Finanziato dall'Unione europea NextGenerationEU

Agenzia Nazionale Politiche Attive del Lavoro

Garanzia Occupabilità Lavoratori

PIANO NAZIONALE DI RIPRESA E RESILIENZA – PNRR

Missione 5 Componente 1 Riforma 1.1

Programma nazionale per la

GARANZIA DI OCCUPABILITÀ DEI LAVORATORI - GOL

Allegato A

STRUMENTI PER L'ATTUAZIONE DELL'ASSESSMENT

PROFILAZIONE QUANTITATIVA

Sommario[1]

[1] Il presente lavoro è stato curato da Giovanna Linfante (Struttura 3 – ANPAL), Debora Radicchia (Struttura 3 - ANPAL) e Enrico Toti (Struttura 1- ANPAL)

1. Modello quantitativo per l'Assessment degli utenti dei CPI

1.1 Dati amministrativi

L'introduzione della cd. DID on-line a partire dal 4 dicembre 2017 ha consentito un miglioramento qualitativo e quantitativo del patrimonio informativo dei Centri per l'impiego e più in generale del mercato del lavoro. In particolare il miglioramento degli archivi amministrativi copre tre aspetti:

- completezza e tempestività delle informazioni sulla disoccupazione amministrativa (sistema DID);
- integrazione tra banche dati amministrative, in particolare i sistemi informativi ANPAL e MLPS-Comunicazioni Obbligatorie (d'ora in avanti MLPS-CO);
- allargamento del patrimonio informativo del disoccupato con le informazioni provenienti dalla profilazione quantitativa così come contemplata dal D.lg. 150.

Questi tre aspetti rappresentano il filo che tiene insieme la presente proposta di revisione dell'attuale modello di profiling quantitativo.

Insieme di riferimento – persone che hanno sottoscritto una DID negli anni 2018 e 2019 (circa 3,4 milioni)

Variabile dipendente/outcome – Per ciascun individuo dell'insieme di riferimento è possibile ricostruire l'insieme dei rapporti di lavoro intervenuti in un determinato periodo di tempo successivamente alla data di rilascio della DID. Compito della variabile dipendente è quello di dare una *misura* oggettiva del grado di difficoltà di (re)inserimento lavorativo del disoccupato, nell'arco di tempo scelto (365 giorni). Rispetto a questo intervallo di tempo sono state costruite due variabili dipendenti.

A. **y0**. La variabile dicotomica assume valore "1" se l'individuo non ha rapporti di lavoro intervenuti nei 365 giorni dopo il rilascio della DID, e valore "0" altrimenti;

B. **y1**. La variabile dicotomica assume valore "1" se l'individuo ha avuto al più 90 giorni complessivi di lavoro contrattualizzato nei 365 giorni dopo il rilascio della DID, estendendo l'orizzonte temporale fino al 454*esimo* giorno successivo se l'individuo era occupato al 365-esimo giorno per meno di 90 giorni, e valore "0" altrimenti;

Variabili esplicative (X) – le variabili esplicative rappresentano l'insieme delle caratteristiche *possedute* dall'individuo al momento della sottoscrizione della DID. Queste informazioni coprono più sfere (anagrafica, istruzione, esperienza lavorativa, contesto familiare ecc.), e la loro scelta necessariamente *influisce* sulla bontà del modello utilizzato nel *predire* la più o meno accentuata difficoltà dell'individuo nel (re)inserimento lavorativo. La capacità esplicativa delle singole variabili nel modello segue un criterio più generale di *parsimonia* e di sfruttamento integrale delle informazioni già censite negli archivi Anpal (DID e SAP) e MLPS-CO.

Alcune variabili, legate per lo più all'esperienza lavorativa ricostruita dalle Comunicazioni Obbligatorie, necessitano in ogni caso di un "calcolo" da sistema.

1.2 Principali differenze rispetto all'attuale modello di profiling

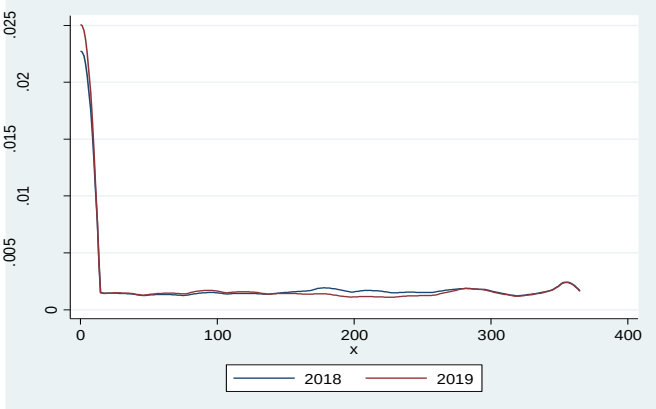
- dati amministrativi Vs. Survey statistiche
- ampliamento dell'insieme delle variabili dipendenti che possono essere scelte per adattarsi a diverse definizioni di minore/maggiore occupabilità
- ampliamento del set di variabili esplicative **X**

2. Descrittive per y e X, Eterogeneità regionali

L'universo di riferimento è rappresentato dall'insieme degli individui con una DID sottoscritta nel biennio 2018-2019, per un totale di circa 3,4 milioni di unità.

NOTA: la scelta di utilizzare i due anni 2018-2019 è maturata durante i lavori di stima dei modelli, poiché essa si è mostrata più efficace nel produrre stime con migliori capacità predittive rispetto ai singoli contesti regionali. La principale obiezione è rappresentata dal fatto che si viene in tal modo a coinvolgere l'anno 2020, con tutta la sua eccezionalità dovuta alla crisi pandemica, per quel che attiene alla misurazione degli outcome per i disoccupati del 2019.

Figura 2.1- Giorni lavorati nei 365 giorni successivi alla DID



Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

In termini relativi, il confronto tra il 2018 e il 2019, comporta in quest'ultimo caso un peggioramento compreso tra l'8,6% (per la variabile y1) e il 9,1% (per la variabile y0). A livello regionale, le variazioni in termini relativi più marcate si hanno nelle regioni del Nord ed in particolare nelle regioni più piccole: Valle d'Aosta, Liguria e PA di Bolzano.

Tavola 2.1: Quota LTU nei 365 giorni successivi alla DID. Confronto 2018-2019 per Regione

	Totale	2018	2019	Delta	Delta%
Piemonte	0,46	0,43	0,48	0,05	11,17
Valle d'Aosta	0,38	0,35	0,41	0,06	16,04
Lombardia	0,45	0,41	0,47	0,05	13,04
PA Bolzano	0,20	0,18	0,22	0,04	21,53
Pa Trento	0,24	0,23	0,25	0,03	12,05
Veneto	0,37	0,34	0,39	0,05	15,47
Friuli Venezia Giulia	0,38	0,36	0,40	0,03	9,51
Liguria	0,39	0,34	0,43	0,09	26,73
Emilia Romagna	0,40	0,39	0,41	0,02	4,47
Toscana	0,39	0,37	0,40	0,04	10,78
Umbria	0,47	0,46	0,48	0,02	4,08
Marche	0,43	0,42	0,45	0,03	7,68
Lazio	0,47	0,44	0,49	0,04	10,06
Abruzzo	0,40	0,39	0,41	0,02	5,71
Molise	0,45	0,44	0,46	0,02	5,29
Campania	0,46	0,44	0,48	0,04	8,18
Puglia	0,45	0,43	0,48	0,05	11,47
Basilicata	0,36	0,36	0,36	0,00	1,16
Calabria	0,49	0,47	0,51	0,04	8,93
Sicilia	0,52	0,51	0,54	0,03	5,87
Sardegna	0,36	0,34	0,38	0,04	10,29
Totale	**0,44**	**0,42**	**0,46**	**0,04**	**9,18**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 2.2: Quota di persone occupate per al più 90 giorni nei 365 giorni successivi alla DID (y1). Confronto 2018-2019 per Regione

