

Un'analisi gerarchica sull'inattività femminile in Italia

Marco Centra^{*}, Andrea Cutillo^{*}, Valentina Gualtieri⁺

Introduzione

Obiettivo di questo lavoro è lo studio delle determinanti dell'inattività femminile, tenendo in debita considerazione il contesto regionale di appartenenza. Nello specifico si intende individuare quali sono le caratteristiche proprie della donna, della sua famiglia attuale e di quella di origine che maggiormente influenzano la probabilità di essere inattiva, controllando per le peculiarità proprie della regione di appartenenza.

Gli strumenti utilizzati sono i modelli gerarchici o di regressione multilivello, che hanno proprio come principale caratteristica quella di tenere in considerazione la struttura gerarchica dei dati oggetto di studio. Questi modelli, in particolare, offrono la possibilità di considerare le unità oggetto di analisi (le donne) non indipendenti le une dalle altre ma raggruppate implicitamente in unità di secondo livello (le regioni), che si distinguono tra loro per caratteristiche proprie: viene quindi considerata una più complessa struttura di variabilità rispetto a un modello non gerarchico.

Il collettivo di analisi sono le donne coniugate o conviventi che hanno partecipato all'Indagine sui Fattori Determinanti l'Inattività Femminile condotta dall'ISFOL nel 2007 sulle donne italiane nella fascia di età 25-45 anni.

Il presente lavoro è il primo step di un più complesso progetto poiché lo strumento adottato per l'analisi dell'inattività femminile è un modello a componenti di varianza nel quale solo l'intercetta è considerata variabile condizionatamente alla regione di appartenenza. Tale modello suppone, quindi, che le variazioni a livello regionale avvengano solo in considerazione della variazione dell'intercetta e quindi che non ci siano variazioni regionali nei coefficienti delle variabili esplicative utilizzate. In futuro si valuterà non solo l'opportunità di considerare ulteriori o differenti covariate, ma anche la possibilità di stimare un modello che consideri effetti casuali di regione sulle esplicative, per analizzare eventuali differenze regionali anche sulla pendenza, e quindi sull'intensità che le covariate esercitano sull'inattività femminile.

1. Il contesto di riferimento

La Strategia di Lisbona ha definito come uno degli obiettivi più qualificanti per l'Unione Europea e per il nostro Paese, da perseguire entro il 2010, il raggiungimento di un tasso di occupazione femminile delle 15-64enni pari al 60 per cento: oltre a volere in parte rivedere il ruolo sociale della donna, è forte l'idea che per sostenere la crescita economica del Paese è necessario aumentare non solo la produttività, ma anche il tasso di occupazione generale. Mentre il tasso di occupazione maschile italiano, pur inferiore alla media dell'Unione Europea, è elevato (70,7 per cento contro 72,5 per cento nel 2007), quello femminile, anche se in crescita, è invece tra i più bassi in Europa: l'Italia, il cui tasso si attesta al 46,6 per cento si trova largamente al di sotto della media europea

^{*} ISFOL, Analisi e Valutazione delle Politiche per l'Occupazione – m.centra@isfol.it

^{*} ISTAT, Condizioni Economiche delle Famiglie – cutillo@istat.it

⁺ ISFOL, Analisi e Valutazione delle Politiche per l'Occupazione – v.gualtieri@isfol.it

(pari al 58,3 per cento), lontano dall'obiettivo finale dettato dalla strategia di Lisbona ed anche dall'obiettivo intermedio fissato già per il 2005 al 57 per cento.

Come specificato in premessa, il presente lavoro si concentra sulle donne di 25-45 anni in coppia. Le ragioni di questa scelta sono legate sia al fatto che la maggior parte della popolazione femminile di 25-45 anni è in coppia (83,3 per cento), sia al fatto che il fenomeno dell'inattività si manifesta principalmente su questa tipologia di donne. Ben il 28,7 per cento delle donne in coppia, infatti, è inattiva contro il 6,1 per cento registrato per quelle non in coppia (Tav. 1).

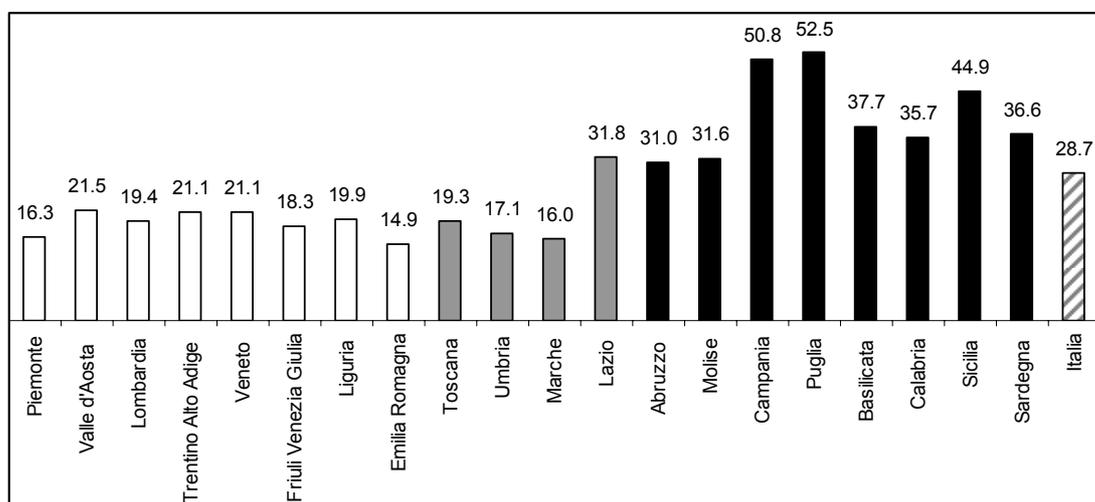
Tav. 1: Donne di 25-45 anni per stato coniugale e condizione lavorativa – Anno 2007

Stato coniugale	Valori in migliaia			% di riga			% di colonna		
	Attive	Inattive	Totale	Attive	Inattive	Totale	Attive	Inattive	Totale
In coppia	5.191.147	2.085.892	7.277.039	71,3	28,7	100,0	78,8	95,8	83,0
Non in coppia	1.398.729	91.082	1.489.811	93,9	6,1	100,0	21,2	4,2	17,0
Totale	6.589.875	2.176.975	8.766.850	75,2	24,8	100,0	100,0	100,0	100,0

Fonte: Indagine Isfol sui Fattori Determinanti dell'Inattività Femminile – Anno 2007

Per questa sottopopolazione si osserva una polarizzazione a livello territoriale simile a quella esistente per la popolazione femminile di 15-64 anni. Nel Nord il tasso di inattività si attesta infatti al 18,6 per cento, al Centro il valore è pari 24,1 per cento e nel Mezzogiorno si raggiunge il 45,3 per cento. L' Emilia Romagna è la regione con il più basso tasso di inattività (14,9 per cento) ed ad essa si contrappongono la Puglia con il tasso più elevato (52,5 per cento) e la Campania (50,8 per cento).

