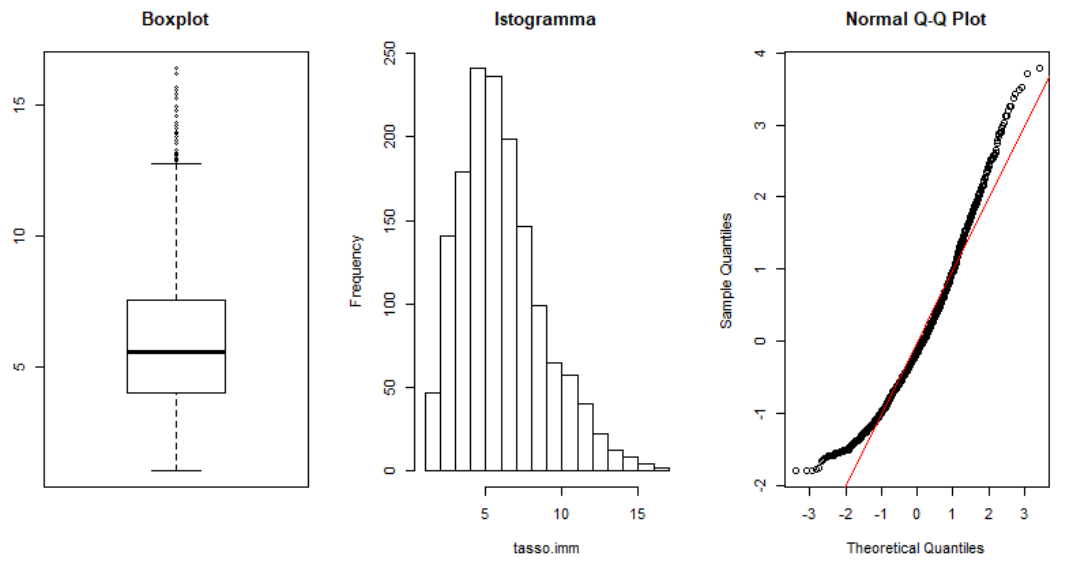


Fig. 4.2: Boxplot, istogramma e Q-Q plot per il tasso di immigratorietà

Tasso di fecondità totale

Il TFT ha un valore minimo di 0,93 e un massimo di 1,76. La media di 1,345 si discosta poco dalla mediana di 1,35. Graficamente non si notano particolari asimmetrie, sono però presenti evidenti outliers su entrambe le code. Il test di Shapiro-Wilk rifiuta di poco l'ipotesi di normalità (p-value=0.002296).

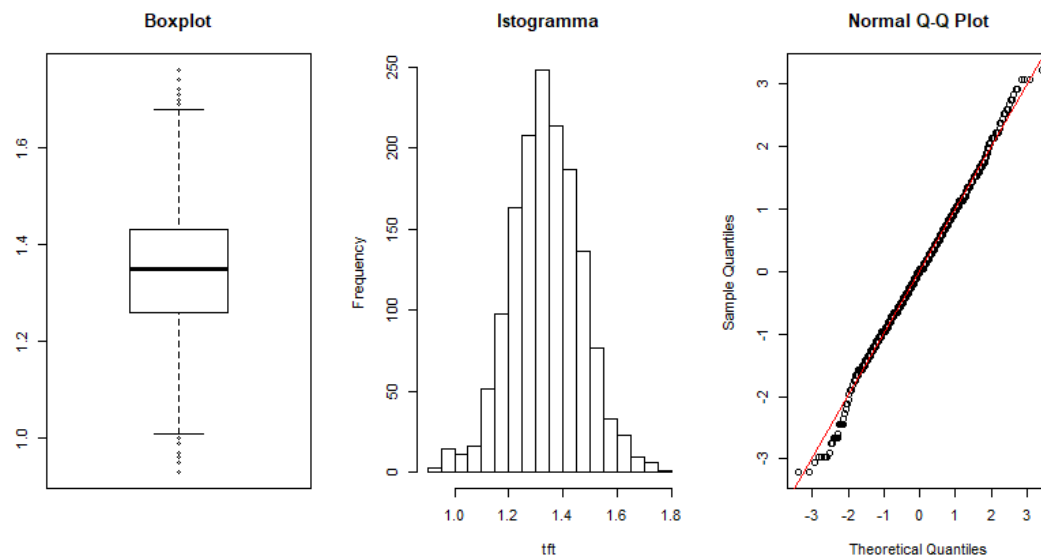


Fig. 4.3: Boxplot, istogramma e Q-Q plot per il tasso di fecondità totale

Età media al parto

L'EMP ha valori tra 29,64 e 33,02 con media uguale alla mediana (31,4). Non appare evidente alcuna asimmetria nella distribuzione. L'ipotesi di normalità è rifiutata al 5% ($p\text{-value}=0.02687$), ma sarebbe accettata all'1%.

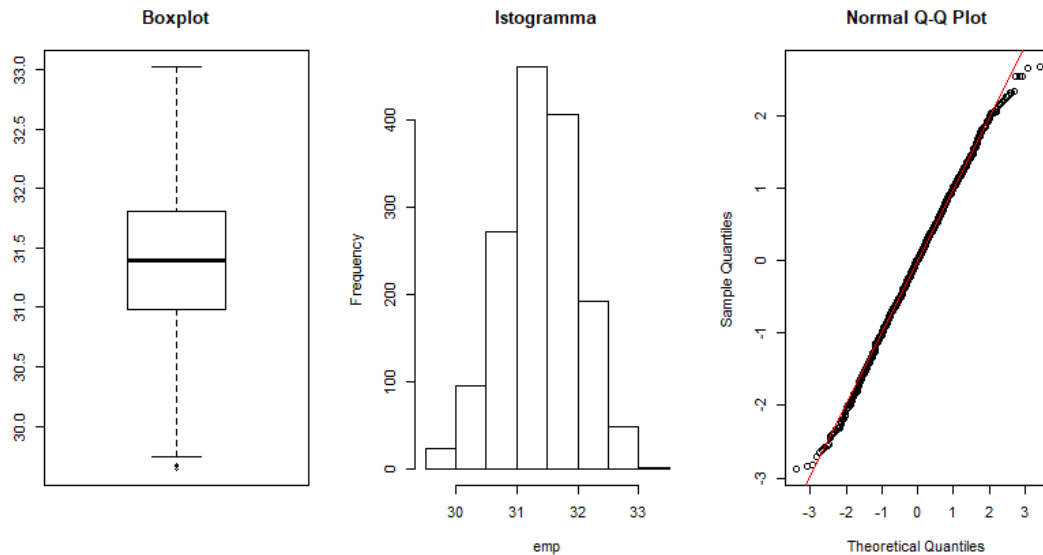


Fig. 4.4: Boxplot, istogramma e Q-Q plot per l'età media al parto

Tasso di natalità

Il tasso di natalità vale tra un minimo di 5,2 e un massimo di 11,8. La media (8,418) è molto simile alla mediana (8,4). La distribuzione empirica è di poco asimmetrica, ma si potrebbe pensare ad una reale simmetria distributiva. Viene rifiutata l'ipotesi di normalità ($p\text{-value} = 7.224e-05$).

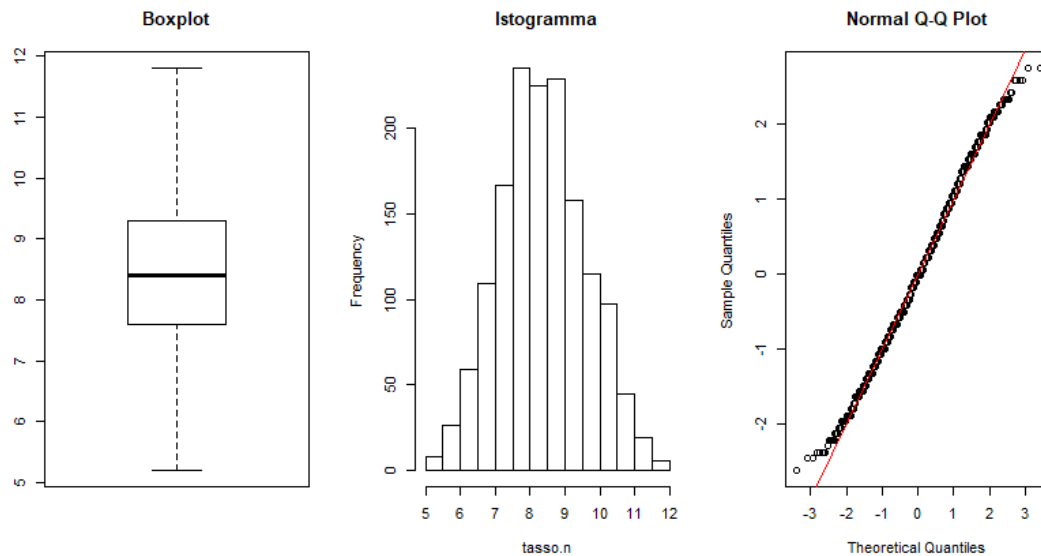


Fig. 4.5: Boxplot, istogramma e Q-Q plot per il tasso di natalità

Età media

L'età media ha valori compresi tra 37,5 e 49,2 con media di 44,36 e mediana di 44,5. E' presente una leggera asimmetria negativa. Viene rifiutata la normalità con $p\text{-value}=1.64e-06$.

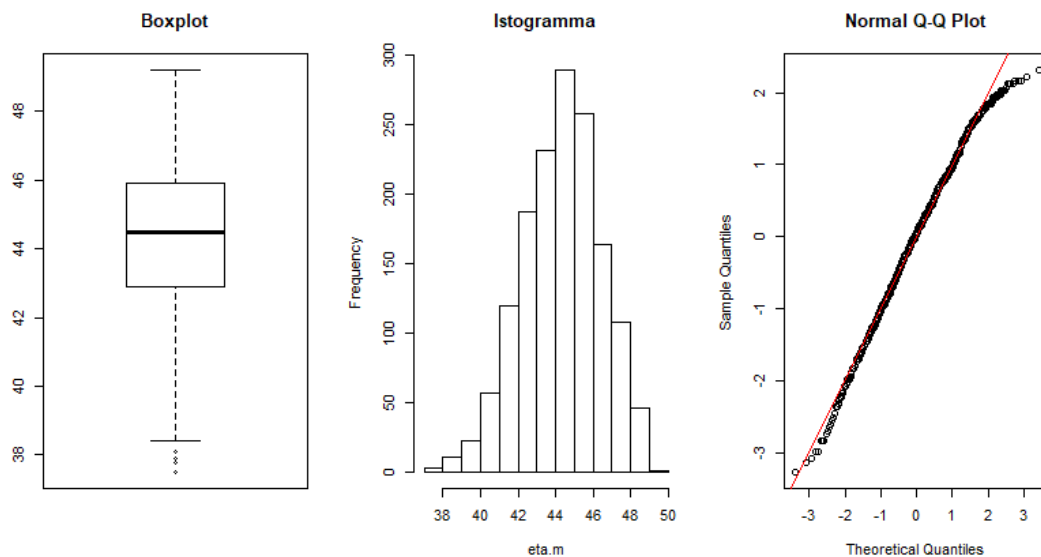


Fig. 4.6: Boxplot, istogramma e Q-Q plot per l'età media

Indice di vecchiaia

L'indice di vecchiaia ha valori osservati tra 70,8 e 260,1. La sua media (166,2) è di poco superiore alla mediana (164,4). Non è chiaro il tipo di asimmetria presente. Il

test di Shapiro-Wilk rifiuta l'ipotesi di normalità (p-value=2.473e-09).