	Totale	2018	2019	Delta	Delta%
Piemonte	0,55	0,52	0,57	0,05	9,95
Valle d'Aosta	0,48	0,46	0,52	0,06	12,71
Lombardia	0,53	0,50	0,55	0,06	11,76
PA Bolzano	0,27	0,24	0,31	0,07	28,67
Pa Trento	0,34	0,31	0,36	0,05	17,07
Veneto	0,45	0,42	0,49	0,07	17,09
Friuli Venezia Giulia	0,47	0,45	0,49	0,04	9,59
Liguria	0,48	0,42	0,52	0,11	25,96
Emilia Romagna	0,50	0,49	0,51	0,02	5,13
Toscana	0,47	0,45	0,50	0,05	11,89
Umbria	0,55	0,54	0,56	0,02	4,48
Marche	0,53	0,52	0,55	0,03	6,49
Lazio	0,56	0,54	0,58	0,04	7,91
Abruzzo	0,51	0,50	0,53	0,03	5,72
Molise	0,55	0,54	0,57	0,03	5,04
Campania	0,55	0,53	0,57	0,04	7,65
Puglia	0,57	0,54	0,59	0,05	8,68
Basilicata	0,47	0,47	0,47	0,00	0,04
Calabria	0,58	0,56	0,60	0,04	6,62
Sicilia	0,61	0,60	0,63	0,03	5,50
Sardegna	0,47	0,44	0,50	0,06	12,74
Totale	**0,53**	**0,51**	**0,55**	**0,04**	**8,62**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Tavola 2.3: Giorni medi lavorati nei 365 giorni successivi alla DID (GG_LAV). Confronto 2018-2019 per Regione

	Totale	2018	2019	Delta	Delta%
Piemonte	96,60	103,22	90,20	-13,02	-12,61
Valle d'Aosta	111,16	116,72	103,47	-13,25	-11,35
Lombardia	104,48	112,47	99,12	-13,36	-11,87
PA Bolzano	163,85	180,06	145,90	-34,16	-18,97
Pa Trento	158,01	166,10	150,15	-15,94	-9,60
Veneto	124,00	133,30	115,08	-18,22	-13,67
Friuli Venezia Giulia	121,77	127,64	116,05	-11,59	-9,08
Liguria	110,32	122,89	99,66	-23,23	-18,90
Emilia Romagna	110,06	112,53	107,88	-4,65	-4,13
Toscana	118,12	123,87	112,47	-11,40	-9,20
Umbria	100,34	103,99	96,95	-7,03	-6,76
Marche	102,33	105,25	99,47	-5,77	-5,49
Lazio	97,76	102,94	93,00	-9,94	-9,66
Abruzzo	115,42	117,82	113,19	-4,63	-3,93
Molise	105,94	107,68	104,06	-3,62	-3,37
Campania	109,25	113,45	105,21	-8,24	-7,26
Puglia	99,26	103,46	94,35	-9,11	-8,80
Basilicata	123,90	123,09	124,67	1,59	1,29
Calabria	96,45	99,00	93,82	-5,18	-5,24
Sicilia	89,55	92,64	86,08	-6,56	-7,08
Sardegna	118,70	124,55	112,87	-11,68	-9,38
Totale	**106,88**	**111,86**	**102,06**	**-9,80**	**-8,76**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Affinché si possa ammettere che la differenza nei livelli di outcome osservati tra le coorti 2018 e 2019 sia interamente attribuibile al mutato contesto di riferimento (leggi pandemia 2020), occorre che le stesse due coorti siano *statisticamente* simili.

Ad eccezione di alcune tra le Regioni più piccole (Basilicata, Valle d'Aosta), le differenze tra le due annualità più importanti riguardo alla composizione per **genere** della platea dei disoccupati (cfr. appendice tavola 2.4), si osservano per la Campania (donne +2,6 punti percentuali nel 2019) e la Lombardia (+1,5 p.p. nel 2019).

Anche l'**età** media per il 2019 (34,4 anni) si mostra significativamente diversa rispetto al 2018 (32,1 anni), con un incremento di +2,3 anni (cfr. appendice tavv. 2.5, 2.6 e 2.7).

La dinamica riguarda sia le donne (+2,1 anni) che gli uomini (+2,4 anni).

Per quanto riguarda il **titolo di studio**, il confronto tra il 2018 e il 2019, anche se con variazioni contenute in termini di punti percentuali, vede un generale aumento per il 2019 di disoccupati con titolo di studio basso e una contestuale riduzione dei disoccupati con titolo di studio medio. Questa dinamica è particolarmente evidente nelle regioni del Mezzogiorno (cfr. appendice tavv. 2.8, 2.9 e 2.10).

Rispetto alla presenza di disoccupati con una o più **esperienze lavorative** nei 12 mesi precedenti la DID, il confronto tra il 2019 e il 2018 comporta un generale aumento nelle regioni del Nord, in particolare la Lombardia (+16,5 p.p.) e un generale decremento nelle regioni del Mezzogiorno in particolare in Campania e in Calabria (cfr. appendice tavola 2.11). Dal 2018 al 2019 aumenta di 0,8 punti percentuali la quota di chi aveva un rapporto di lavoro attivo al momento del rilascio della DID: dal 5,7% al 6,5%.

Come ultimo confronto si prende in esame la variabile relativa alla **presenza di figli** all'interno del nucleo familiare del disoccupato. Il 2019 presenta una percentuale sensibilmente più elevata di nuclei con figli con particolare riferimento alle Regioni Lombardia, Emilia Romagna, Lazio, Calabria e Sicilia (cfr. appendice tavv. 2.12, 2.13 e 2.14).

In conclusione, la struttura delle due coorti di disoccupati 2018 e 2019 presenta delle differenze che diventano più o meno significative se calate nel contesto regionale. La Lombardia è la Regione che presenta le

differenze più marcate rispetto alle caratteristiche dei soggetti analizzate.

Tenere insieme le due annualità nella stima del modello consente quindi di mediare le differenze con la conseguenza di ottenere stime più robuste e meno sensibili a variazioni di segno amministrativo che possono avere un peso non trascurabile in alcuni contesti territoriali soprattutto per il 2018.

3. Modelli variabile dipendente dicotomica

Nell'applicazione classica la variabile dipendente, oggetto di stima, è rappresentata da una condizione di presenza (=1) e di assenza (=0) di una determinata condizione rilevata su ciascuna unità in un determinato contesto spaziale e temporale. La condizione rilevata (1, 0) è direttamente associata alla difficoltà di (re)inserimento lavorativo.

L'utilizzo di una variabile dipendente dicotomica viene associata a modelli di stima di tipo Logit o Probit:

$p(y) = \vartheta(Y = 1|X)$

Tra i punti di forza di una tale modellistica vi è il fatto che il valore del profiling è sempre contenuto nell'intervallo [0, 1], ed inoltre il valore atteso del profiling stimato *è pari al valore atteso di y* :

$E[\vartheta(Y = 1|X)] = E(y)$

Tra i punti di debolezza del modello vi è il forte legame con la **scelta** della variabile dipendente.

3.1 La scelta della variabile dipendente/outcome

Punto di partenza è la definizione dell'outcome: qual è la variabile indicatrice con cui si vuole misurare la *difficoltà* di (re)inserimento lavorativo di un soggetto che si reca presso un centro per l'impiego?

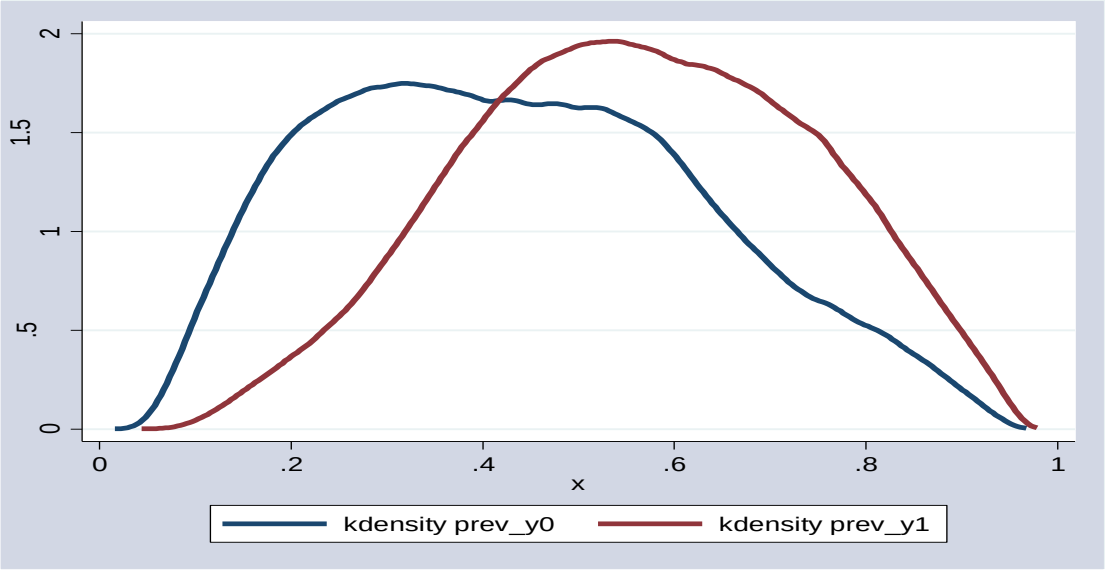
Un modello più restrittivo (modello LTU) considera come outcome la variabile y_0 (cfr. 1.1) che rappresenta una condizione di "estrema" debolezza/difficoltà di (re)inserimento lavorativo. Il modello probabilistico stima, infatti, la probabilità che l'i-esimo individuo, avente determinate caratteristiche descritte dall'insieme di variabili esplicative utilizzate nel modello **X**, non avrà giorni di lavoro

contrattualizzati nell'anno successivo alla data di ingresso nella disoccupazione.