Fig. 1: Tasso di inattività¹ per le donne in coppia di 25-45 anni per regione - Anno 2007



Fonte: Indagine Isfol sui Fattori Determinanti dell'Inattività Femminile – Anno 2007

Diverse possono essere le determinanti dell'inattività femminile italiana. Tra queste, si possono citare anzitutto le preferenze individuali e le necessità imposte da

¹ Nell'indagine Isfol sono escluse dalla rilevazione le donne di 25-45 anni studentesse, ritirate dal lavoro o in altre condizioni. La stima è quindi fatta su una popolazione ridotta e sono rilevate come inattive soltanto le casalinghe.

particolari situazioni, quali la cura della casa e della famiglia (attività che ricadono generalmente sulle donne piuttosto che sugli uomini, ISTAT, 2007) o la situazione economica della famiglia. Sono poi da considerare sia l'atteggiamento del partner verso l'occupazione femminile che l'influenza della famiglia di origine: la condizione occupazionale della madre ha, ad esempio, un effetto considerevole sulle future probabilità di lavorare delle figlie una volta che queste raggiungono l'età lavorativa.

Tra i fattori che spingono l'Italia così in alto nel tasso di inattività femminile ci sono inoltre effetti particolarmente importanti determinati dalla situazione del mercato del lavoro "locale" in cui ci si muove. E' vero, infatti, che dove il gap tra domanda ed offerta di lavoro è più elevato, il divario tra le possibilità occupazionali degli uomini e quelle delle donne aumenta, facendo sì che una consistente parte di queste ultime non si offra neppure sul mercato, nella convinzione di non trovare comunque lavoro o una remunerazione che corrisponda alle proprie aspettative.

Un altro tratto peculiare dell'Italia, rispetto a molti altri Paesi europei, è che le donne spesso rinunciano a rientrare nel mercato del lavoro dopo la maternità. Anche questa decisione può essere determinata da particolari situazioni locali: dove è minore l'offerta di strutture per i primi anni di vita del bambino, quali gli asili nido o le scuole materne, è possibile che una donna esca dal mercato del lavoro per assistere i figli senza avere più l'opportunità o la determinazione di rientrarvi dopo alcuni anni di inattività.

E' infine da considerare un importante fattore culturale, ancora relativo al contesto in cui si vive e ci si muove: l'atteggiamento femminile verso l'inattività lavorativa è infatti differente non solo da persona a persona e da famiglia a famiglia, ma può anche esserlo da territorio a territorio, con le regioni del Centro-Nord molto più chiuse a questa possibilità rispetto alle regioni del Mezzogiorno.

2. Il modello

In questo lavoro si utilizzano modelli gerarchici (o di regressione multilivello, Goldstein, 1995; Snijders e Bosker, 1999) che hanno come principale caratteristica quella di tenere in considerazione la struttura gerarchica dei dati oggetto di studio. Le unità su cui si rileva il fenomeno oggetto di studio (unità di primo livello, le donne) risultano naturalmente aggregate in gruppi differenti (unità di secondo livello, le regioni). In un contesto di struttura gerarchica dei dati, ove i legami tra il fenomeno e le variabili che lo descrivono dipendono anche dalla natura dei gruppi delle unità statistiche, è utile che le analisi siano condotte in modo da tenere conto di tale caratteristica della popolazione indagata. L'analisi non gerarchica assume che ci sia indipendenza tra le osservazioni e ciò può risultare non vero nel caso in cui si ha una implicita struttura in gruppi: non considera quindi l'esistenza di una più complessa struttura di variabilità. Quando infatti ci si trova in presenza di unità statistiche annidate in unità di livello gerarchicamente superiore, la variabilità complessiva si genera dalle due fonti, *between groups* e *within groups*; in altri termini, le unità statistiche appartenenti a uno stesso gruppo sono soggette a comportamenti simili a causa delle caratteristiche del contesto che le raggruppa. La stima di modelli di regressione multilivello per l'analisi della dipendenza provvede quindi all'integrazione tra le dimensioni micro e macro, evidenziando e studiando le relazioni esistenti tra l'individuo e il contesto territoriale in cui vive: il processo in opera è quindi funzione sia di caratteristiche individuali che del gruppo cui l'individuo appartiene. Essendo la

variabile di studio una variabile binaria (attività vs inattività), nel nostro caso si è scelto di utilizzare un modello di regressione logistica a due livelli.

Il modello utilizzato è un modello a componenti di varianza per variabili di risposta dicotomiche (Anderson e Aitkin, 1985), nel quale il legame tra variabile dipendente e variabili esplicative si basa sull'ipotesi che le caratteristiche socio-economiche locali siano una sovrastruttura che condiziona sia le variabili di primo livello, sia le scelte individuali relative alla variabile oggetto di studio. Nel modello solamente l'intercetta è considerata variabile. Si suppone, dunque, che le diverse situazioni regionali siano associate a variazioni nell'intercetta, mentre non ci sono variazioni nei coefficienti delle altre variabili che rappresentano il legame con l'inattività sul mercato del lavoro.

Nello specifico, l'interesse consiste nel valutare da un lato le probabilità individuali di essere inattive sul mercato del lavoro, dall'altra il contributo specifico delle regioni di residenza a tale probabilità. La strategia di analisi si sviluppa in tre passi.