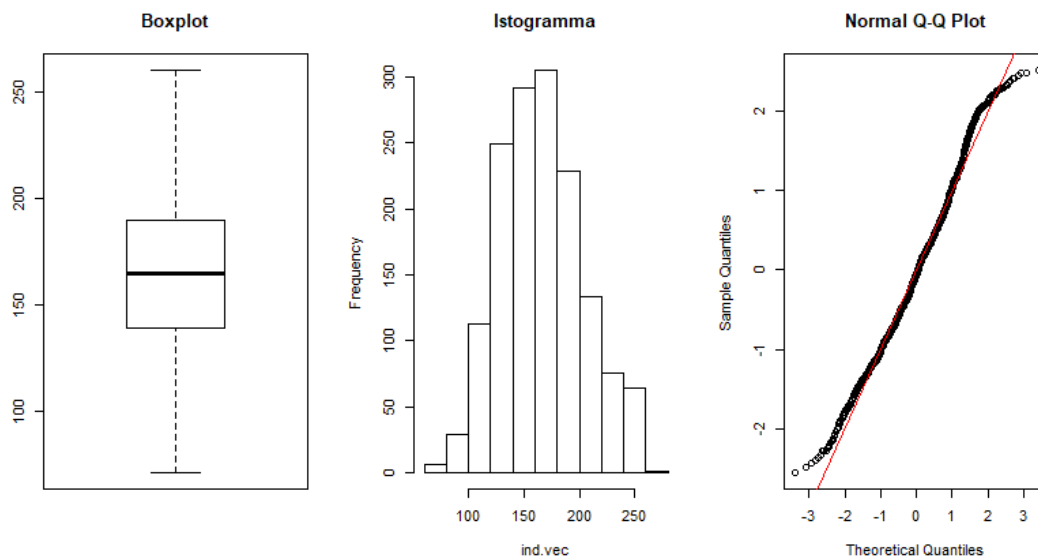


Fig. 4.7: Boxplot, istogramma e Q-Q plot per l'indice di vecchiaia

Indice di dipendenza strutturale

L'indice di dipendenza strutturale presenta valori tra 40,6 e 67,3. Ha media pari a 54,54 e mediana a 54,2. Si rifiuta la normalità (p-value=3.154e-07).

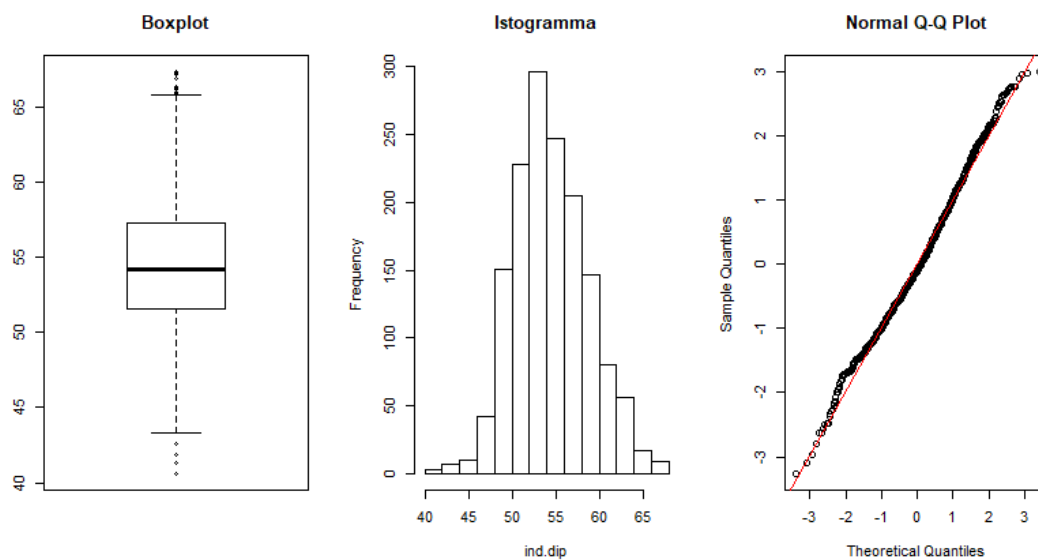


Fig. 4.8: Boxplot, istogramma e Q-Q plot per l'indice di dipendenza strutturale

Indice di ricambio della popolazione attiva

L'indice di ricambio della popolazione attiva varia empiricamente tra 63,8 e 219,5. Ha media di 132,1 e mediana di 133,1. Anche in questo caso si rifiuta l'ipotesi di normalità ($p\text{-value}=1.401e-06$).

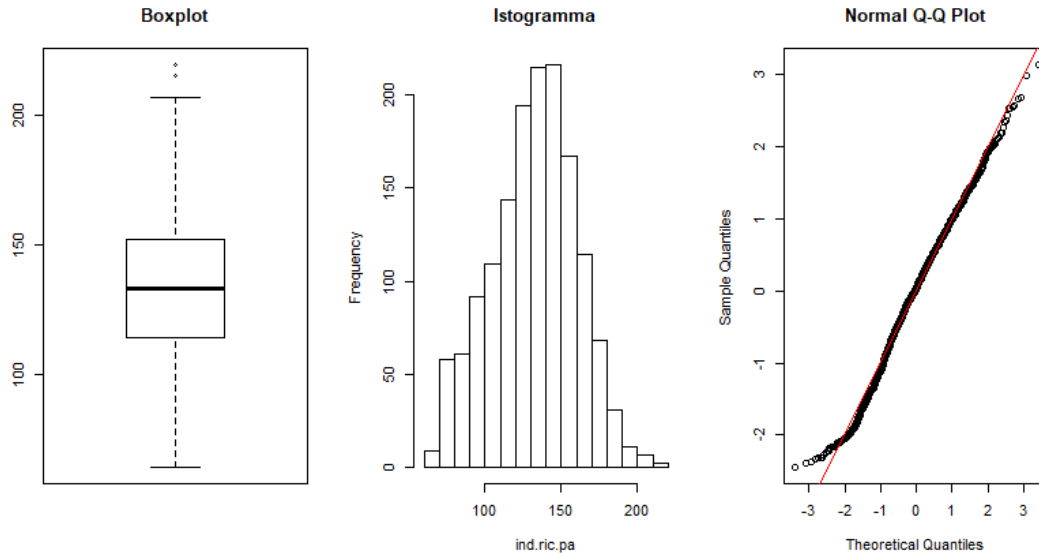


Fig. 4.9: Boxplot, istogramma e Q-Q plot per l'indice di ricambio della popolazione attiva

4.4.2 Serie storiche

Andiamo ad analizzare l'andamento nel tempo dei tassi di emigratorietà e immigratorietà per il periodo di riferimento. I grafici relativi agli altri indicatori sono invece presenti in Appendice. Si segnala che per tutti i casi in appendice, si ha hanno in ascissa gli anni traslati di -1 (quindi $t-1$ invece di t) per mantenere anche in questo caso un punto di vista di causalità. Per semplicità mostreremo la serie storica a livello grafico dei soli 5 aggregati territoriali classificati secondo l'Istat.

La serie storica dei tassi di emigratorietà per macroaree è la seguente:

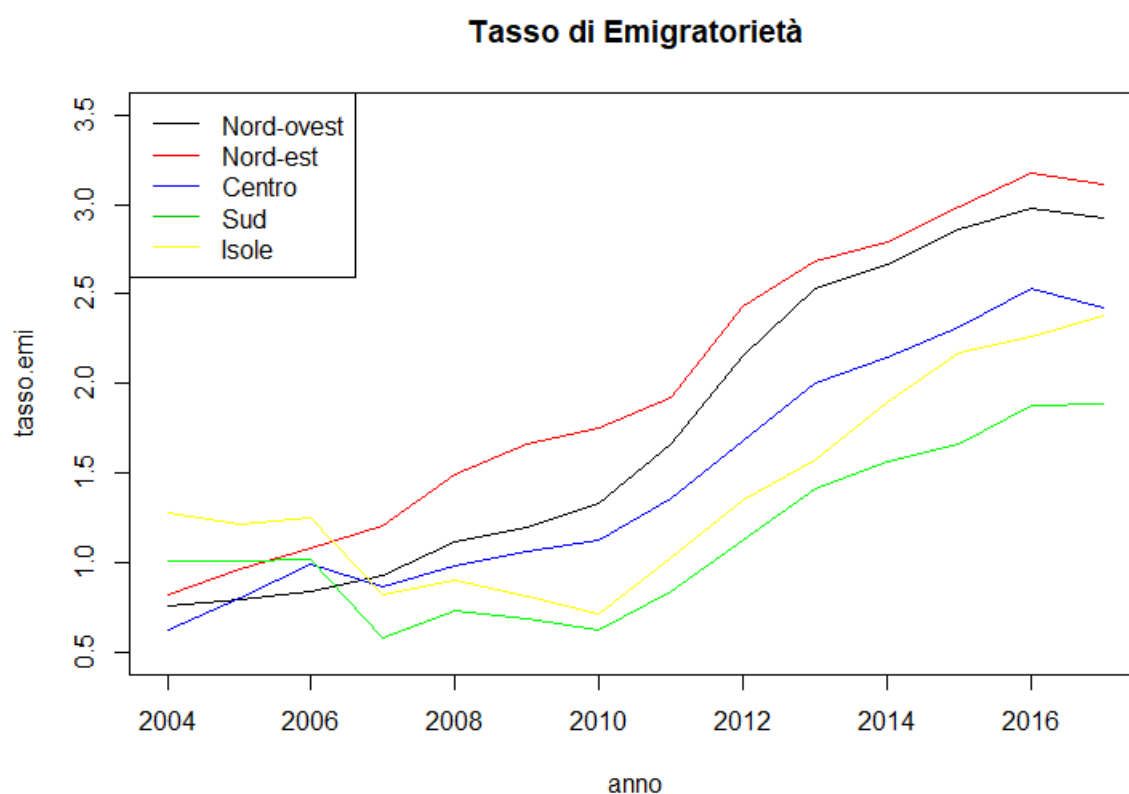


Fig. 4.10: Serie storiche dei tassi di emigratorietà

Si può notare per tutti i territori un trend per lo più crescente. Se all'inizio del periodo il Sud Italia e le Isole avevano livelli leggermente superiori rispetto alle due aree del Nord e al Centro, successivamente la situazione si è invertita. Infatti Nord-ovest, Nord-est e il Centro hanno continuato a crescere in termini di tasso di emigratorietà superando verso il 2006 le altre aree. A differenza di quanto si può vedere per le emigrazioni in termini assoluti, il Nord-ovest ha mantenuto un tasso di emigratorietà sempre leggermente inferiore rispetto al Nord-est. Questo è dovuto al fatto che nel Nord-ovest vi siano più abitanti rispetto alle altre macroaree del paese, e ciò va ad abbassare i tassi migratori ai vari anni. Si vede inoltre, che verso fine periodo il trend ha smesso generalmente di essere positivo ad eccezione del Sud. La serie storica relativa al tasso di immigratorietà è invece:

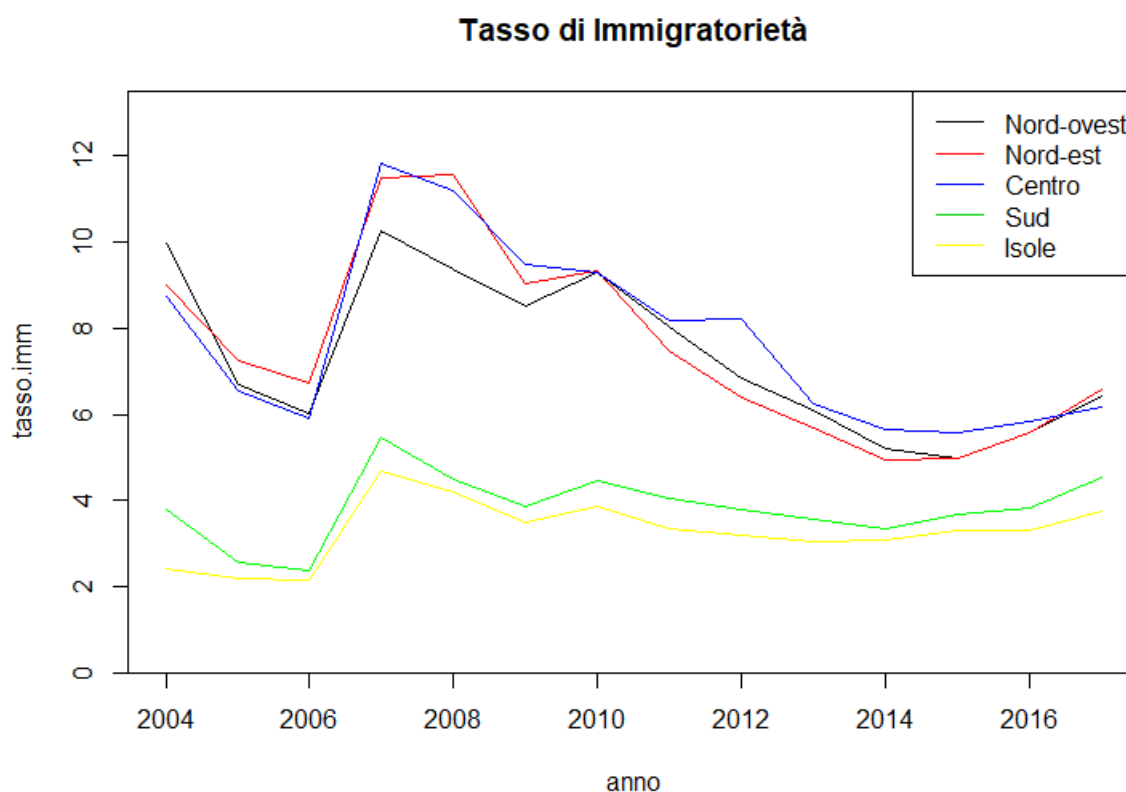


Fig. 4.11: Serie storiche dei tassi di immigratorietà

Si osserva che i tassi di immigratorietà stavano decrescendo a partire dal 2004, ma con l'inizio della crisi economica si è avuta una crescita rapidissima dei tassi per poi tornare gradualmente a decrescere gli anni a seguire. Salta all'occhio il fatto che si possano distinguere sostanzialmente due gruppi di macroaree: il gruppo composto da Nord-ovest, Nord-est e Centro mostrano sempre dei valori più elevati rispetto al Sud e alle Isole, con un particolare divario raggiunto nel 2007, anno di inizio della crisi.

4.4.3 Analisi bivariata

Si è interessati a vedere intuitivamente le relazioni tra i due tassi di migratorietà con quelle che saranno le variabili risposta nei modelli. In questo contesto l'analisi bivariata ha l'obiettivo di valutare la distribuzione congiunta tra le concomitanti e le risposte. Vengono utilizzati a livello grafico dei diagrammi di dispersione con i tassi di migratorietà sulle ascisse e le variabili risposte calcolate all'anno successivo sulle ordinate. Si calcola successivamente, per ogni coppia di variabili, l'indice di correlazione di Spearman poichè non assume la normalità delle variabili. Infine si effettuerà il test di Wilcoxon-Mann-Whitney per verificare l'ipotesi di uguaglianza tra le distribuzioni. Questo citato è un test non parametrico per ranghi che si basa sulle mediane delle distribuzioni.

Tasso di fecondità totale

E' presente una leggera correlazione negativa tra il tasso di emigratorietà e il TFT pari a $\rho = -0.004$ confermata dal test di Wilcoxon che rifiuta l'ipotesi di uguaglianza delle distribuzioni (p-value=0.001145). Più evidente è invece la correlazione positiva tra il tasso di immigratorietà e il TFT che ammonta a $\rho = 0.459$, anche questo verificato ampiamente tramite il test di Wilcoxon (p-value<2.2e-16). Questi risultati indicano che all'aumentare delle emigrazioni si osserva una lieve, quasi impercettibile, diminuzione del TFT, mentre all'aumentare del fenomeno immigratorio è forte l'aumento del TFT.

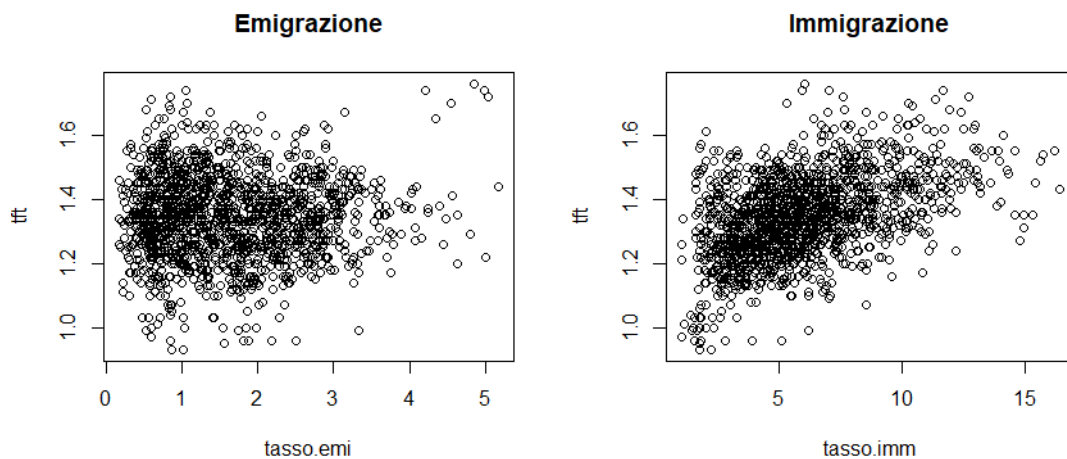


Fig. 4.12: Plot dei tassi di migratorietà con il tasso di fecondità totale

Età media al parto

Si osserva una modesta correlazione positiva tra emigrazione ed EMP ($\rho = 0.419$) e negativa tra immigrazione ed EMP ($\rho = -0.155$). Entrambe confermate dal test di Wilcoxon (p-value<2.2e-16).

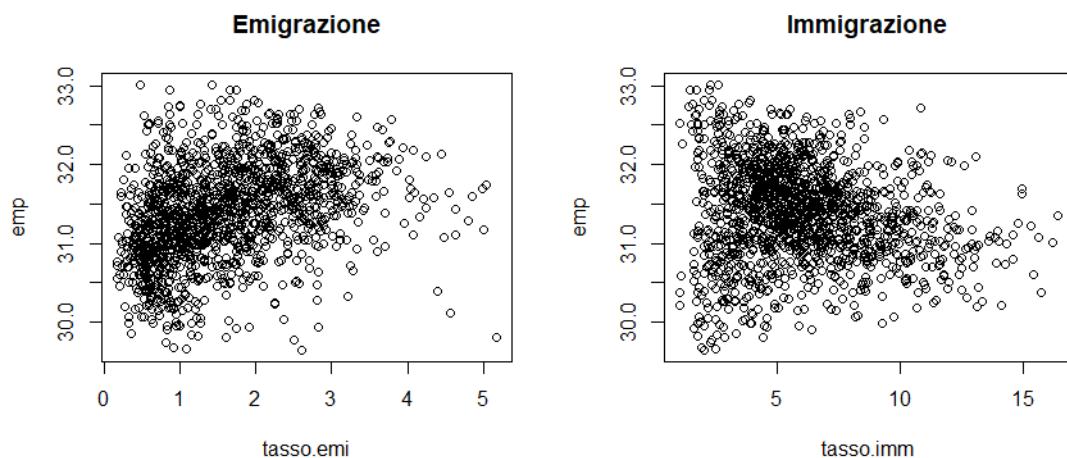


Fig. 4.13: Plot dei tassi di migratorietà con l'età media al parto

Tasso di natalità

Coerentemente con il TFT, si ha correlazione negativa tra emigrazione e tasso di natalità ($\rho = -0.473$) e correlazione positiva con l'immigrazione ($\rho = 0.285$). Ancora una volta il test di Wilcoxon segnala un p-value molto significativo ($<2.2e-16$).