Un secondo outcome riguarda la variabile y_1 (cfr. 1.1). Questa variabile aderisce meglio all'idea di debole o forte occupabilità. La stima $\Pr(y_1 = 1 \mid X)$ avviene con lo **stesso modello utilizzato** per l'outcome precedente. Quindi, in particolare, questo si traduce nel fatto che si utilizza lo stesso insieme di covariate X.

Consideriamo le due distribuzioni $\vartheta(y_0 = 1 \mid X)$, $\vartheta(y_1 = 1 \mid X)$. Si vede chiaramente come il modello y_0 sia più "panciuto" nella coda sinistra (bassi valori dell'indice, ossia migliori livelli di occupabilità), rispetto al modello y_1 che invece presenta una distribuzione più spostata verso la coda destra.

Figura 3.1.1 - Distribuzione dei valori P(y) per y=y_0 e y=y_1



	media	dev. St.	mediana	Q1	Q3
y0	0,436	0,197	0,424	0,278	0,578
y1	0,528	0,187	0,525	0,385	0,672

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

L'obiettivo che ci poniamo è quello di utilizzare entrambe le distribuzioni per *meglio* selezionare l'insieme degli individui *"forti"* (veri negativi). Si affronterà questo aspetto nel successivo capitolo 4. Il presente capitolo si chiude invece con un'analisi dei modelli stimati.

3.2 Stime del modello

Il modello logistico stima, su circa 3,4 milioni di individui che hanno rilasciato una DID nel 2018 e nel 2019, la probabilità di diventare disoccupato di lunga durata (y_0) ovvero di lavorare per al più 90 giorni (y_1), utilizzando 15 variabili (X) e la loro interazione (cfr. appendice tav. 3.2.1). Il processo restituisce per ciascuna caratteristica individuale (genere, età, titolo di studio, esperienza lavorativa pregressa, ecc.) un coefficiente che indica il peso e il segno che una specifica caratteristica ha nel determinare la probabilità di diventare disoccupato di lunga durata (cfr. appendice figura 3.2.1).

Come già specificato, la scelta delle caratteristiche osservabili X da utilizzare nel modello si è basata sulla decisione di integrare e valorizzare molte delle informazioni già disponibili dagli archivi amministrativi delle DID, delle MLPS-CO e del patrimonio informativo proveniente dalla profilazione quantitativa prevista dal D.lg. 150.

L'utilizzo di archivi di dati amministrativi ha richiesto una fase di preprocessing molto importante, sia per quanto riguarda la presenza di dati mancanti, sia per la valutazione delle incoerenze intrarecord che spesso interessavano anche la stessa informazione registrata in archivi diversi, ma anche nella definizione degli aggregati da inserire nel modello.

Nel dettaglio, le informazioni anagrafiche quali, genere, età, anni presenza in Italia, ma anche il titolo di studio[2], la condizione occupazionale dell'anno precedente dichiarata e le caratteristiche familiari (numero componenti, presenza di figli e di figli minori) sono quelle raccolte durante la profilazione quantitativa D.lg. 150, mentre le informazioni relative all'eventuale esperienza lavorativa pregressa, in termini anche di contratto, settore, qualifica prevalente e numero datori di lavoro cambiati, sono state ricostruite utilizzando l'archivio del MLPS CO, limitando l'analisi ai 24 mesi precedenti la sottoscrizione della DID, in modo da definire un arco temporale non troppo lontano dalla dichiarazione della disponibilità. Nel caso di nessun rapporto di lavoro

[2] Per il titolo di studio è stata applicata la stessa ricodifica prevista nell'attuale profiling D.lg. 150, ovvero tenendo distinto il tipo di scuola (qualifica professione, diploma istituto professionale, istituto tecnico, liceo) e il corso di laurea (triennale, magistrale, specialistica, vecchio ordinamento) per tipo facoltà.

alle dipendenze attivato nei due anni precedenti la sottoscrizione della DID, tali variabili sono valorizzate con la categoria "Mai lavorato". L'attività prevalente è determinata in termini di maggiore durata in giorni del rapporto di lavoro.

Dall'archivio delle SAP sono state definite alcune variabili relative alle competenze informatiche e linguistiche, e al possesso di una patente. Solo quest'ultima è risultata però significativa per il modello di stima.

Per quanto riguarda, invece, la ricostruzione dell'ambito territoriale, nel modello entra come variabile di controllo la provincia del CPI competente preferendola al domicilio/residenza dell'individuo poiché più aderente al contesto territoriale del mercato del lavoro di riferimento dell'individuo.

La fase di pulizia del dataset di analisi ha, inoltre, previsto l'esclusione di individui: con età inferiore ai 16 anni e superiore ai 64 anni alla data di rilascio della DID; appartenenti alle liste speciali (L. 68); con valori mancanti in alcune variabili, tra cui il titolo di studio, gli anni di presenza in Italia per gli stranieri, e la qualifica prevalente.

Il modello restituisce la quasi totalità dei coefficienti statisticamente significativi, e anche il loro segno conferma quanto osservato a livello descrittivo: il rischio di diventare un disoccupato di lunga durata (y_0=1) è maggiore per le donne rispetto gli uomini, si riduce all'aumentare del livello di istruzione e in presenza di esperienze lavorative pregresse, mentre rispetto all'età il rischio di diventare disoccupati di lunga durata è dapprima decrescente per poi aumentare nelle fasce di età più adulte. Per quanto riguarda la dimensione territoriale, i disoccupati che risiedono nel Mezzogiorno hanno maggiore probabilità di diventare disoccupati di lunga durata, benché il dato sia sensibile alle diverse caratteristiche degli utenti dei CPI nelle singole Regioni. Il carico familiare, soprattutto la presenza di figli, aumenta molto il rischio di restare disoccupato di lunga durata per le donne.

Con leggere differenze, le stesse conclusioni si estendono al caso della variabile y_1, ovvero al rischio di lavorare al più 90 giorni nell'anno successivo alla DID.

Si rinvia all'appendice statistica per i dettagli tecnici dei modelli Logit stimati.

4. Utilizzo combinato modelli LTU e Y1: identificazione delle platee "work-ready" e "weaker"

I valori stimati per i due modelli (y_0, y_1) rappresentano per ciascun individuo la probabilità di trovarsi, dopo un anno dall'ingresso nello stato di disoccupazione (DID), rispettivamente senza aver mai lavorato neppure un giorno (y_0) o al più aver lavorato 90 giorni (y_1). Ogni considerazione/conclusione tratta dal modello vale in "media", per cui ci aspetta che individui che hanno un valore più basso dell'indice, per i due modelli, siano potenzialmente più occupabili, abbiano cioè caratteristiche di occupabilità migliori, rispetto ad individui dotati di livelli di probabilità più elevati. Tuttavia, come detto, questo aspetto vale in media mentre la reale "situazione" occupazionale dell'individuo può risentire di situazioni e/o caratteristiche non osservabili o misurabili dal modello.

I modelli di stima, lineari o non lineari, sono dunque soggetti ad errori nella propria capacità predittiva, e ciò che vale in media può invece comportare situazioni critiche laddove si procedesse all'applicazione di regole automatiche di classificazioni basate sull'indice attribuito al singolo individuo.

I metodi quantitativi qui analizzati devono necessariamente essere utilizzati a **supporto** del processo complessivo di *assessment* in cui è rilevante l'integrazione con metodi qualitativi.

Muovendoci all'interno di questo paradigma, qui si propone un utilizzo del profiling quantitativo atto a selezionare la parte di platea di disoccupati che si possono ritenere, con un margine di errore atteso ritenuto accettabile, particolarmente forti in termini di occupabilità (**work-ready**), e la platea di disoccupati per i quali invece si può ritenere che siano particolarmente fragili/difficili in termini di occupabilità. Queste due coorti si posizionano nelle due code distributive delle funzioni di probabilità stimate per y_0 e y_1, con una limitata percentuale attesa di popolazione coinvolta.