1) Il primo passo prevede la determinazione dei parametri del modello:

$$Y_{ir} = \alpha + u_r + e_i \quad (1)$$

$Y_{ir} = 1$ se la generica donna i della regione r non partecipa attivamente al mercato del lavoro ($Y_{ir}^* \geq 0$);

$Y_{ir} = 0$ se la generica donna i della regione r partecipa attivamente al mercato del lavoro ($Y_{ir}^* < 0$).

L'asterisco rappresenta l'utilità individuale, non osservabile, ad essere inattive sul mercato del lavoro.

e_i è il parametro che rappresenta la variabilità residua dovuta agli individui, supposto $iid(0, \sigma_e^2) \forall i, r$;

u_r è il parametro che rappresenta la variabilità residua dovuta al contesto, supposto $iid(0, \sigma_u^2) \forall r$.

Un modello del genere, senza variabili esplicative, è detto modello nullo, utile per confermare l'esistenza di un effetto contesto; quando infatti non si controlla per alcuna covariata, le differenze tra le donne residenti in diverse regioni, dovute sia a caratteristiche di primo che di secondo livello, sono completamente descritte dai termini u_r , ossia dagli effetti casuali relativi alle unità di secondo livello. Gli u_r , che costituiscono un elemento distintivo dei modelli gerarchici, esprimono, dunque, l'effetto residuo esercitato da ciascuna regione nei riguardi della variabile risposta. Tali stime possono essere interpretate come misura del contributo unico di ciascun gruppo e fanno sì che l'intercetta vari al variare della regione.

2) Il secondo passo prevede l'introduzione nel modello delle sole variabili che descrivono il comportamento dell'individuo (variabili di primo livello), al fine di

esplicitare l'effetto del contesto condizionatamente a caratteristiche individuali osservabili, e l'espressione (1) si modifica nel modo seguente:

$$Y_{ir} = \alpha + \mathbf{x}_i' \boldsymbol{\beta} + u_r + e_i \quad (2)$$

nel quale \mathbf{X} è il vettore delle variabili esplicative relative agli individui e $\boldsymbol{\beta}$ il relativo vettore di parametri.

3) Il passo finale riguarda l'introduzione nel modello delle variabili relative ai gruppi (variabili di secondo livello):

$$Y_{ir} = \alpha + \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_r' \boldsymbol{\theta} + u_r + e_i \quad (3)$$

nel quale \mathbf{Z} è il vettore delle variabili esplicative relative ai gruppi e $\boldsymbol{\theta}$ il relativo vettore di parametri.

Il modello completo permette così di valutare l'effetto delle covariate di primo e di secondo livello, nonché, tramite gli effetti casuali di secondo livello (u_r), l'effetto residuo esercitato dalle singole regioni.

L'analisi degli u_r permette di determinare l'effetto di ciascuna regione sulle probabilità di inattività fornendo una sorta di graduatoria tra regioni: la stima dell'effetto casuale senza esplicative determina, infatti, l'effetto complessivo del territorio di analisi; stimando il modello con sole esplicative di primo livello si ottiene l'effetto delle singole regioni sulle scelte individuali una volta che si controlla per il fatto che le donne all'interno di una stessa regione abbiano caratteristiche individuali più o meno simili; stimando infine il modello con esplicative anche di secondo livello si ottiene l'effetto delle singole regioni sulle scelte individuali depurato dalle caratteristiche del mercato del lavoro e dalla disponibilità di strutture di sostegno. Nell'ipotesi "ideale" che le esplicative utilizzate siano perfettamente descrittive della situazione dei mercati del lavoro regionali e delle possibilità di accesso a strutture di sostegno nella cura dei figli e degli anziani la variabilità residua sarebbe descrittiva della componente culturale di atteggiamento generale verso l'attività/inattività femminile dei diversi territori.

Poiché i termini e_i e u_r sono supposti tra loro indipendenti, le utilità individuali risultano indipendenti, condizionatamente al gruppo e alle esplicative utilizzate, mentre risulta:

$$\begin{aligned} \text{Var}(Y_{ir} | \mathbf{X}_i, \mathbf{Z}_r) &= \sigma_u^2 + \sigma_e^2 \\ \text{Covar}(Y_{ir}, Y_{i'r'} | \mathbf{X}, \mathbf{Z}) &= \begin{cases} 0 & \text{per } r \neq r' \\ \sigma_u^2 & \text{per } r = r' \end{cases} \end{aligned}$$

Ciò equivale a dire che le utilità individuali delle donne appartenenti alla stessa regione sono positivamente correlate.

Per mettere in evidenza come le probabilità di essere inattive delle donne appartenenti allo stesso gruppo siano, o meno, positivamente correlate si usa il coefficiente di correlazione intra-classe che assume la seguente forma:

$$\rho = \frac{\sigma_u^2}{\sigma_u^2 + \sigma_e^2} \quad (4)$$

Il valore di tale coefficiente è tanto più elevato quanto più i gruppi sono diversi tra loro; rappresenta infatti quanta parte della variabilità del fenomeno oggetto di studio sia spiegata dall'appartenenza ad aree territoriali differenti. L'ipotesi che σ_u^2 sia nullo equivale all'ipotesi che non ci siano differenze tra gruppi.

Utilizzando la riparametrizzazione di Rampichini (1992), i parametri α , β e θ vengono espressi in unità di σ_e^2 . In tal modo, la 3, da stimare tramite un modello logit, diventa:

$$Y_{ir} = \alpha + \mathbf{x}_i' \boldsymbol{\beta} + \mathbf{z}_r' \boldsymbol{\theta} + \gamma u_r + e_i$$

dove $\gamma = \sigma_u / \sigma_e$.