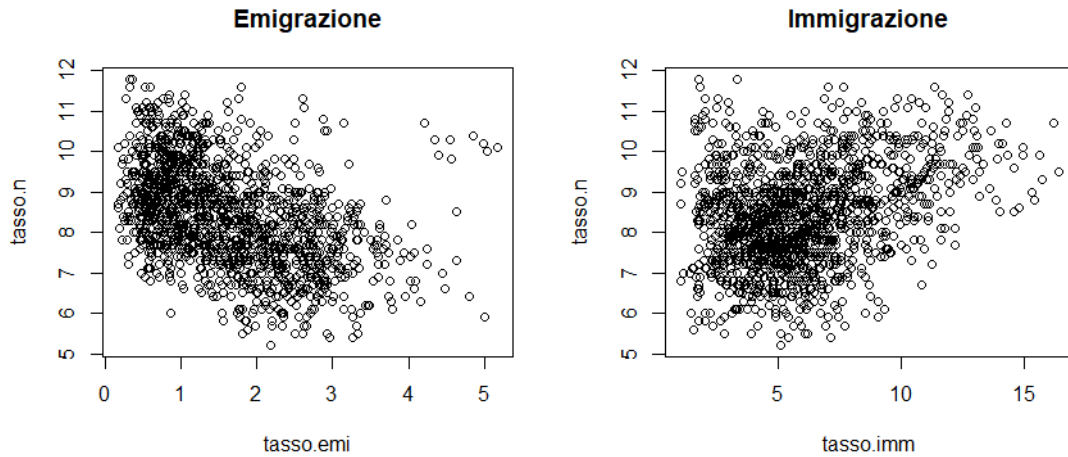


Fig. 4.14: Plot dei tassi di migratorietà con il tasso di natalità

Età media

E' presente correlazione positiva sia tra il tasso di emigratorietà con l'età media ($\rho = 0.353$), sia tra quest'ultimo con il tasso di immigratorietà ($\rho = 0.238$). Il test rifiuta l'ipotesi nulla con $p\text{-value} < 2.2e-16$.

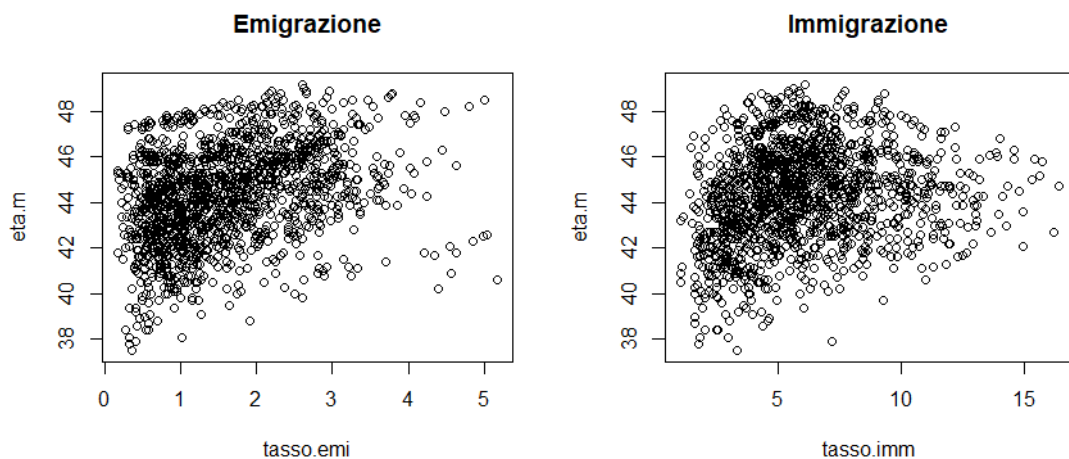


Fig. 4.15: Plot dei tassi di migratorietà con l'età media

Indice di vecchiaia

Si calcola un valore di correlazione positiva modesto molto simile tra i due tassi di migratorietà con l'indice di vecchiaia (rispettivamente di $\rho = 0.209$ per l'emigratorietà

e $\rho = 0.214$ per l'immigratorietà). Il test rifiuta l'ipotesi nulla con $p\text{-value} < 2.2e-16$.

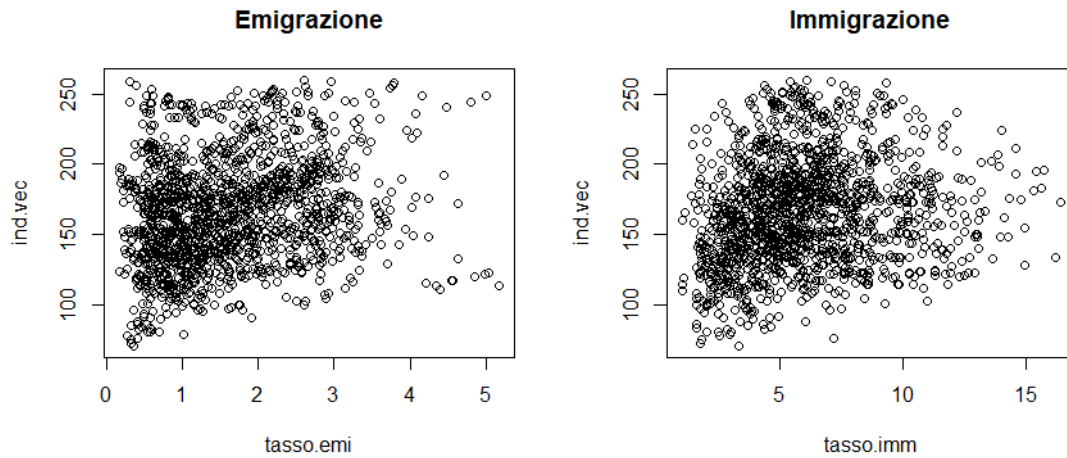


Fig. 4.16: Plot dei tassi di migratorietà con l'indice di vecchiaia

Indice di dipendenza strutturale

Si calcola un valore di correlazione positiva tra emigrazione e indice di dipendenza strutturale $\rho = 0.49$, lo stesso vale tra immigrazione e indice di dipendenza strutturale $\rho = 0.266$. Entrambi mostrano un $p\text{-value} < 2.2e-16$ nel test associato.

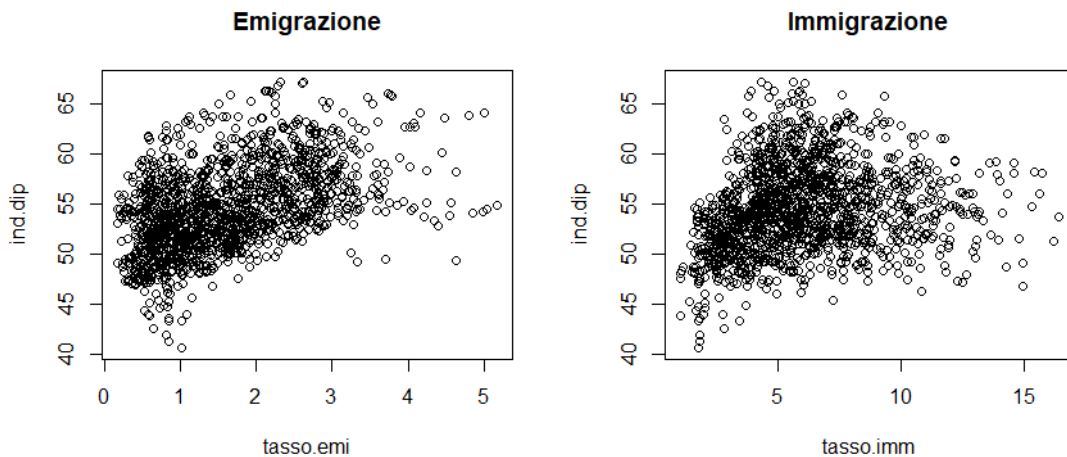


Fig. 4.17: Plot dei tassi di migratorietà con l'indice di dipendenza strutturale

Indice di ricambio della popolazione attiva

Come per tutti gli altri indici di struttura risulta una correlazione positiva tra i tassi di emigratorietà e immigratorietà con l'indice in questione (rispettivamente pari a $\rho = 0.194$ e $\rho = 0.408$). Il test di Wilcoxon ancora una volta scarta l'ipotesi nulla con $p\text{-value} < 2.2e-16$.

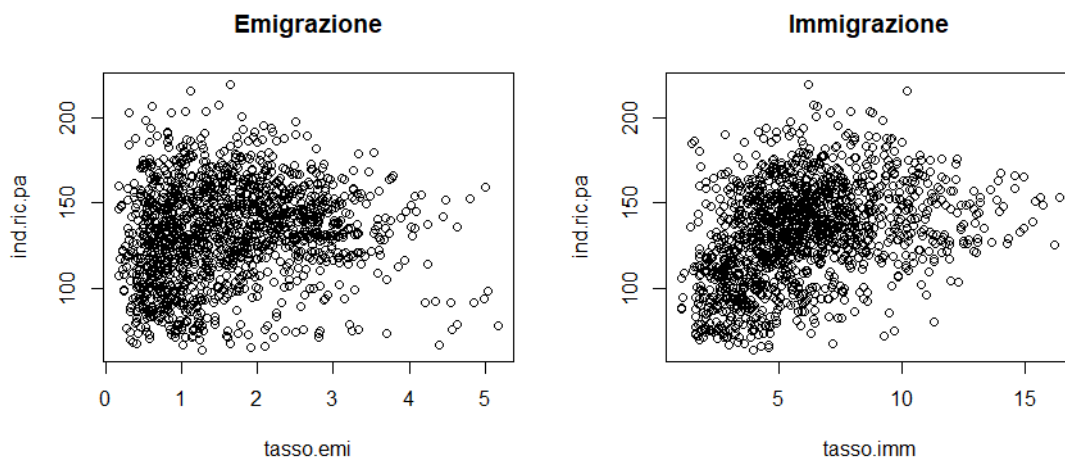


Fig. 4.18: Plot dei tassi di migratorietà con l'indice di ricambio della popolazione attiva

Si segnala anche che le variabili esplicative che verranno incluse nei modelli non sono eccessivamente correlate tra loro, di conseguenza, non si impongono importanti limitazioni nei modelli che si utilizzeranno nell'analisi.

4.5 Applicazione dei modelli

Si vuole ora applicare ognuno dei tre modelli presentati a livello teorico. Per presentare i risultati vengono riportati per ciascun indicatore le stime dei parametri significativi relativi alle esplicative, le stime dei relativi errori standard e gli intervalli di confidenza al 95%. La notazione utilizzata è quella scientifica con l'esponentiale. Si è deciso poi di presentare gli intervalli di confidenza basati sulla quantità $r_p(\beta_i)$, ovvero sulla versione unilaterale del rapporto di verosimiglianza profilo (alternativamente si potevano utilizzare gli intervalli di confidenza basati sulla statistica di Wald). In seguito si indicheranno i parametri con le seguenti diciture: *tasso.emi* è il tasso di emigratorietà, *tasso.imm* il tasso di immigratorietà, *anno* l'anno t-1, *tasso.d* il tasso di disoccupazione, *pil.reg* il pil della regione e *tasso.d : pil.reg* l'eventuale interazione tra il tasso di disoccupazione e il pil della regione. Per agevolare la trattazione, nelle tabelle non verranno riportati i valori riferiti alle variabili dicotomiche che esprimono il fattore "provincia", di conseguenza i risultati faranno sempre riferimento alla prima provincia italiana per ordine alfabetico (Agrigento). Tali variabili vanno infatti ad esprimere soltanto le diverse intercette che presentano le singole province e ciò non risulta rilevante ai fini dell'analisi.