4.1 Falsi Negativi e Falsi Positivi

Un criterio per valutare la capacità predittiva di un modello probabilistico si basa sull'incidenza dei falsi negativi e dei falsi positivi per un dato livello (soglia) della probabilità stimata.

In generale i software per la stima dei modelli Logit/Probit forniscono alcune statistiche di sintesi sulla bontà predittiva del modello. Una di esse si basa sulla quota di individui correttamente classificati rispetto ad un valore soglia di p(y) pari a 0,5: sono considerati correttamente classificati gli individui per i quali il valore stimato di p(y)>=0,5 e l'outome osservato è pari a 1, e gli individui per i quali il valore stimato di p(y)<0,5 e l'outcome osservato è pari a 0. Sulla base di questa statistica, complessivamente il modello y_0 classifica correttamente il 67,6% dei casi, mentre la capacità complessiva di predizione del modello y_1 scende al 65,7% (tavola 4.1.1).

Tavola 4.1.1 – Statistiche sull'accuratezza dei modelli Logit per y_0 e y_1

Classified + if predicted Pr(D) >=0,5
True D defined as y != 0

		y0	y1
Sensitivity	Pr(+ D)	55,83%	69,10%
Specificity	Pr(-~D)	76,76%	61,90%
Positive predictive value	Pr(D +)	65,01%	67,03%
Negative predictive value	Pr(~D -)	69,19%	64,12%
False + rate for true ~D	Pr(+~D)	23,24%	38,10%
False - rate for true D	Pr(- D)	44,17%	30,90%
False + rate for classified +	Pr(~D +)	34,99%	32,97%
False - rate for classified -	Pr(D -)	30,81%	35,88%
Correctly classified		67,63%	65,70%
AUC (ROC curve)		0,729	0,715

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Un ulteriore criterio per valutare la capacità predittiva del modello si basa sull'incidenza dei cosiddetti falsi negativi e falsi positivi (vedi oltre) al variare dei valori soglia di p(y) da 0 a 1.

Definiamo con

$$\tau_i = \frac{i}{100} \ (per\ i = 1,..,100) \qquad [1]$$

i valori soglia mobili della probabilità stimata (livello di profiling), valori che vanno da 0,01 (per i = 1) a 1,00 (per i=100), si incrementa cioè il valore soglia di un punto percentuale al crescere di i.

Per facilitare la scrittura indichiamo con "y" il generico outcome (variabile dipendente) che assume i valori 0 (negativo) e 1 (positivo).

Nella presente Indichiamo con p(y) il valore (profiling) stimato dal modello logit, che come detto corrisponde alla probabilità di osservare l'outcome positivo di y:

p(y) = pr(y=1|X)

Falsi negativi. Per ogni unità-individuo per cui vale:

p(Y)<= τ_i,

viene fatta l'assunzione di ritenere l'individuo come un potenziale negativo rispetto ad y, cioè:

Se p(y)≤ τ_i → E(y)=0 [2]

Si definisce "Falso negativo", per un dato valore soglia τ_i, l'unità-individuo per la quale invece si verifica

p(y)≤ τ_i e y=1 [3]

Chiaramente all'aumentare dei valori soglia τ_i, cioè al crescere di i, aumenta il numero dei falsi negativi. In particolare per $\tau_i = 1$, il numero dei falsi negativi corrisponde al numero complessivo dei positivi (y=1). Definiamo come indice di tolleranza dei falsi negativi la quantità:

$$\alpha_i = \frac{\sum_{j\epsilon D_i} I(p(y)\leq \tau_i; y=1)}{\sum_j I(D_i)} \qquad [4]$$

Dove D_i rappresenta l'insieme caratterizzato da $p(y) \leq \tau_i$ e l'indice j identifica l'unità-individuo della popolazione. La [4] rappresenta l'incidenza dei falsi negativi (numeratore del rapporto) rispetto alla popolazione complessiva che soddisfa la condizione p(y)<= τ_i. Anche il valore α_i, presenta in generale un andamento crescente al crescere del valore soglia τ_i, fino ad arrivare al valore E(y) per i=1.

Falsi positivi: Per ogni unità-individuo per cui vale p(y)> τ_i, viene fatta l'assunzione di un outcome (Y) positivo cioè:

Se p(y)> τ_i → E(y)=1 [5]

Si definisce "Falso positivo" per un dato valore soglia τ_i, l'unità-individuo per la quale si verifica

p(y)> τ_i e y=0 [6]

Chiaramente all'aumentare dei valori soglia τ_i, cioè al crescere di i, si riduce il numero dei falsi positivi. Per $\tau_i \sim 1$, il numero dei falsi positivi è pari a zero. Definiamo come indice di tolleranza dei falsi positivi la quantità:

$$\gamma_i = \frac{\sum_{j \epsilon G_i} I(p(y) > \tau_i; y=1)}{\sum_j I(G_i)} \qquad [7]$$

Dove G_i rappresenta l'insieme caratterizzato da $p(y) > \tau_i$ e l'indice j identifica l'unità-individuo della popolazione. La [7] rappresenta l'incidenza dei falsi positivi (numeratore del rapporto) rispetto alla popolazione complessiva che soddisfa la condizione $p(y) \geq \tau_i$. Anche il valore γ_i, presenta in generale un andamento decrescente rispetto a τ_i, fino ad arrivare al valore 0 per i=1.

La figura 4.1 riproduce i valori α_i e γ_i, in base 100 (asse sx) e i valori della quota di popolazione interessata, sempre in base 100 (asse dx), al variare della soglia τ_i (asse delle ascisse). La quota di popolazione interessata è pari al numero di unità in D_i (vale la condizione $p(y) \leq \tau_i$) sul totale della popolazione (istogramma grigio) e al numero di unità in G_i (vale la condizione $p(y) > \tau_i$) sul totale della popolazione (istogramma giallo). L'interesse per valori bassi di τ_i è rivolto ai falsi negativi (α_i e D_i/T), mentre per valori alti di τ_i l'interesse si sposta sui falsi positivi (γ_i e G_i/T). Che cosa osserviamo se fissiamo un valore comune per gli indici di tolleranza α_i e γ_i? Supponiamo di fissare i margini di tolleranza ad un valore massimo di 0,2 (20%). Per quali valori di τ_i possiamo attenderci di osservare un'incidenza di falsi negativi e di falsi positivi, rispettivamente, non superiore al 20%.

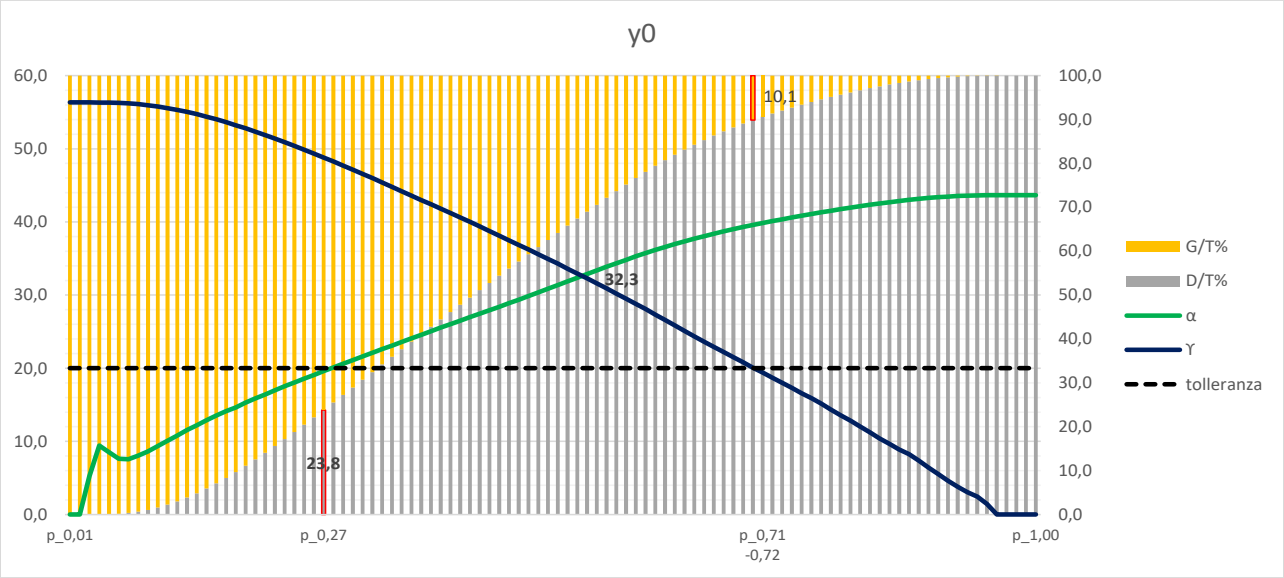
Tracciando una retta parallela all'asse delle ascisse di ordinata (asse sinistra) pari a 0,2, si ottengono i valori cercati τ_i, nel punto di intersezione di questa retta con la retta α_i e γ_i rispettivamente.