Sotto queste ipotesi i coefficienti e_i ed u_r si distribuiscono

$$\begin{aligned} (e_i | \mathbf{X}_i, \mathbf{Z}_r, u_r) &\sim N(0,1) - IID \quad \forall i, r \\ u_r &\sim N(0,1) - IID \quad \forall r \end{aligned}$$

Inoltre, il coefficiente di correlazione intraclasse diventa

$$\rho = \frac{\gamma^2}{1 + \gamma^2}$$

Tramite la stima del coefficiente di correlazione intraclasse, è possibile vedere quanta parte della variabilità del fenomeno è dovuta a caratteristiche di contesto: il valore di tale coefficiente è tanto più elevato quanto più i gruppi sono diversi tra loro; il coefficiente di correlazione intraclasse misura infatti la proporzione di variabilità dovuta all'effetto di raggruppamento e, quindi, di dipendenza tra le osservazioni annidate in unità dello stesso livello. Fornisce così una misura complessiva del grado di omogeneità tra osservazioni appartenenti allo stesso gruppo. Attraverso la diminuzione della variabilità complessiva degli effetti casuali di secondo livello passando dal modello con esplicative di primo livello al modello completo, è possibile determinare quanta parte della variabilità di contesto sia dovuta alle esplicative utilizzate e quanta sia quella residua.

3. I dati e le variabili utilizzate

I dati utilizzati per realizzare le analisi provengono dall'Indagine sui Fattori Determinanti l'Inattività Femminile condotta dall'ISFOL nel 2007, integrata con informazioni relative al contesto territoriale di fonte ISTAT. La rilevazione condotta

dall'ISFOL prevede un campione per quote di 6.000 donne, stratificato secondo le principali variabili sociodemografiche; le stime sono prodotte grazie ad uno stimatore calibrato.

La popolazione oggetto di studio dell'indagine è costituita dalle donne italiane nella fascia di età 25-45 anni, che include la quota maggiore di donne attive; sono state volutamente escluse dall'indagine sia le donne molto giovani, che spesso non possono essere effettivamente collocate nell'area delle attive/inattive poiché probabilmente ancora in fase di formazione oppure attive ma con una posizione/condizione professionale non ben definita, e le donne con più di 45 anni che non sono più in età di "scelta"; la loro condizione, infatti, è ormai generalmente consolidata.

Nel presente lavoro ci limitiamo inoltre allo studio delle donne coniugate/conviventi. Questa scelta, già motivata in precedenza, è dovuta sia dal fatto che la maggior parte delle donne nella fascia di età in questione sono in coppia, sia dal fatto che l'essere o meno in coppia risulta fattore discriminante sull'incidenza dell'inattività. E' lecito avanzare l'ipotesi che molte delle motivazioni che possono spingere una donna in coppia a non lavorare siano legate, oltre che ai maggiori impegni derivanti dalle responsabilità familiari, al coniuge/convivente, ossia alla sua condizione, alle sue caratteristiche individuali ed al suo parere sul ruolo della donna nella famiglia e nella società.

Le variabili che vengono utilizzate nel modello sono distinte in variabile dipendente, variabili di primo livello e variabili di secondo livello.

La variabile dipendente è una dummy che registra se la donna è attiva sul mercato del lavoro (occupata o in cerca di occupazione) o inattiva (casalinga). Come già evidenziato, vengono escluse le donne impegnate in percorsi di istruzione, che ancora non hanno cominciato la propria carriera lavorativa.

3.1 Variabili di primo livello (individuali)

Tra le esplicative di primo livello, ci sono l'età, il titolo di studio e il numero di componenti in famiglia; l'esistenza di percettori di reddito da lavoro o pensione in famiglia (ad esclusione della donna stessa), per verificare se la presenza di altri introiti monetari nella famiglia possa non incentivare l'offerta femminile sul mercato del lavoro. Un'informazione molto importante è relativa alla presenza in famiglia di figli, che è stata divisa in: presenza di figli non in età dell'obbligo scolastico che frequentano un asilo nido o una scuola dell'infanzia; presenza di figli non in età dell'obbligo per i quali non si ricorre ad alcuna forma di assistenza da servizi formali; presenza di figli tra i 6 e i 14 anni, fascia di età in cui i bambini e i ragazzi ancora necessitano di una assidua presenza della madre; presenza di figli di 15 anni e più. Analogamente, si è considerata la presenza in famiglia di anziani non autosufficienti, suddivisi in assistiti o non assistiti da strutture pubbliche o private.

Avendo ristretto il collettivo di studio alle coniugate/conviventi abbiamo potuto utilizzare nel modello anche informazioni relative al coniuge/convivente: tra queste, il titolo di studio, la condizione e posizione nella professione; relativamente a questo ultimo punto, è interessante verificare se il fatto che il coniuge/convivente sia un lavoratore autonomo possa incrementare l'incertezza sugli introiti futuri, favorendo di conseguenza l'attività femminile. Sono state inoltre utilizzate ulteriori due informazioni: una relativa all'aiuto che il coniuge/convivente fornisce alla donna nella gestione della famiglia e della casa, l'altra relativa all'opinione di questo sul fatto se sia normale che le

donne lavorino. In entrambi i casi, l'ipotesi è che la risposta affermativa renda più probabile la partecipazione femminile attiva al mercato del lavoro.

Per studiare l'effetto della famiglia di origine sui percorsi individuali femminili, ipotizzando che vi sia una stretta relazione tra le azioni che una donna intraprende in età adulta con le caratteristiche della famiglia d'origine e l'ambiente nei quali si è formata in età giovanile, si utilizzano anche il più alto titolo di studio dei genitori, come proxy del livello socio-economico della famiglia di origine, la condizione prevalente della madre (con l'ipotesi che una madre prevalentemente occupata nel corso della sua vita sia di spinta all'attività della figlia) e un'informazione relativa al fatto che le donne che frequentavano la famiglia dell'intervistata quando questa era bambina erano prevalentemente lavoratrici.

3.2 Variabili di secondo livello (regionali)

Le variabili di secondo livello utilizzate sono state scelte con l'obiettivo di controllare per la situazione del mercato del lavoro e la disponibilità di servizi assistenziali a livello regionale. L'ipotesi è che dove la situazione del mercato del lavoro femminile è sfavorevole, alcune donne possano non ritenere conveniente entrare attivamente sul mercato del lavoro e che dove maggiori sono i servizi di assistenza ai bambini ed adulti non autosufficienti, maggiori sono invece le probabilità di partecipazione femminile.