Poichè i modelli ols e glm presentano risultati molto simili per quanto riguarda le stime e il legame identità per il modello glm gamma risulta sempre il preferibile in termini di AIC, i risultati per i primi due modelli verranno trattati assieme prendendo come riferimento il modello ols; in fase di confronto tra i modelli si indicherà quale dei due risulta preferibile. In questo senso è da evidenziare che i modelli ols e glm restituiscono come significative le stesse variabili esplicative. I modelli selezionati e presentati nelle tabelle sottostanti non hanno come finalità unicamente

l'aspetto interpretativo, ma fanno fortemente leva sull'aspetto del bilanciamento tra adattamento ai dati e parsimonia nel numero di variabili, nonchè possono eventualmente essere utilizzati a fini previsivi. Per questo motivo per alcuni indicatori, i parametri relativi alle migrazioni possono assumere dei valori stimati controintuitivi. Tali aspetti verranno analizzati e spiegati, indicando eventuali accortezze per adattare i modelli a fini puramente interpretativi.

Tasso di fecondità totale

Il modello ols per questo indicatore segnala un effetto negativo del tasso di emigratorietà per il TFT e positivo per il tasso di immigratorietà. A parità delle altre variabili, infatti, un aumento unitario del tasso di emigrazione comporta una diminuzione del TFT di 0,01178, mentre un aumento del tasso di immigratorietà implica un aumento unitario di 0,008451 sul TFT. Si nota che il tasso di emigratorietà è modestamente significativo (p-value=0.001189), invece il tasso di immigratorietà è fortemente significativo (p-value<2e-16) (nel glm sono entrambi molto significativi). Le variabili confondenti operano in direzioni opposte. Il tasso di disoccupazione agisce negativamente sul TFT, al contrario un maggiore pil regionale favorisce il presentarsi di un TFT elevato. E' presente anche un effetto di interazione tra le variabili confondenti. Il modello "first difference" conferma quanto detto per l'effetto delle migrazioni. Un aumento del tasso di emigratorietà prevede infatti una variazione negativa del TFT e un aumento del tasso di immigratorietà una variazione positiva. Per questo modello, il pil regionale non è più significativo (e nemmeno l'interazione con il tasso di disoccupazione). Quest'ultimo risultato trova coerenza con gli altri modelli, poichè anche in questi le stime del coefficiente relativo al pil regionale era molto vicino ad essere nullo.

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	8,583	1,659	5,3284	11,8383
tasso.emi	-0,0118	0,0036	-0,0189	-0,0047
tasso.imm	0,0085	0,0009	0,0067	0,0102
anno	-0,0037	0,0008	-0,0053	-0,002
tasso.d	-0,0066	0,0019	-0,0104	-0,0028
pil.reg	$1,001 * 10^{-5}$	$1,988 * 10^{-6}$	$6,109 * 10^{-6}$	$1,391 * 10^{-5}$
tasso.d:pil.reg	$2,709 * 10^{-7}$	$9,243 * 10^{-8}$	$8,954 * 10^{-8}$	$4,522 * 10^{-7}$

Tab. 4.3: OLS tasso di fecondità totale: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	8,311	1,598	5,1836	11,1445
tasso.emi	-0,0116	0,0035	-0,0185	-0,0048
tasso.imm	0,0083	0,0009	0,0065	0,0101
anno	-0,0035	0,0008	-0,0051	-0,0019
tasso.d	-0,0062	0,0019	-0,0099	-0,0025
pil.reg	$9,188 * 10^{-6}$	$1,996 * 10^{-6}$	$5,247 * 10^{-6}$	$1,313 * 10^{-5}$
tasso.d:pil.reg	$2,553 * 10^{-7}$	$9,152 * 10^{-8}$	$7,563 * 10^{-8}$	$4,35 * 10^{-7}$

Tab. 4.4: GLM tasso di fecondità totale: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	9,492	0,7345	8,0514	11,0933
tasso.emi	-0,0082	0,0041	-0,0163	$-8,413861 * 10^{-5}$
tasso.imm	0,0036	0,0007	0,0022	0,0049
anno	-0,0047	0,0004	-0,0054	-0,004
tasso.d	-0,0015	0,0007	-0,0029	$-5,031701 * 10^{-5}$

Tab. 4.5: OLS "first difference" tasso di fecondità totale: Stime, errori standard e i.c al 95%

Età media al parto

Nei modelli per l'EMP non risulta significativa nessuna delle due variabili confondenti. Questo potrebbe indicare che non ci sia evidenza empirica per una possibile relazione tra fattori economici ed età media al parto. Secondo il modello ols, un aumento unitario del tasso di emigratorietà comporta una diminuzione di 0,03063 anni dell'EMP e un aumento unitario del tasso di immigratorietà implica una sua diminuzione di 0,01234 anni. Tale risultato sembrerebbe entrare in contrasto con le analisi preliminari, dove si evidenziava invece una correlazione positiva tra emigrazioni e EMP. Si pensa quindi che le altre variabili introdotte nel modello riescano a discernere l'effetto delle migrazioni. In particolare, andando ad eliminare la variabile *anno* e il fattore provincia, si nota che l'effetto del tasso di emigratorietà sull'EMP torna coerente con le analisi preliminari. Di conseguenza, l'effetto positivo attribuibile precedentemente all'emigrazione è qui in realtà associato alla variabile indicante il tempo e al fattore provincia. Bisogna dunque stare molto attenti a trarre delle conclusioni. La variabile temporale va infatti ad attribuire un peso diverso alle migrazioni a seconda dell'anno che si va a considerare (in maniera analoga il fattore provincia). Si riporta ad esempio che, senza la variabile *anno* e il fattore provincia, un aumento unitario del tasso di emigratorietà porta ad un aumento di 0,2257 dell'EMP, mentre il tasso di immigratorietà comporta ancora una diminuzione dell'EMP ($\beta_2 = -0.0288$). In conclusione si pensa allora che generalmente l'emigrazione per le province italiane porti in termini assoluti ad una diminuzione dell'età media al parto, ma "standardizzando" l'effetto del tempo e delle province, questo effetto risulta opposto. Si potrebbe inoltre pensare di introdurre un termine di interazione tra i tassi di migratorietà e la variabile *anno* per comprendere meglio la relazione tra i due fenomeni. Successivamente, in questa analisi il modello "first

difference” non segnala alcuna significatività dei parametri, oltre al fatto che il coefficiente di determinazione R^2 è prossimo a zero. Per questo si decide di scartare tale modello.

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-138,3	3,426	-145,0475	-131,605
tasso.emi	-0,0306	0,0094	-0,049	-0,0123
tasso.imm	-0,0123	0,0023	-0,0168	-0,0079
anno	0,0842	0,0017	0,0808	0,0875

Tab. 4.6: OLS età media al parto: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-100,389	3,385	-145,5132	-132,2318
tasso.emi	-0,0318	0,0092	-0,05	-0,0137
tasso.imm	-0,0125	0,0023	-0,0169	-0,0081
anno	0,0844	0,0017	0,0811	0,0878

Tab. 4.7: GLM età media al parto: Stime, errori standard e i.c al 95%

Tasso di natalità

Coerentemente con quanto segnalato dall’effetto sul TFT, i modelli in questione indicano un impatto negativo del tasso di emigratorietà e un impatto positivo del tasso di immigratorietà sul tasso di natalità. Questo risultato era attendibile, in quanto sia TFT che tasso di natalità sono misure di intensità delle nascite. Analizzando da vicino si osserva nel modello ols semplice che un aumento unitario del tasso di emigratorietà induce una diminuzione di 0,1593 del tasso di natalità, mentre un aumento unitario del tasso di immigratorietà implica un suo aumento di 0,06674 (sempre al netto delle altre esplicative). Come testimoniato anche dal modello ”first difference”, anche in questo caso si può notare che il tasso di natalità è più suscettibile al variare delle emigrazioni rispetto alle immigrazioni. D’altra parte però, il coefficiente relativo all’immigrazione è sempre più significativo rispetto a quello dell’emigrazione. Per quanto riguarda i fattori economici, si può vedere che il tasso di disoccupazione ha un effetto negativo sul tasso di natalità, ma risulta significativo soltanto nel modello ”first difference”; un pil regionale elevato è invece in tutti e tre i casi un fattore che va a favorire l’aumento della natalità (proprio come avviene per il TFT).

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	327,9	8,442	311,3045	344,4246
tasso.emi	-0,1576	0,0219	-0,2005	-0,1147
tasso.imm	0,0668	0,0055	0,056	0,0776
anno	-0,1591	0,0043	-0,167	-0,1507
pil.reg	$4,872 * 10^{-5}$	$1,054 * 10^{-5}$	$2,8038 * 10^{-5}$	$6,9409 * 10^{-5}$

Tab. 4.8: OLS tasso di natalità: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	324,4	8,508	307,7966	341,112
tasso.emi	-0,1593	0,0218	-0,2016	-0,1168
tasso.imm	0,0667	0,0055	0,0559	0,0776
anno	-0,1571	0,0043	-0,1655	-0,1487
pil.reg	$1,8 * 10^{-5}$	$1,037 * 10^{-5}$	$-2,5016 * 10^{-6}$	$3,8527 * 10^{-5}$

Tab. 4.9: GLM tasso di natalità: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	51,1	4,586	42,0982	60,0934
tasso.emi	-0,0836	0,258	-0,1343	-0,033
tasso.imm	0,0229	0,0044	0,0142	0,0316
anno	-0,0255	0,0023	-0,03	-0,021
tasso.d	-0,015	0,0047	-0,0242	-0,0059
pil.reg	$2,24 * 10^{-5}$	$1,113 * 10^{-5}$	$5,5939 * 10^{-7}$	$4,424 * 10^{-5}$

Tab. 4.10: OLS "first difference" tasso di natalità: Stime, errori standard e i.c al 95%

Età media

Per questo indice, si ha secondo il modello ols semplice che un aumento unitario del tasso di emigratorietà induce una diminuzione di 0,158 anni dell'età media della popolazione italiana, invece l'aumento unitario del tasso di immigratorietà porta ad un aumento dell'età media di 0,01087 anni. Come per l'EMP l'effetto dell'emigrazione è distorto dalla variabile anno e dal fattore provincia. Infatti togliendo questi due elementi, un aumento unitario del tasso di emigratorietà comporta un aumento di 0,56 anni dell'età media e di 0,09389 anni per un aumento unitario del tasso di immigratorietà. In linea di massima è logico pensare che l'emigrazione in Italia vada ad alzare l'età media della popolazione perchè negli anni presi in considerazione l'età media degli emigrati risultava minore rispetto a quella della popolazione residente. Per adattare quindi un modello per soli fini interpretativi sono perseguibili varie strade: eliminare le variabili tempo e il fattore provincia o introdurre dei termini di interazione tra migrazioni e tempo/fattore provincia. Per quanto riguarda il modello "first difference", non risulta significativa al 95% la variabile *tasso.emi*. E' presente un effetto negativo delle immigrazioni sull'età media.