La figura 4.1.1, in corrispondenza del margine di errore fissato a 0,2, mostra come per $y=y_0$ (LTU) i valori τ_i corrispondono rispettivamente a 0,27 e 0,71. Detto in altri termini, se volessimo considerare come "negativi" ($y_0=0$)[3] gli individui che presentano un indice di profiling $pr(y_0) \leq 0,27$,

[3] L'accezione "negativo" attiene all'outcome utilizzato e dunque in questo caso i "negativi" sono gli individui più forti in termini di occupabilità, ovvero nella fattispecie individui con una bassa probabilità di essere LTU.

allora ci si deve attendere che circa una persona su cinque (20%) sarebbe in realtà erroneamente classificata, potendo risultare invece come caso "positivo" (y_0=1); mentre se volessimo considerare come "positivi" (y0=1) gli individui che presentano un indice di profiling pr(y_0)>0,71, allora allora ci si deve attendere che circa una persona su cinque (20%) sarebbe in realtà erroneamente classificata, potendo risultare invece come caso "negativo" (y_0=0).

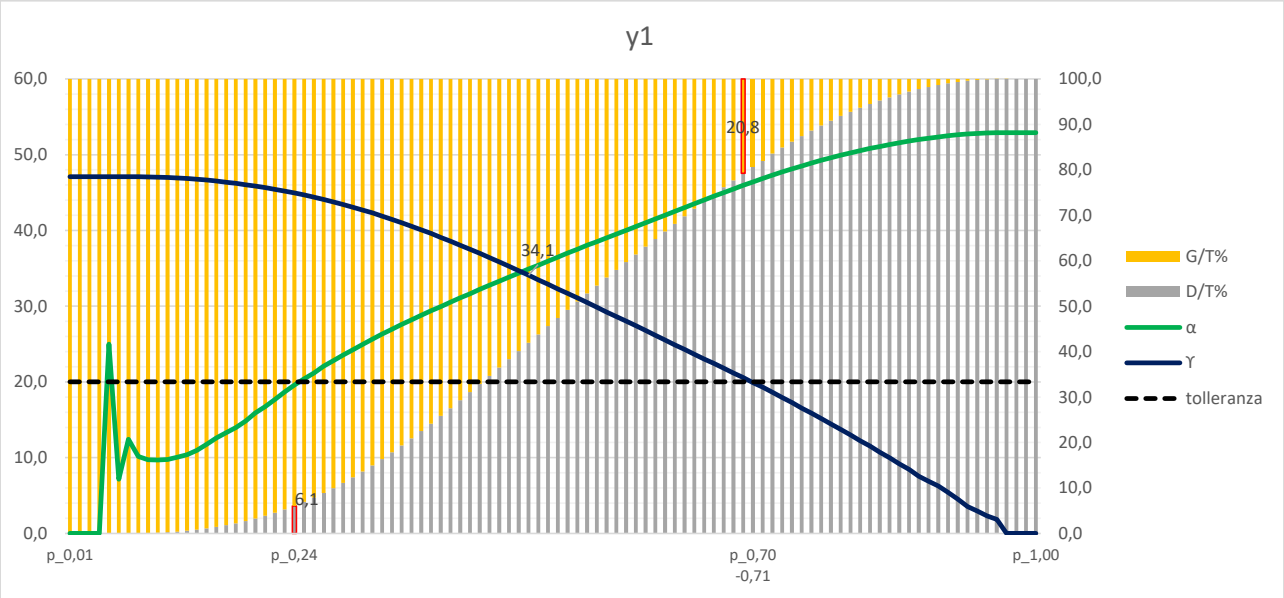
Figura 4.1.1: Falsi negativi, falsi positivi e quota popolazione interessata per y=y_0



Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Allo stesso modo la figura 4.1.2 riporta l'andamento delle grandezze α_i, D_i/T, γ_i e G_i/T per l'outcome y_1. Il confronto tra i due outcome mostra come, a parità di margine di tolleranza, per y_1 si ha un valore inferiore della popolazione più forte (D/T), pari al 6,1% contro il 23,8% visto per y_0, e un valore superiore della quota di popolazione debole (G/T) pari al 20,8% contro il 10,1% osservato per y_0. I valori soglia per y_1 sono pari rispettivamente a 0,24 e 0,70.

Figura 4.1.2: Falsi negativi, falsi positivi e quota popolazione interessata per $y=y_1$



Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

4.2 Popolazioni "Work-ready" e "Weaker"

Come visto nel paragrafo precedente, il modello LTU (y_0) consente di racchiudere una percentuale più alta di popolazione (23,8%) che può essere considerata più forte in termini di occupabilità, e una percentuale più bassa di popolazione (10,1%) che può essere classificata estremamente debole (tavola 4.2.1).

Tavola 4.2.1: Stima valore p(y) per falsi negativi e falsi positivi, con tolleranza 20%. Modello y_0 e y_1.

Modello	**Falsi negativi**		**Falsi Positivi**	
	P(y)	**Pop%**	**P(y)**	**Pop%**
y0	0,27	23,8	0,71	10,1
y1	0,24	6,1	0,70	20,8

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Ribaltando le stesse percentuali sul modello y_1 si ha una situazione opposta: una percentuale maggiore della popolazione (20,8%) è classificabile come più debole, e una parte esigua della popolazione (6,1%) è classificabile come più forte. È possibile allora sfruttare l'informazione congiunta dei due modelli per arrivare a degli insiemi *più*

robusti, cercando il giusto compromesso tra il margine di errore e la percentuale di popolazione coinvolta. In particolare, volendo partire dal modello base $y=y_0$ (LTU), l'obiettivo è quello di *selezionare* dalla popolazione $D_{27}(y_0)$[4] coloro i quali, anche con riferimento alla variabile y_1, appartengono alla parte sinistra della distribuzione del profiling.

Readiness to work

Punto di partenza: $D_{27}(y_0) = \{j: pr(y_0) \leq 0,27\}$

La popolazione così individuata può mascherare tra i "veri" negativi, situazioni di "probabili" positivi rispetto alla variabile y_1, e dunque situazioni comunque deboli in termini di occupabilità. Vale a dire il fatto di non essere LTU potrebbe essere comunque associato a situazioni di *scarsa* occupazione ($y_1=1$). Possiamo allora combinare le due distribuzioni di probabilità, $pr(y_0)$ e $pr(y_1)$, per definire la categoria dei "work-ready".

La proposta è quella di selezionare la popolazione "work-ready" nella intersezione tra $D_{27}(y_0)$ e $D_{p*}(y_1)$ in cui, in accordo con le notazioni già utilizzate, $D_{p*}(y_1)$ rappresenta la popolazione con un livello di profiling stimato $pr(y1) \leq 0,36$. Questo valore è tale che $D_{36}(y_1)/T \leq D_{27}(y_0)/T = 23,8\%$.

Complessivamente la popolazione attesa work-ready è pari a circa il 19,6% della platea complessiva degli utenti CPI che rilasciano la DID in un anno.

Weakner

Il modello LTU, come detto, stima come casi positivi ($y_0=1$) situazioni di estrema difficoltà di inserimento lavorativo. Per questo motivo la coda destra della distribuzione di $pr(y_0)$ appare di per sé sufficiente per selezionare i casi più difficili. In ottica conservativa si può prendere a riferimento la popolazione $G_{70}(y_0)$ come fascia di popolazione più problematica, pari al 10,1% della popolazione complessiva. Un approccio meno conservativo potrebbe portare a considerare la popolazione $G_{71}(y_1)$ che invece raccoglie circa il 20,8% della popolazione.

Indeterminatezza

[4] In accordo con le notazioni utilizzate nel paragrafo 4.2 il termine D_{27} identifica la popolazione le cui unità-individuo soddisfano la condizione $pr(y0=1|X) \leq 0,27$, ovvero il 23,8% di individui per i quali il livello di profiling stimato è inferiore a 0,27.

Tutti i restanti casi, che possono ricomprendere una quota della popolazione tra il 70% e il 59% a seconda se per la determinazione dei weakner si adotta una definizione più o meno restrittiva. Va detto che il livello di profiling descrive anche in questa area di indeterminatezza delle zone di più o meno rischio occupazionale. Da questo punto di vista il livello di difficoltà stimato dal modello quantitativo offrirà un ulteriore elemento per l'assessment complessivo dell'individuo.