Le variabili di contesto utilizzate sono: tasso di disoccupazione delle persone con 16 anni o più (fonte: Istat, Rilevazione Continua sulle Forze di Lavoro, anno 2007); tasso di occupazione delle persone con 16 anni o più (fonte: Istat, Rilevazione Continua sulle Forze di Lavoro, anno 2007); differenziale di genere nel tasso di disoccupazione delle persone con 16 anni o più (fonte: Istat, Rilevazione Continua sulle Forze di Lavoro, anno 2007); indice di copertura territoriale di asili nido (per 100 persone) (fonte: Istat, Indagine censuaria sugli interventi e i servizi sociali dei comuni, anno 2005); quota di anziani non autosufficienti e disabili in strutture assistenziali (ogni 10.000 anziani) (fonte: Istat, L'assistenza residenziale e socio-assistenziale in Italia, anno 2004). Questo ultimo indicatore può anche essere visto come un indicatore di offerta assistenziale, dato che i posti disponibili sono sempre occupati, a meno di fisiologiche rotazioni o sostituzioni nell'utenza. E' stato inoltre considerato il differenziale percentuale di genere nel reddito netto da lavoro individuale (fonte: Istat, Indagine EUSILC sui Redditi e le Condizioni di Vita, anno 2006).

Le maggior parte delle variabili regionali sono state sintetizzate, tramite l'Analisi in Componenti Principali (ACP), in un'unica variabile che è stata successivamente inserita nel modello gerarchico. Partendo dalle relazioni tra un determinato numero di variabili rilevanti su n soggetti, l'ACP permette, infatti, di individuare eventuali dimensioni soggiacenti alle relazioni stesse, che riassumono l'interdipendenza tra le variabili originali. L'ACP individua quindi nuove variabili sintetiche, chiamate fattori, che formano un nuovo sottospazio e che sono: indipendenti tra loro; combinazioni lineari delle variabili iniziali; spiegano, ognuno in ordine decrescente, il massimo della variabilità della matrice originale dei dati. Il primo fattore è quindi la migliore approssimazione della matrice dei dati di partenza, il secondo è quello che presenta la seconda migliore approssimazione e così via. Il fatto che i nuovi fattori sono indipendenti tra di loro evita il problema della multicollinearità.

L'ACP è stata applicata alle variabili elencate, ad eccezione del differenziale nel reddito da lavoro², ed ha fornito un primo fattore che spiega più dell'85 per cento della variabilità totale. Questo elevato valore autorizza ad utilizzare nelle analisi successive solo il primo fattore ottenuto con l'ACP, che è positivamente correlato con il tasso di occupazione, l'indice di copertura territoriale degli asili e con la quota di anziani non autosufficienti e disabili in strutture e negativamente correlato con il tasso di disoccupazione e con il differenziale di genere nel tasso di disoccupazione.

Il primo fattore può essere quindi identificato come un indice "di bontà" del territorio poiché al crescere di questo aumentano il tasso di occupazione e la disponibilità di strutture assistenziali mentre diminuisce il tasso di disoccupazione e la disparità tra genere sempre nel tasso di disoccupazione. Nel modello, oltre alla variabile sintetica, è stata quindi inserita anche la variabile che misura il differenziale nel reddito netto da lavoro individuale.

4. Risultati del modello

Come esplicitato nel paragrafo 2, la stima del modello avviene in tre passi: si stima il modello nullo, a sole intercette casuali, successivamente vengono inserite le variabili esplicative di livello individuale e infine vengono inserite anche le informazioni regionali.

Il primo passo da compiere nell'analisi consiste quindi nello scindere la variabilità totale dell'inattività femminile nelle due fonti, individuale e di contesto, tramite il modello nullo a componenti di varianza senza covariate.

La stima del parametro γ^3 è pari a 0,2868; (γ^2), che corrisponde ad una stima del rapporto tra variabilità di gruppo e variabilità individuale non osservata, è pari a 0,0823; il coefficiente di correlazione intra-classe (ρ) risulta pari a 0,076, il che significa che il 7,6 per cento della variabilità complessiva del fenomeno è spiegato dalla diversa regione di residenza. Il valore di ρ non molto elevato potrebbe essere imputato al fatto che la regione di residenza nasconda una forte eterogeneità sub-regionale.

Tav. 2: Modello multilivello nullo, stima della probabilità di inattività

	Coeff.	S.E	TEST T	P value
Intercetta	0,4412	0,09	5,10	0,00
γ	0,2868	0,06		
ρ	0,0760			
Numero osservazioni	4.772			

² Questa variabile è stata inserita nell'analisi fattoriale ma, essendo risultata poco correlata con le altre, non permette una chiara identificazione e lettura delle componenti principali. E' stato quindi ritenuto opportuno inserirla nel modello come variabile singola.

³ Per valutare la significatività del coefficiente relativo al parametro γ , Anderson e Aitkin (1985) hanno suggerito di studiare il profilo di verosimiglianza relativa tramite il test del rapporto di verosimiglianza tra modello con e senza effetti casuali: questo è risultato significativo, indicando che effettivamente la regione di residenza della donna influenza l'attività/inattività sul mercato del lavoro.

Una volta che vengono introdotte le variabili esplicative sia di primo che di secondo livello il parametro γ passa ad un valore di 0,1055. Il coefficiente di correlazione intraclasse, risulta ora pari a 0,011 contro lo 0,076 del modello nullo⁴. L'introduzione delle covariate di secondo livello relative al mercato del lavoro e alla disponibilità di servizi assistenziali rende quasi nullo l'effetto del raggruppamento regionale, indicando che le informazioni utilizzate sono fortemente caratterizzanti dell'effetto regionale sul fenomeno dell'inattività.