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-452,2	8,838	-469,5021	-434,8282
tasso.emi	-0,158	0,0193	-0,1959	-0,12
tasso.imm	0,0109	0,0048	0,0014	0,0204
anno	0,2461	0,0045	0,2374	0,2549
tasso.d	0,1318	0,0103	0,1117	0,1519
pil.reg	$-3,005 * 10^{-5}$	$1,059 * 10^{-5}$	$-5,0815 * 10^{-5}$	$-9,276 * 10^{-6}$
tasso.d:pil.reg	$-5,989 * 10^{-6}$	$4,923 * 10^{-7}$	$-6,955 * 10^{-6}$	$-5,0235 * 10^{-6}$

Tab. 4.11: OLS età media: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-459,9	8,553	-476,6382	-443,0947
tasso.emi	-0,1606	0,0187	-0,1971	-0,124
tasso.imm	0,0102	0,0048	0,0008	0,0197
anno	0,25	0,0043	0,2415	0,2585
tasso.d	0,1278	0,01	0,1082	0,1473
pil.reg	$-3,215 * 10^{-5}$	$1,052 * 10^{-5}$	$-5,2756 * 10^{-5}$	$-1,1536 * 10^{-5}$
tasso.d:pil.reg	$-5,941 * 10^{-6}$	$4,894 * 10^{-7}$	$-6,9 * 10^{-6}$	$-4,9827 * 10^{-6}$

Tab. 4.12: GLM età media: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-17,15	1,136	-19,3759	-14,92
tasso.imm	-0,0084	0,0011	-0,0105	-0,0062
anno	0,0087	0,0006	0,0076	0,0098
tasso.d	-0,0043	0,0012	-0,0066	-0,002
pil.reg	$6,854 * 10^{-6}$	$2,775 * 10^{-6}$	$1,4094 * 10^{-6}$	$1,23 * 10^{-5}$

Tab. 4.13: OLS "first difference" età media: Stime, errori standard e i.c al 95%

Indice di vecchiaia

Per l'indice di vecchiaia si osserva un comportamento dei modelli simile a quello per l'età media. Infatti per l'ols semplice il tasso di emigratorietà ha un effetto negativo sull'indice (contrariamente a quanto osservato in analisi esplorativa), invece il tasso di immigratorietà ha un effetto positivo. Anche in questo caso la causa è da attribuire alla variabile *anno* e al fattore provincia. Eliminandole si osserva che, sempre per il modello ols semplice, un aumento unitario di *tasso.emi* porta ad un aumento di 6,312 (in punti percentuale) dell'indice di vecchiaia e un aumento unitario di *tasso.imm* restituisce un aumento di 1,628 dello stesso indice. Sembrerebbe quindi che ad emigrare siano maggiormente le persone under 14 rispetto agli over 65, mentre tra gli immigrati sembrano esserci più over 65 rispetto agli under 14. Come nel caso dell'età media, il modello "first difference" riporta un effetto delle emigrazioni sull'indice non significativo e un effetto delle immigrazioni negativo.

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-5504	212,2	-5920,236e	-5087,578
tasso.emi	-1,274	0,464	-2,1839	-0,3633
tasso.imm	0,309	0,1163	0,0808	0,5371
anno	2,805	0,1074	2,5946	3,0161
tasso.d	2,827	0,2461	2,3441	3,3096
pil.reg	-0,0007	0,0003	-0,0012	-0,0002
tasso.d:pil.reg	-0,0001	$1,182 * 10^{-5}$	-0,0002	-0,0001

Tab. 4.14: OLS indice di vecchiaia: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-5868	172,4	-6206,52	-5528,521
tasso.emi	-1,248	0,3777	-1,986	-0,5056
tasso.imm	0,2446	0,1052	0,0404	0,4492
anno	2,990	0,08736	2,8183	3,1618
tasso.d	2,342	0,2042	1,9428	2,7422
pil.reg	-0,0008	0,0002	-0,0013	-0,0004
tasso.d:pil.reg	-0,0001	$1,076 * 10^{-5}$	-0,0001	$-9,7542 * 10^{-5}$

Tab. 4.15: GLM indice di vecchiaia: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-818,7	22,06	-861,9655	-775,4063
tasso.imm	-0,1222	0,0211	-0,1636	-0,0808
anno	0,409	0,011	0,3875	0,4305
pil.reg	0,0003	$5,165 * 10^{-5}$	0,0002	0,0004

Tab. 4.16: OLS "first difference" indice di vecchiaia: Stime, errori standard e i.c al 95%

Indice di dipendenza strutturale

In questo caso il modello ols semplice prevede che il tasso di emigratorietà abbia un effetto positivo sull'indice di dipendenza strutturale e il tasso di immigratorietà abbia un effetto negativo. Volendo offrire dei risultati validi unicamente per semplicità interpretativa, si può ancora una volta eliminare l'effetto del tempo e delle province. Sotto questo modo di procedere, un aumento di *tasso.emi* fa aumentare l'indice di dipendenza strutturale di 2,148 punti percentuali, mentre un aumento di *tasso.imm* induce un aumento dello stesso indice di 0,2307 punti percentuali. Per motivazioni analoghe il modello "first difference" con tutte le variabili segnala inizialmente un effetto positivo delle emigrazioni e positivo delle immigrazioni.

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-282,1	28,53	-338,0374	-226,1142
tasso.emi	0,8727	0,0624	0,7503	0,9951
tasso.imm	-0,0752	0,0156	-0,1059	-0,0445
anno	0,1642	0,0144	0,1358	0,1925
tasso.d	-0,1831	0,0331	-0,248	-0,1182
pil.reg	0,0002	$3,418 * 10^{-5}$	0,0001	0,0003
tasso.d:pil.reg	$1,142 * 10^{-5}$	$1,589 * 10^{-6}$	$8,3057 * 10^{-6}$	$1,454 * 10^{-5}$

Tab. 4.17: OLS indice di dipendenza strutturale: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-261,8	28,31	-317,0933	-206,4052
tasso.emi	0,9084	0,0625	0,7859	1,0311
tasso.imm	-0,0716	0,0158	-0,1024	-0,0407
anno	0,1538	0,0143	0,1258	0,1818
tasso.d	-0,1809	0,0327	-0,2448	-0,1169
pil.reg	0,0002	$3,473 * 10^{-5}$	0,0002	0,0003
tasso.d:pil.reg	$1,188 * 10^{-5}$	$1,618 * 10^{-6}$	$8,7057 * 10^{-6}$	$1,505 * 10^{-5}$

Tab. 4.18: GLM indice di dipendenza strutturale: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-31,25	5,054	-41,1711	-21,3382
tasso.emi	0,2137	0,0284	0,158	0,2693
tasso.imm	-0,0461	0,0049	-0,0557	-0,0365
anno	0,0155	0,0025	0,0106	0,0204
tasso.d	0,0269	0,0053	0,0165	0,0374
pil.reg	$3,561 * 10^{-5}$	$1,379 * 10^{-5}$	$8,563 * 10^{-6}$	$6,2657 * 10^{-5}$
tasso.d:pil.reg	$-3,543 * 10^{-5}$	$8,745 * 10^{-6}$	$-5,2592 * 10^{-5}$	$-1,8277 * 10^{-5}$

Tab. 4.19: OLS "first difference" indice di dipendenza strutturale: Stime, errori standard e i.c al 95%

Indice di ricambio della popolazione attiva

Come nel caso dell'indice di vecchiaia, anche qui, per tutti i modelli, il tasso di emigratorietà ha un effetto negativo sull'indice di ricambio e il tasso di immigratorietà ne ha uno positivo. Non considerando il tempo e il fattore provincia, si osserva che un aumento unitario di *tasso.emi* fa aumentare di 1,122 punti percentuali l'indice, mentre un aumento unitario di *tasso.imm* lo fa aumentare di 1,927 punti percentuali. Si conclude quindi che sia le emigrazioni che le immigrazioni vadano a sfavorire il ricambio generazionale.

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-6672	265,1	-7192,131	-6151,952
tasso.emi	-7,479	0,5797	-8,6165	-6,3421
tasso.imm	1,51	0,1453	1,2254	1,7955
anno	3,397	0,1342	3,1336	3,6602
tasso.d	2,347	0,3074	1,7442	2,9503
pil.reg	-0,0034	0,0003	-0,004	-0,0028
tasso.d:pil.reg	-0,0001	$1,477 * 10^{-5}$	-0,0001	$-7,6223 * 10^{-5}$

Tab. 4.20: OLS indice di ricambio della popolazione attiva: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	-7156	208,4	-7563,59	-6748,33
tasso.emi	-6,394	0,4323	-7,2165	-5,5644
tasso.imm	1,41	0,1386	1,1353	1,686
anno	3,642	0,1057	3,4354	3,849
tasso.d	1,853	0,2652	1,3312	2,3753
pil.reg	-0,0034	0,0003	-0,004	-0,0029
tasso.d:pil.reg	-0,0001	$1,403 * 10^{-5}$	-0,0001	$-7,2832 * 10^{-5}$

Tab. 4.21: GLM indice di ricambio della popolazione attiva: Stime, errori standard e i.c al 95%

Parametro	Stima	Errore Standard	I.C Inferiore 95%	I.C Superiore 95%
Intercetta	703,2	52,05	601,06	805,2981
tasso.emi	-1,697	0,2921	-2,2702	-1,1241
tasso.imm	0,0985	0,0502	$5,1262 * 10^{-5}$	0,197
anno	-0,3483	0,0288	-0,3991	-0,2976
tasso.d	-0,133	0,055	-0,2408	-0,0251
pil.reg	-0,0005	0,0001	-0,0008	-0,0002
tasso.d:pil.reg	0,0004	$9,005 * 10^{-5}$	0,0002	0,0005

Tab. 4.22: OLS "first difference" indice di ricambio della popolazione attiva: Stime, errori standard e i.c al 95%

4.6 Adattamento e confronto tra modelli

In questa sezione si vuole capire quale sia la bontà di adattamento ai dati dei modelli presi in considerazione e confrontare tali modelli con il criterio AIC. Per tutti i modelli in analisi si osserva un p-value relativo al test F di Fisher-Snedocor praticamente nullo. Questo risultato porta a preferire ognuno dei modelli ai corrispettivi modelli con sola intercetta. L'unica eccezione è data dal modello "first difference" per l'EMP, il quale, oltre a non presentare alcun parametro significativo, non risulta complessivamente significativo al 95%.