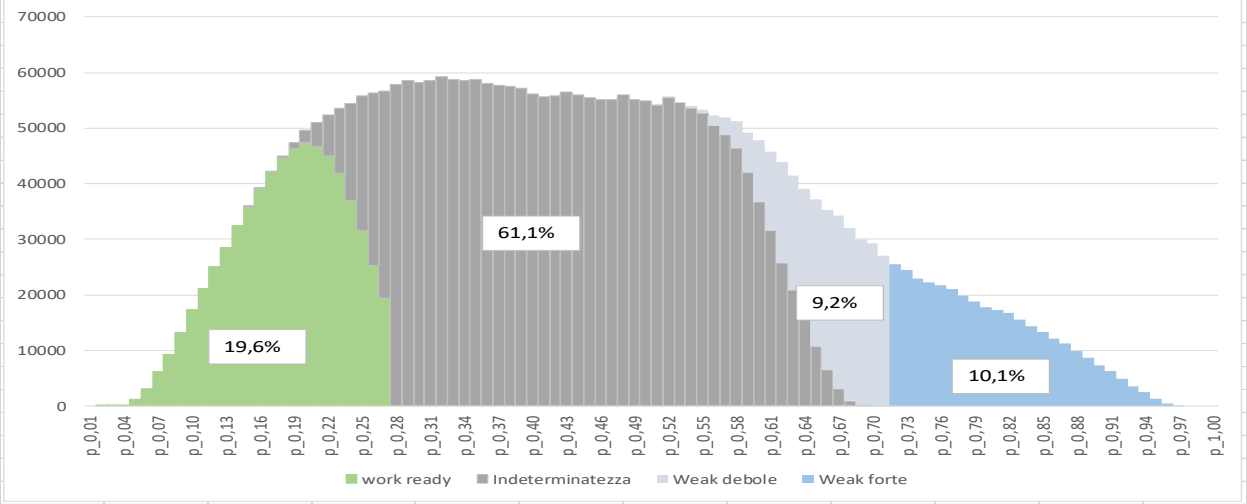
Tavola 4.2.2: Definizione Work-ready e Weaker

		% Pop	Popolazione
$P(y_0)<=0,27$	$p(y_1)<=0,36$	19,6	Work-Ready
	$p(y_1)> 0,36$	4,1	Indeterminatezza
$0,27<P(y_0)<=0,71$	$p(y_1)<= 0,70$	57,0	Indeterminatezza
$0,27<P(y_0)<=0,71$	$p(y_1)> 0,70$	9,2	Weaker debole
$p(y_0)>0,71$		10,1	Weaker forte

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Considerando un flusso annuale di (nuovi) disoccupati di circa 1,8 milioni, il sistema classifica come "work-ready" circa 350 mila individui.

Figura 4.2.1 – Definizione delle popolazioni "work ready" e "weaker".



Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

4.3 Simulazione – per alcuni strati della popolazione

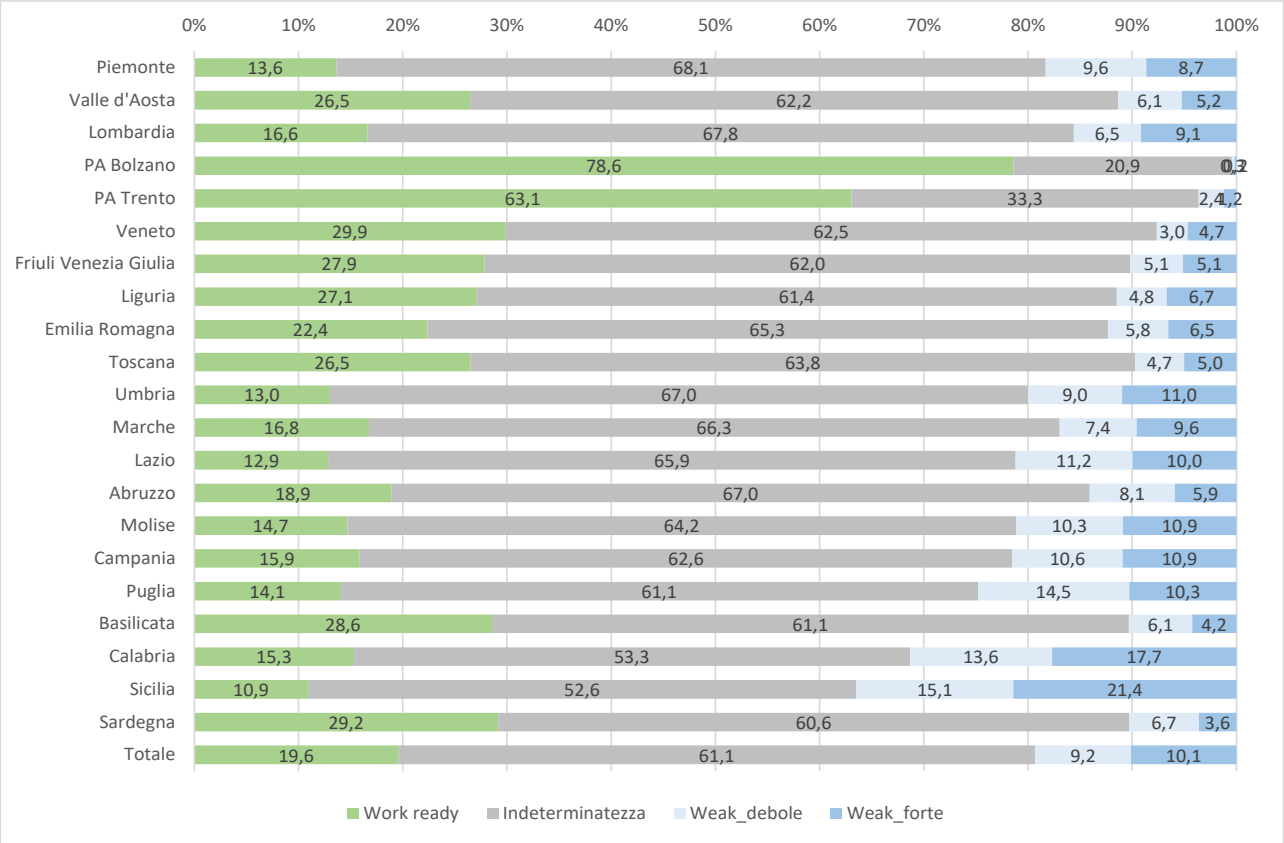
In questo paragrafo si procederà a delle simulazioni per misurare il peso che le fasce di popolazione definite al precedente assumono in determinate sub-popolazioni. In particolare, si farà riferimento sia al contesto regionale, sia a specifici target del Programma GOL: donne, giovani under 30, beneficiari di Naspi, beneficiari di RdC, over 55, disoccupati da oltre 6 mesi. Rimandiamo invece nelle singole schede regionali il dettaglio dell'incrocio Regione e Target GOL.

Regioni

La simulazione mostra una situazione molto diversificata a livello regionale.