Relativamente ai risultati ottenuti per le singole esplicative di primo livello, il titolo di studio ha, come atteso, una forte influenza sulla probabilità di non partecipare attivamente al mercato del lavoro; questa decresce in maniera significativa al crescere del titolo di studio, ed è particolarmente vero per le donne con almeno un titolo universitario. Inoltre, pur essendo l'analisi limitata ad un collettivo di donne di classe di età ristretta (le 25-45enni), si rilevano differenze nella probabilità di essere inattive dovute all'età: risulta infatti che le donne di 41-45 anni hanno maggiori probabilità di essere inattive rispetto alle classi di età più giovani. Dalle stime ottenute si osserva peraltro che rispetto alla classe delle 41-45enni, la classe che ha maggiore probabilità di avere donne attive al suo interno è quella più giovane, delle 25-30enni. Questo risultato può essere dovuto anche alla selezione a priori del collettivo utilizzato, nel quale come inattive sono considerate le sole casalinghe, meno frequenti quando le donne giovani, pur coniugate, non hanno generalmente ancora figli. Relativamente a questo ultimo punto, l'aver figli ha generalmente un effetto positivo sull'inattività femminile. Questo effetto è particolarmente rilevante quando i figli sono molto piccoli (0-5 anni), sia che questi frequentino sia che non frequentino un asilo nido o una scuola dell'infanzia. Il coefficiente relativo alla dummy che indica la presenza in casa di figli di 14 anni e più, pur con il segno positivo, non risulta comunque statisticamente differente da 0. Questo significa che è possibile che le madri che erano al di fuori dell'offerta di lavoro per accudire i figli una volta che questi crescono possono in alcuni casi rientrarvi. Non risulta invece avere un effetto sull'attività/inattività la presenza in famiglia di anziani o disabili non autosufficienti. Passando all'analisi di alcune caratteristiche relative al coniuge/convivente si osserva che all'aumentare del titolo di studio di questo aumenta la probabilità per le donne di non partecipare al mercato del lavoro. Non risulta invece avere effetto, a parità di altre condizioni, la condizione e la posizione nella professione del coniuge. Se l'intervistata dichiara di essere aiutata nei lavori domestici e di cura dei figli dal coniuge o convivente la donna ha minore probabilità di essere inattiva. Lo stesso discorso vale quando l'aiuto nella gestione della casa o nella cura dei figli viene da parenti, colf, badanti, baby sitter: anche in questo caso la probabilità di essere inattive diminuisce considerevolmente, anche se non si sa se come causa o come conseguenza dell'aiuto ricevuto. Infine, riguardo alla famiglie di origine, il fatto che le figure femminili frequentanti la famiglia della donna quando questa era bambina fossero prevalentemente lavoratrici è un esempio per le scelte della donna in età adulta, avendo

⁴ L'introduzione delle sole esplicative di primo livello ha fatto invece diminuire leggermente il coefficiente di correlazione intraclasse, che arriva al valore di 5,03 per cento: poiché l'effetto dell'introduzione di variabili di primo livello sulla variabilità di secondo livello non è prevedibile a priori (Longford, 1993), questo sta a significare che all'interno di una stessa regione le donne hanno caratteristiche individuali e familiari più simili tra loro rispetto alle donne delle altre regioni. Tale relativa omogeneità nelle caratteristiche individuali non ha comunque eliminato l'esistenza di diversi effetti regionali.

un impatto negativo sulla probabilità di essere inattiva, mentre non risulta statisticamente significativo il fatto che la madre fosse prevalentemente lavoratrice nel corso della sua vita in età da lavoro.

Per le covariate di secondo livello, mentre il differenziale di genere nel reddito da lavoro non mostra avere effetti sulle probabilità individuali di essere inattive, la variabile di sintesi, definita come indicatore di “bontà” del territorio verso il lavoro femminile, risulta avere un forte impatto negativo.

Tav. 3: Modello multilivello con covariate di primo livello, stima della probabilità di inattività

	Coeff.	S.E	TEST T	P value
Intercetta	1,1504	0,35	3,29	0,00
Classe di età (Base=41-45 anni)				
<i>25-30 anni</i>	-0,5925	0,11	-5,37	0,00
<i>31-35 anni</i>	-0,1752	0,08	-2,25	0,04
<i>36-40 anni</i>	-0,2821	0,07	-4,10	0,00
Presenza di percettori ad esclusione della donna	-0,4055	0,07	-5,53	0,00
Presenza di figli 0-5 anni che vanno all'asilo/scuola dell'infanzia	0,2858	0,08	3,57	0,00
Presenza di figli 0-5 anni che non vanno all'asilo/scuola dell'infanzia	0,2071	0,07	3,08	0,01
Presenza di figli 6-14 anni	0,1026	0,06	1,86	0,08
Presenza di figli 14 anni o più	0,0836	0,07	1,13	0,27
Presenza di anziani non autosufficienti o disabili affidati a strutture pubbliche/private	-0,0491	0,39	-0,13	0,90
Presenza di anziani non autosufficienti o disabili non affidati a strutture pubbliche/private	-0,1707	0,12	-1,43	0,17
Titolo di studio (Base=Al massimo scuola secondaria di primo grado)				
<i>Scuola secondaria di secondo grado</i>	-0,4477	0,06	-7,34	0,00
<i>Università</i>	-2,1323	0,11	-19,08	0,00
Titolo di studio del coniuge (Base=Al massimo scuola secondaria di primo grado)				
<i>Scuola secondaria di secondo grado</i>	0,5081	0,06	7,92	0,00
<i>Università</i>	0,7035	0,09	8,02	0,00
Condizione e posizione nella professione del coniuge/convivente (Base=Dip. a tempo indeterminato)				
<i>Dipendente a tempo determinato</i>	-0,1215	0,10	-1,24	0,23
<i>Autonomo</i>	0,0883	0,06	1,51	0,15
<i>Non occupato</i>	0,0457	0,12	0,38	0,71
Si avvale di aiuti nella gestione del lavoro domestico e nella cura dei familiari	-0,2943	0,07	-4,17	0,00
Massimo titolo di studio dei genitori (Base=Al massimo scuola primaria)				
<i>Scuola secondaria di primo grado</i>	-0,1466	0,06	-2,40	0,03
<i>Scuola secondaria di secondo grado</i>	-0,1120	0,08	-1,47	0,16
<i>Universitario</i>	0,0203	0,12	0,16	0,87
Madre "prevalentemente" lavoratrice	-0,0409	0,05	-0,77	0,45
La maggior parte delle donne con cui era in contatto da piccola lavoravano	-0,1516	0,06	-2,73	0,01
Il coniuge/convivente ritiene che sia normale che una donna lavori	-0,5002	0,10	-4,84	0,00
Il coniuge/convivente collabora al lavoro domestico e la cura dei figli	-0,2308	0,05	-4,58	0,00