Relativamente alla bontà di adattamento, si osservano degli indici di determinazione R^2 e $R^2_{aggiustato}$ sempre maggiori di 0,8 per il modello ols normale semplice. Questi valori testimoniano un elevato adattamento di questi modelli ai dati osservati. Per quanto riguarda il modello "first difference" invece, i coefficienti di determinazione vanno da un minimo di 0,2 nel modello per il TFT, ad un massimo di 0,7 nel modello per l'indice di ricambio. Anche in questo caso il modello per l'EMP mostra un adattamento quasi nullo ai dati. Generalmente quindi il modello "first difference" si adatta peggio ai dati rispetto al modello ols semplice. Nonostante questo il modello "first difference" risulta interessante in quanto offre una visione differente del fenomeno in analisi. Per confrontare i modelli ols normale semplice con i glm, si fa uso del criterio di informazione AIC. In tutti i casi si vede che, secondo questo criterio, i glm risultano sempre leggermente migliori. Si può quindi pensare che le variabili risposta seguano più fedelmente una determinata distribuzione gamma rispetto ad una gaussiana.

Conclusione

In questo testo si è andati a vedere come affrontare il problema dell'effetto delle migrazioni sulle dinamiche demografiche. Abbiamo quindi presentato due dei possibili approcci per studiare il fenomeno. Tra questi si è applicato il metodo relativo all'effetto delle migrazioni su alcuni semplici indicatori demografici tramite l'utilizzo di modelli di regressione. Per ciascun indicatore si è riusciti a quantificare gli effetti dei tassi di emigratorietà ed immigratorietà sulla determinazione puntuale e non degli indicatori. Chiaramente il metodo utilizzato è uno dei tanti possibili. Ogni metodo infatti presenta dei pregi e dei limiti. Il punto di forza del metodo utilizzato consiste nella facilità applicativa e in una discreta facilità interpretativa (dovendo però non considerare alcune variabili). Presentiamo dunque di seguito quelli che sono invece i maggiori limiti e i suggerimenti di questa trattazione.

5.1 Critiche e suggerimenti

Un valido metodo che si sarebbe potuto utilizzare, consiste nel trattamento dei dati come serie storiche. In questo modo il fattore tempo sarebbe stato tenuto maggiormente sotto controllo. Dei possibili modelli in questo senso sarebbero gli ARIMA e, più specificatamente, dei modelli autoregressivi AR. Una strada percorribile sarebbe stata infatti quella della selezione del miglior modello ARIMA per i dati in analisi. In particolare, si potrebbe applicare un modello ARIMA a ciascuna provincia. Con questo modo di procedere si tiene conto in miglior modo della correlazione tra le osservazioni ad ogni istante t . Un punto su cui si è infatti soffermati sull'analisi è stato quello della necessità di dover introdurre dei diversi modelli per sole finalità interpretative; problema che probabilmente non si sarebbe posto con un modello di tipo ARIMA.

Un aspetto che sarebbe poi interessante valutare è l'effetto delle migrazioni su altri aspetti demografici come la mortalità o la composizione per genere. Poichè i cambiamenti di mortalità avvengono a lungo termine, sarebbe necessario estendere il periodo d'osservazione. Si potrebbero, inoltre, valutare gli effetti delle migrazioni su altri indicatori non trattati in questa sede, o più accuratamente, applicare il metodo presentato nel capitolo 2 ad un caso reale. Lo studio si potrebbe poi estendere al caso delle migrazioni interne (quindi non solo internazionali) per vedere se anche all'interno del territorio nazionale sia presente un impatto demografico dovuto alla mobilità.