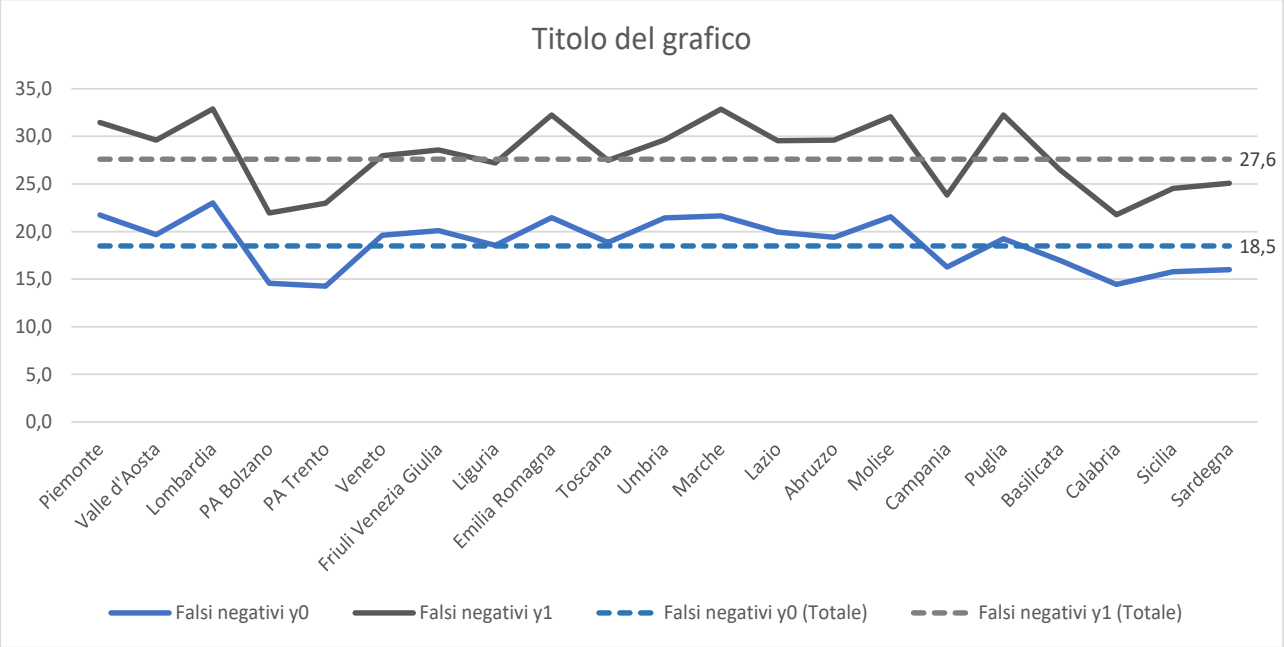
Prendendo ad esempio la popolazione "work ready" si va da un minimo del 13% (Piemonte e Umbria) ad un massimo del 78,6% per la PA di Bolzano, o comunque trascurando le regioni più piccole, si arriva a valori prossimi al 30% per la regione Veneto e per la regione Sardegna. La percentuale di falsi negativi oscilla tra il 14% per le Province Autonome e il 23,2% per la Lombardia.

Figura 4.2.1 – Distribuzione popolazione per fasce di rischio e regione.



Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Figura 4.2.2 – Work Ready: Falsi negativi (%) per y_0 e y_1 per regione.



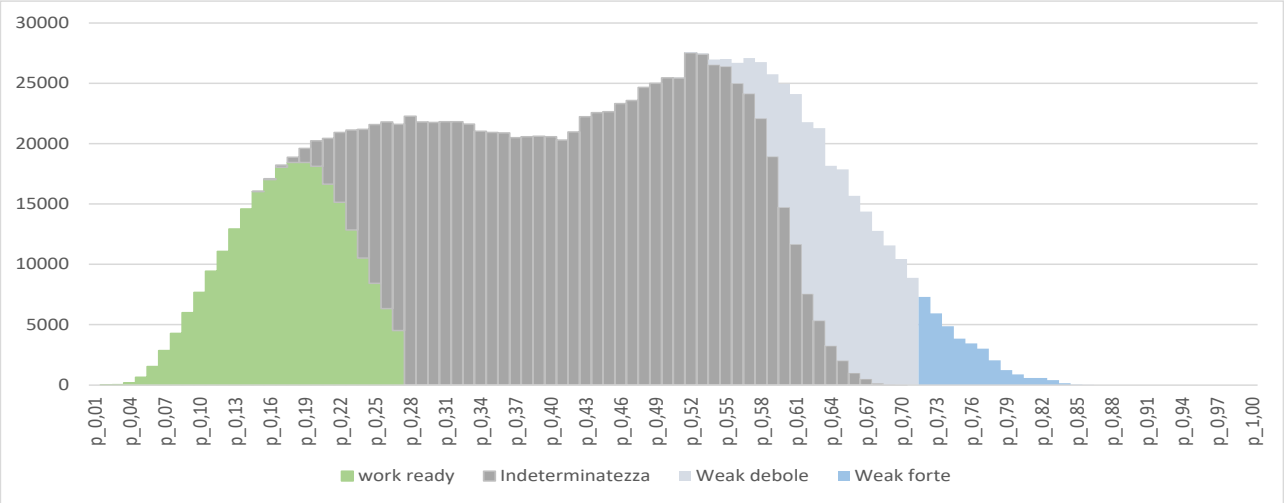
Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Target GOL

Consideriamo alcuni target del Programma GOL. L'individuazione dei gruppi target all'interno della platea dei disoccupati amministrativi avverrà per il tramite di *proxy*.

Per ciascun target verrà indicata la quota di popolazione in esso ricompresa, la suddivisione nelle tre zone - work-ready, indeterminatezza e weakness -, e la percentuale di falsi negativi per y_0 all'interno della categoria work-ready.

Target 1: Giovani under 30

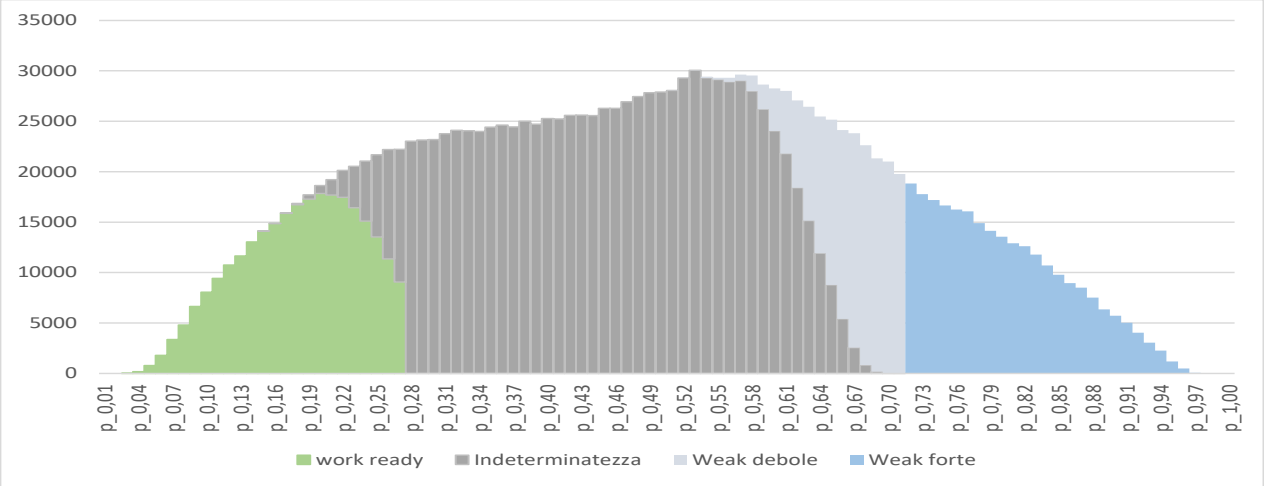


	% su pop T	% Pop su pop Under 30	Mean pr(y0)	Mean pr(y1)
Work-ready	38,1	19,1	0,170	0,273
Indeterminatezza	41,5	65,1	0,427	0,543
Weak debole	55,9	13,1	0,640	0,749
Weak forte	10,1	2,6	0,744	0,816
Totale	**39,0**	**100,0**	**0,414**	**0,525**

	Work-ready	Inderterminatezza	Weak debole	Weak forte
Titolo istruzione basso	12,6	57,2	24,3	5,9
Titolo istruzione medio	20,7	68,7	9,4	1,2
Titolo istruzione alto	28,3	69,7	1,6	0,3
Totale	**19,2**	**65,1**	**13,2**	**2,6**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

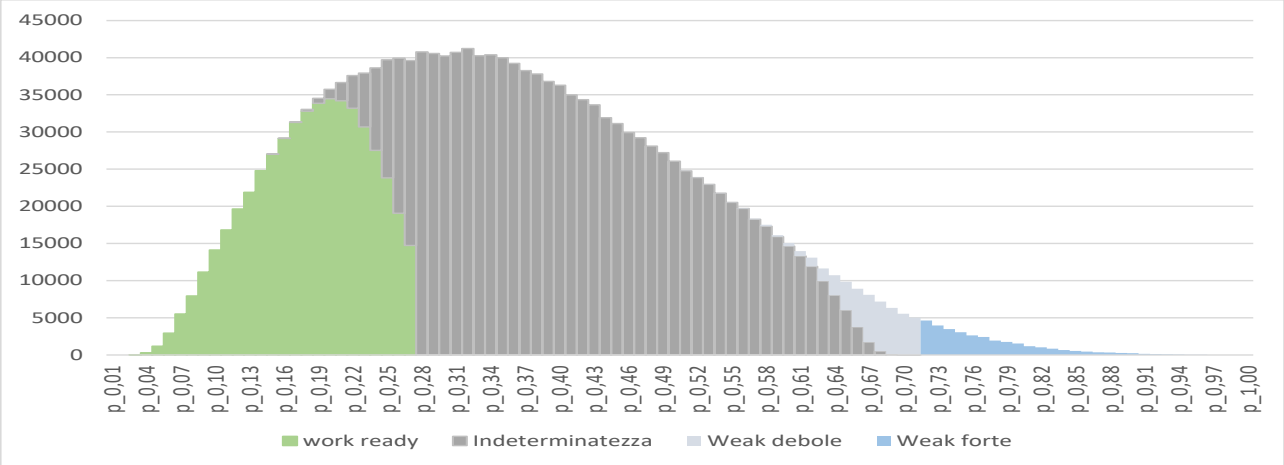
Target 2: Donne



	% su pop T	% Pop su pop Donne	Mean pr(y0)	Mean pr(y1)
Work-ready	40,6	15,7	0,177	0,252
Indeterminatezza	48,3	58,2	0,447	0,542
Weak debole	61,2	11,1	0,660	0,748
Weak forte	75,2	15,0	0,796	0,841
Totale	**50,7**	**100,0**	**0,481**	**0,564**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