Differenziale (%) di genere nel reddito netto da lavoro	0,0048	0,01	0,42	0,68
Primo fattore ACP (indice di “bontà” del territorio)	-0,2353	0,05	-4,44	0,00
γ	0,1055	0,04		
ρ	0,0110			
Numero di osservazioni	4.772			

4.1 Gli effetti casuali di secondo livello

Come detto, un interessante spunto di analisi tipico dei modelli multilivello è rappresentato dallo studio degli effetti casuali relativi alle unità di secondo livello: gli u_r rappresentano infatti l’effetto esercitato sulle probabilità individuali dalla r-sima regione, una volta controllato per l’effetto di tutte variabili esplicative, cosa che permette di studiare l’impatto di ciascuna regione sulle probabilità individuali di presentare l’una o l’altra determinazione della variabile dipendente. In questa maniera è peraltro possibile confrontare tra loro le varie regioni, creando una sorta di “graduatoria” tra regioni⁵.

Tav.4: Effetti casuali di secondo livello ottenuti tramite modello nullo, modello con covariate di primo livello e modello con covariate di primo e secondo livello

	Modello nullo		Modello con covariate di primo livello		Modello con covariate di secondo livello	
	Stima	E.S.	Stima	E.S.	Stima	E.S.
Piemonte	-1.2727*	0.264	-1.4158*	0.347	-1.1980**	0.628
Val d’Aosta	-0.3275	0.567	-0.5462	0.674	0.1960	0.891
Lombardia	-0.7751*	0.202	-0.5548**	0.275	0.2961	0.528
Trentino Alto Adige	-0.3418	0.292	-0.8751**	0.383	0.2867	0.668
Veneto	-0.9039*	0.237	-1.0471*	0.312	-1.0261**	0.583
Friuli Venezia Giulia	0.5946**	0.309	0.1621	0.414	1.0569	0.699
Liguria	0.0951	0.263	0.1101	0.353	0.7827	0.633
Emilia Romagna	-1.3112*	0.260	-1.0237*	0.348	-0.2423	0.628
Toscana	-0.5219*	0.227	-0.0691	0.307	0.5565	0.574
Umbria	0.3205	0.330	0.3741	0.430	0.4745	0.719
Marche	-0.1795	0.322	0.3112	0.419	1.0842	0.706
Lazio	0.6343*	0.254	1.2274*	0.343	1.8651*	0.617
Abruzzo	1.3142*	0.303	1.2566*	0.404	1.4280**	0.687
Molise	0.6307**	0.357	0.7586	0.462	0.1383	0.752
Campania	1.2950*	0.279	1.5399*	0.373	0.5298	0.660
Puglia	0.6823**	0.352	0.5787	0.450	-0.4082	0.745
Basilicata	1.0220*	0.392	0.7641	0.499	-0.0410	0.787
Calabria	0.8270*	0.292	0.5304	0.384	-0.7121	0.678
Sicilia	1.1570*	0.286	0.8616*	0.373	0.0225	0.661
Sardegna	1.3647*	0.329	1.1877*	0.439	0.9015	0.725

* Significativo al 95 per cento. **Significativo al 90 per cento.

⁵ Anche la significatività statistica degli effetti casuali di secondo livello può essere valutata tramite il test t di Wald (Chiandotto e Giusti, 2005)

In generale, nel modello nullo, non controllando quindi per alcuna covariata, sia di primo che di secondo livello, tutte le regioni ad eccezione di cinque (Val d'Aosta, Trentino Alto Adige, Liguria, Umbria e Marche) risultano avere un effetto significativamente differente da 0; le regioni che esercitano gli effetti più positivi (valori più elevati e negativi degli effetti casuali di secondo livello) sulla probabilità delle 25-45enni coniugate di essere attive sul mercato del lavoro sono, nell'ordine, Emilia Romagna, Piemonte, Veneto e Lombardia; viceversa, quelle che influenzano maggiormente l'inattività sono la Sardegna, l'Abruzzo, la Campania, la Sicilia e la Basilicata. Una volta che si controlla per le covariate individuali, di primo livello, diminuisce il numero di regioni che risultano avere un effetto statisticamente significativo (almeno al 90 per cento), che diventano dieci: in particolare, risultano avere un effetto negativo sull'inattività il Piemonte, il Veneto e l'Emilia Romagna, mentre hanno un impatto positivo la Campania, l'Abruzzo, il Lazio e la Sardegna. Infine, nel modello finale, che controlla per le covariate sia di primo che di secondo livello, hanno un impatto negativo sulla probabilità di essere inattive le sole Piemonte e Veneto (entrambe con un livello di significatività statistica del 90 per cento) e presentano invece un impatto positivo il Lazio (al 95 per cento) e l'Abruzzo (al 90 per cento). Se si confrontano le stime ottenute con il modello nullo con quelle ottenute dal modello finale, si osserva in generale come le regioni settentrionali godano nel primo caso del fatto che il grado di modernizzazione socio-economica contribuisce ad aumentare la probabilità di attività femminile sul mercato del lavoro; controllando invece l'effetto dell'eterogeneità regionale attraverso l'introduzione delle due variabili di contesto (quella di sintesi e quella relativa al gap di genere sui redditi da lavoro), gli effetti casuali di secondo livello ne risultano "penalizzate" ed il loro valore diminuisce, fino a non essere generalmente statisticamente differenti da 0. Lo stesso accade, partendo però da un'influenza negativa sull'attività femminile, per le regioni meridionali. Si è fatto cenno al fatto che, una volta controllate le covariate individuali e quelle di contesto, i valori degli effetti casuali di secondo livello potrebbero rappresentare una sorta di "atteggiamento culturale" della regione nei confronti dell'inattività femminile. In questo caso, come detto, risulterebbero avere un effetto positivo sull'inattività femminile solamente il Lazio e l'Abruzzo ed un effetto negativo le sole Piemonte e Veneto. E' però probabile che la variabilità residua dovuta alle regioni è imputabile non tanto ad atteggiamenti culturali, quanto piuttosto alla presenza di caratteristiche omesse che possano spiegare il diverso posizionamento di queste quattro regioni rispetto alle altre.