5.2 Appendice

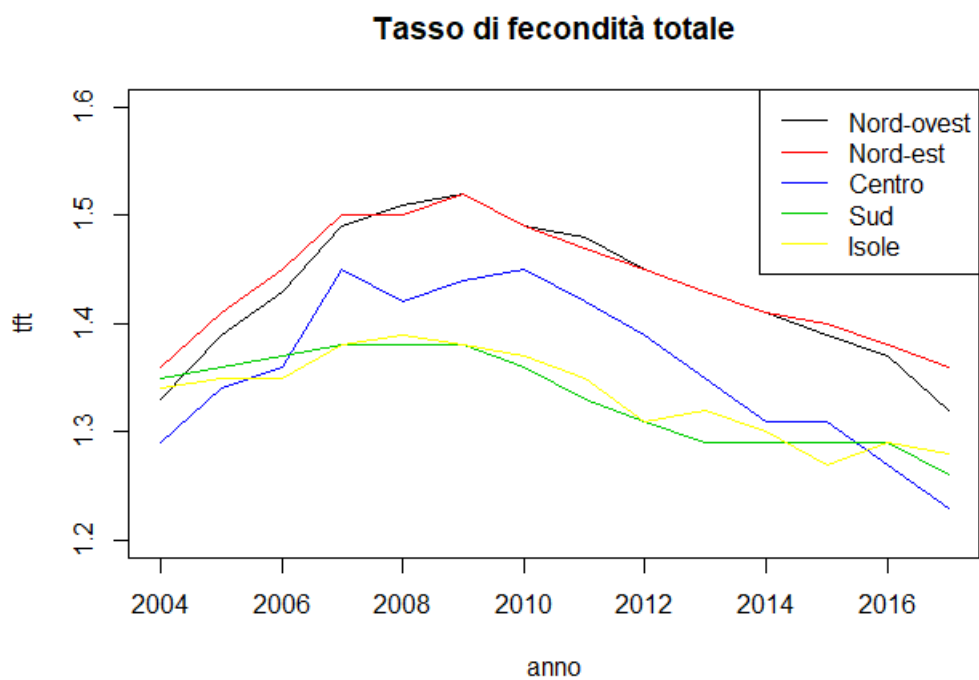


Fig. 5.1: Serie storiche dei tassi di fecondità totale

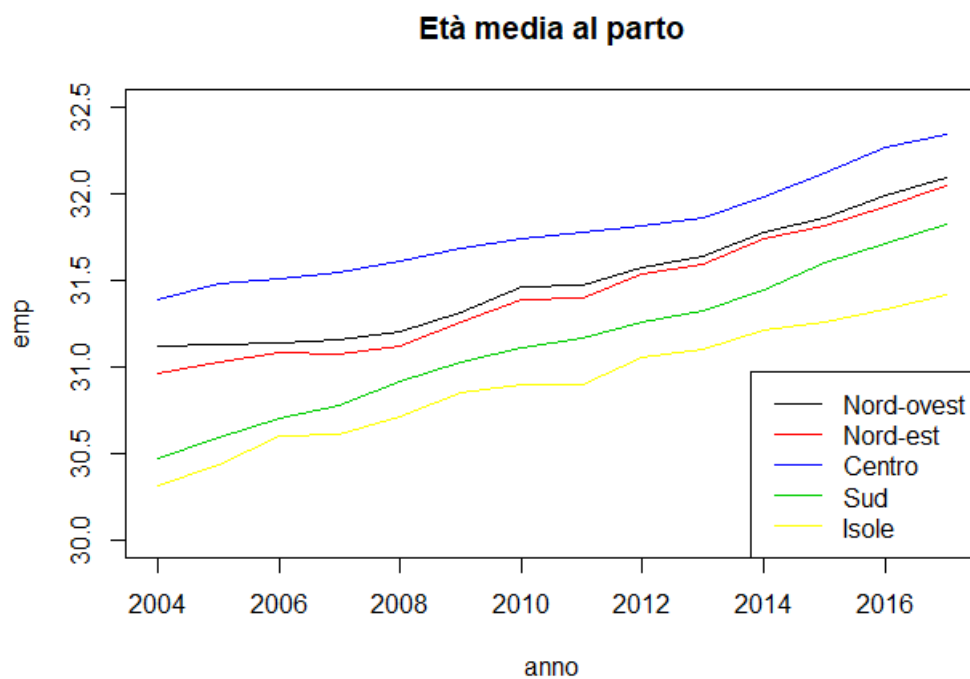


Fig. 5.2: Serie storiche delle età medie al parto

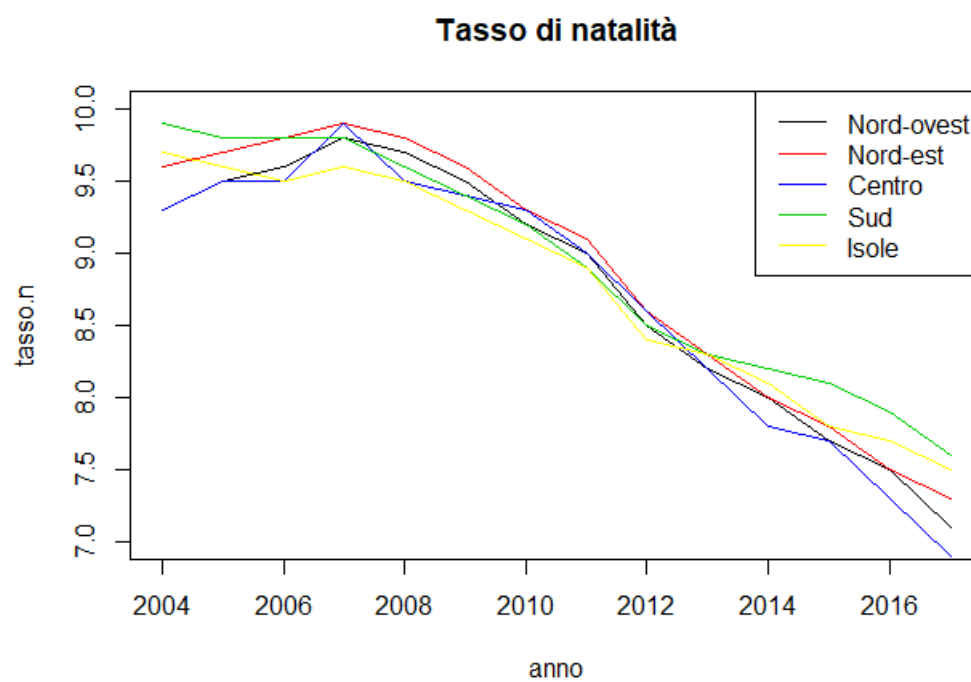


Fig. 5.3: Serie storiche dei tassi di natalità

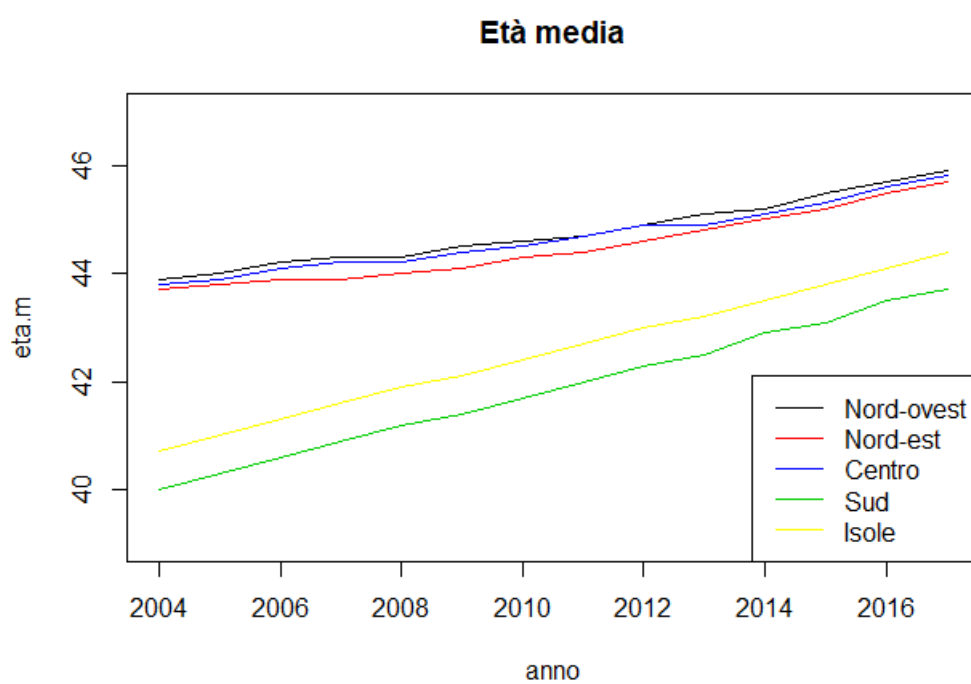


Fig. 5.4: Serie storiche delle età medie

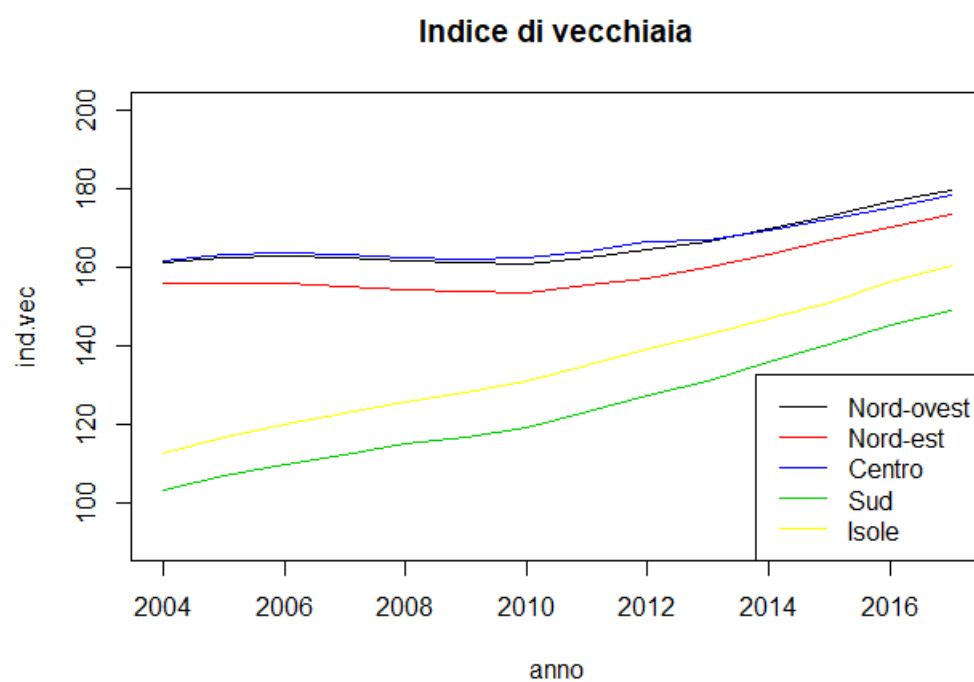


Fig. 5.5: Serie storiche degli indici di vecchiaia

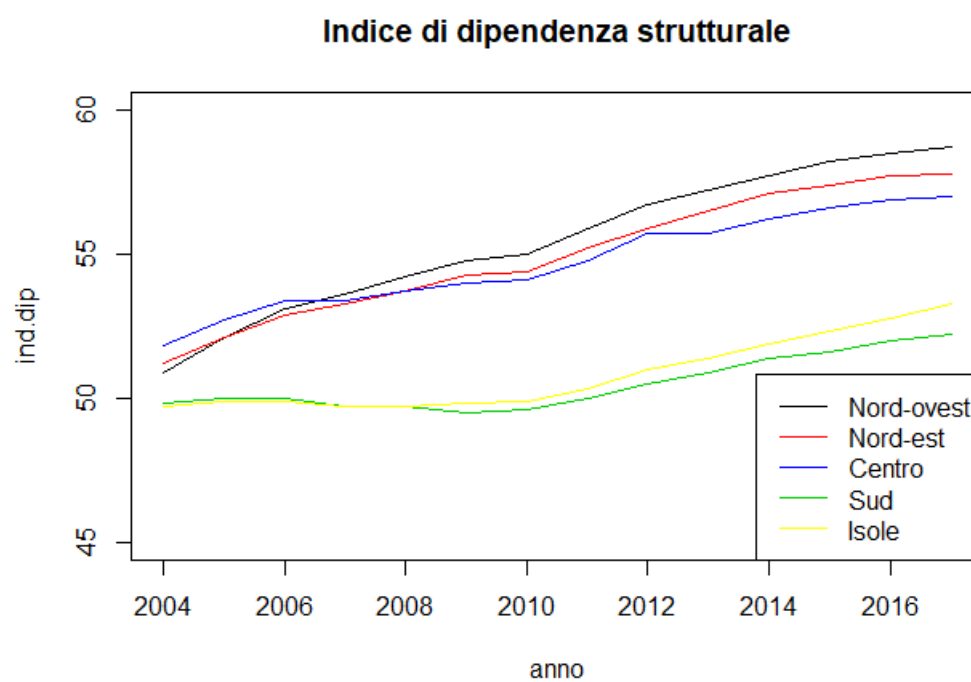


Fig. 5.6: Serie storiche degli indici di dipendenza strutturale

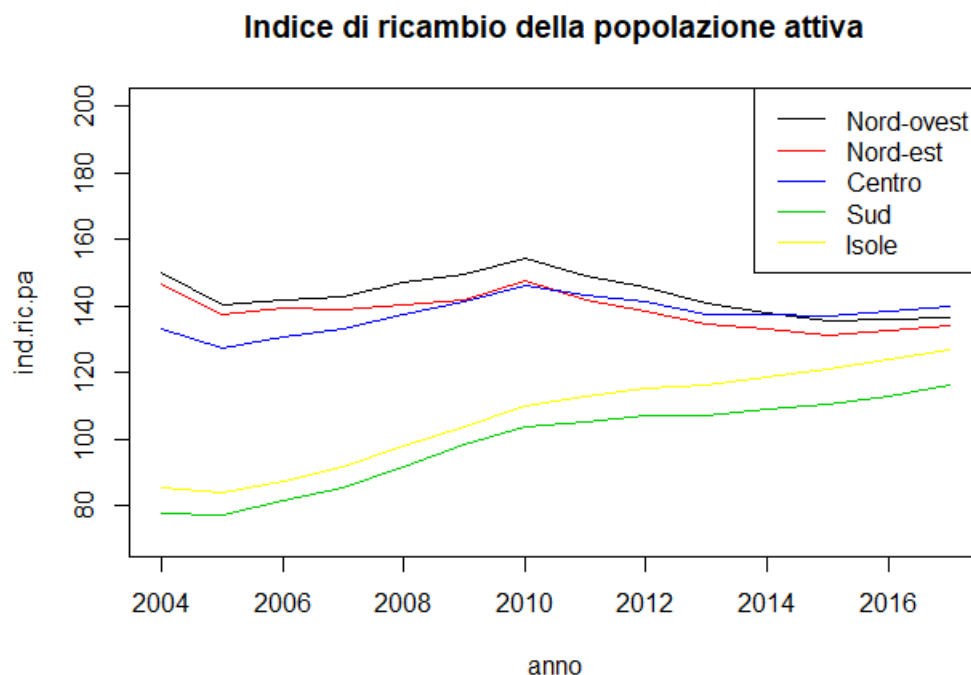


Fig. 5.7: Serie storiche degli indici di ricambio della popolazione attiva

5.3 Bibliografia

Agnieszka Fihel, Anna Janicka, Weronika Kloc-Nowak, 2018. The direct and indirect impact of international migration on the population ageing process: a formal analysis and its application to Poland

Alessandra Salvan, Nicola Sartori, Luigi Pace, 2019. Modelli Statistici 2, Anno Accademico 2019-2020

Anagrafe degli italiani residenti all'estero (AIRE), 2019. Annuario delle statistiche ufficiali del ministero dell'interno

Davide Batzella, 2017. Nuove province in Sardegna, cosa cambia per Sud Sardegna e Area Metropolitana di Cagliari?

Istat, 2016. Glossario. URL: <https://www.istat.it/it/files//2016/01/Glossariodef.pdf>

Istat, 2020. URL: <https://www.dati.istat.it>

-Conti nazionali; Conti e aggregati economici territoriali; Valori pro capite

-Indicatori di fecondità; Principali indicatori

-Migrazioni; Migrazioni (Trasferimenti di residenza); Emigrati - province d'origine

-Migrazioni; Migrazioni (Trasferimenti di residenza); Immigrati - province di destinazione

-Popolazione residente al 1° gennaio; Indicatori demografici

-Popolazione residente al 1° gennaio; Italia, regioni, province

Jean-Michel Lafleur, Mikolaj Stanek, 2017. South-North Migrations of EU Citizens in Times of Crisis

Massimo Anelli, Nicoletta Balbo, 2018. Fertility Drain or Fertility Gain? The Effect of Emigration on Fertility in Italian Municipalities during the Great Recession

Russell King, 2014. Recent Immigration to Italy: Character, Causes and Consequences

Unione Province d'Italia, 2004. URL: <https://www.provinceditalia.it/monza-e-brianza-fermo-e-barletta-andria-trani/>