Conclusioni

In questo lavoro si sono analizzate le cause dell'inattività femminile in Italia nelle donne 25-45enni coniugate o conviventi. L'analisi ha considerato sia caratteristiche individuali proprie della donna che caratteristiche della sua famiglia attuale e della famiglia di origine. Si è inoltre tenuto in considerazione l'effetto che la regione di residenza, con proprie caratteristiche in termini di mercato del lavoro e disponibilità di servizi di aiuto alle famiglie, può avere sulle scelte femminili. I modelli utilizzati sono modelli di regressione logistica multilivello a intercetta casuale. Questi provvedono all'integrazione tra le dimensioni micro e macro, evidenziando le relazioni esistenti tra l'individuo e il contesto territoriale in cui vive: il processo in opera è quindi funzione sia di caratteristiche individuali che del gruppo cui l'individuo appartiene.

Tra i risultati ottenuti, la prima cosa da rilevare è che il 7,6 per cento della variabilità relativa al fenomeno dell'attività/inattività è dovuta al differente contesto territoriale. Inoltre, l'indicatore utilizzato come "bontà" del territorio nei confronti dell'attività femminile, risultato di una sintesi fattoriale tra indicatori del mercato del lavoro e di offerta di servizi assistenziali, altamente correlati tra loro, ha un forte impatto sulle decisioni della donna.

Per quanto riguarda l'effetto delle caratteristiche individuali, il titolo di studio ha, come atteso, una forte influenza sulla probabilità di partecipare attivamente al mercato del lavoro, che cresce in maniera significativa al crescere del titolo di studio, mentre l'aver figli ha generalmente un effetto positivo sull'inattività femminile, particolarmente rilevante quando i figli sono molto piccoli (0-5 anni), sia che questi frequentino sia che non frequentino un asilo nido o una scuola dell'infanzia. Se l'intervistata dichiara di essere aiutata nei lavori domestici e di cura dei figli dal coniuge o convivente la donna, a parità di tutte le altre condizioni, ha meno probabilità di essere inattiva. Lo stesso discorso vale quando l'aiuto nella gestione della casa o nella cura dei figli viene da parenti, colf, badanti, baby sitter: anche in questo caso la probabilità di essere inattive diminuisce considerevolmente. Infine, relativamente alle famiglie di origine, il fatto che le donne che frequentavano la famiglia della donna quando questa era bambina fossero prevalentemente lavoratrici risulta essere un esempio per le scelte della donna in età adulta, avendo un forte impatto negativo sulla probabilità di essere inattiva.

Tramite lo studio degli effetti casuali di secondo livello si è potuto analizzare l'effetto esercitato sulle probabilità individuali da una generica regione. Nel modello nullo, non controllando quindi per alcuna covariata, tutte le regioni ad eccezione di cinque (Val d'Aosta, Trentino Alto Adige, Liguria, Umbria e Marche) risultano avere un effetto statisticamente rilevante; le regioni che esercitano gli effetti più positivi sulla probabilità delle 25-45enni coniugate di essere attive sul mercato del lavoro sono, nell'ordine, Emilia Romagna, Piemonte, Veneto e Lombardia; viceversa, quelle che influenzano maggiormente l'inattività sono la Sardegna, l'Abruzzo, la Campania, la Sicilia e la Basilicata. Quando invece si controlla sia per l'eterogeneità individuale che per le caratteristiche regionali, diminuisce in maniera significativa l'effetto residuo esercitato dal territorio, e hanno un impatto positivo residuo sulla probabilità di essere attive le sole Piemonte e Veneto e presentano invece un impatto negativo il Lazio e l'Abruzzo.

I risultati presentati in questo contributo sono comunque da considerarsi un primo step di una ricerca più estesa, in quanto si vuole verificare l'opportunità di considerare ulteriori o differenti variabili esplicative; nello specifico, si vuole anche verificare se l'effetto residuo esercitato da alcune regioni nel modello completo sia dovuto ad una componente culturale regionale o piuttosto all'esistenza di variabili omesse. Inoltre, si vuole stimare un modello che consideri effetti casuali di regione non solo sull'intercetta, ma anche sui coefficienti delle esplicative utilizzate, proprio per analizzare eventuali differenze regionali sull'intensità che queste esercitano nello spiegare l'inattività femminile.

Riferimenti bibliografici

Aitkin M., Longford N. (1986), Statistical modelling issues in school effectiveness studies, *Journal of Royal Statistics Society A*, **149**, part 1, pp.1-43.

Anderson D.A. e Aitkin M. (1985), Variance component models with binary response: interviewer variability, *Journal of Royal Statistics Society B*, **47**, pp. 203--210.

Burstein L., Linn R.L. e Capell F.J. (1978) Analyzing multilevel data in the presence of heterogeneous within-class regressions, *Journal of Educational Statistics*, **3**, pp. 347-383.

Chiandotto B., Giusti C. (2005), L'abbandono degli Studi Universitari, in *Modelli statistici per l'analisi della transizione università-lavoro* (a cura di Crocetta C.), Cleup, Padova

Goldstein H. (1995). *Multilevel Statistical Models*, Edward Arnold, London.

ISTAT (2007) I tempi della vita quotidiana, (a cura di M. C. Romano), *Argomenti* n.32, Roma.

Longford N.T. (1993) *Random Coefficient Models*, Clarendon Press, London

Rampichini C. (2002) Introduzione ai modelli multilivello, relazione presentata a "La rappresentazione dei soggetti collettivi: fra analisi scientifica e senso comune", Università di Udine 3-5 ottobre 2002

Rampichini, C. (1992), Metodi Statistici per la valutazione dell'efficienza ed efficacia dei corsi di Formazione Professionale, *Working Papers* n. 38, Dipartimento Statistico, Università' di Firenze.

Snijders T., Bosker R. (1999). *An Introduction to Basic and Advanced Multilevel Modeling*, Sage, London.