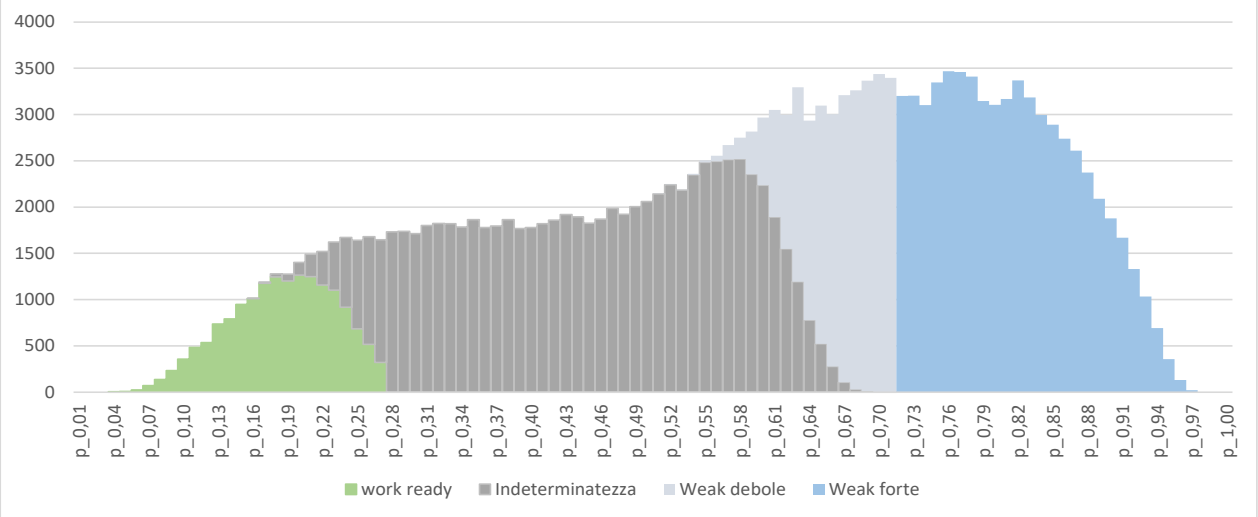
Target 3: Naspi



	% su pop T	% Pop su pop Naspi	Mean pr(y0)	Mean pr(y1)
Work-ready	75,4	28,6	0,177	0,263
Indeterminatezza	56,5	66,9	0,407	0,503
Weak debole	14,9	2,7	0,667	0,736
Weak forte	9,3	1,8	0,764	0,807
Totale	**51,6**	**100,0**	**0,355**	**0,446**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

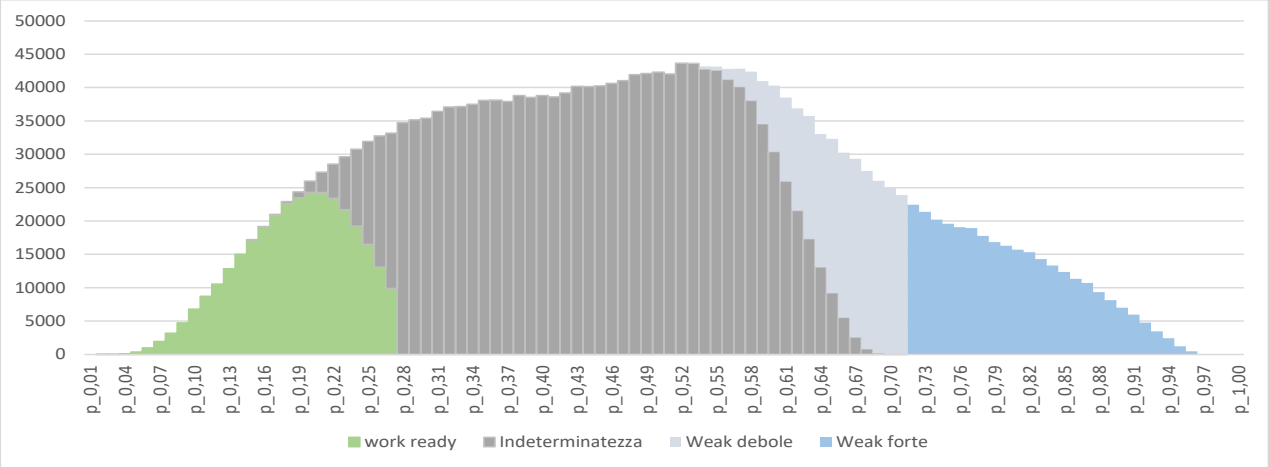
Target 4: Reddito di cittadinanza (*)



	% su pop T	% Pop su pop Rdc	Mean pr(y0)	Mean pr(y1)
Work-ready	N.d.	8,7	0,180	0,287
Indeterminatezza	N.d.	41,8	0,446	0,555
Weak debole	N.d.	16,3	0,660	0,751
Weak forte	N.d.	33,3	0,809	0,854
Totale	**N.d.**	**100,0**	**0,578**	**0,663**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

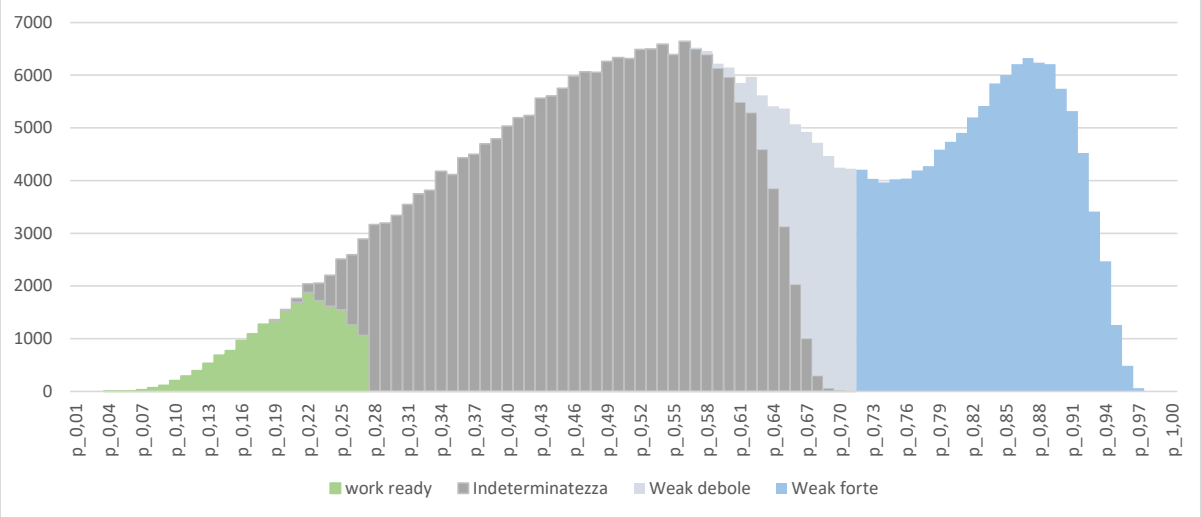
Target 5: LTU 6 mesi



	% su pop T	% Pop su pop Ltu	Mean pr(y0)	Mean pr(y1)
Work-ready	48,6	13,5	0,183	0,281
Indeterminatezza	71,6	62,2	0,438	0,539
Weak debole	86,7	11,3	0,653	0,747
Weak forte	90,6	13,0	0,797	0,843
Totale	**70,4**	**100,0**	**0,475**	**0,567**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

Target 6: Over 55



	% su pop T	% Pop su pop Over55	Mean pr(y0)	Mean pr(y1)
Work-ready	3,0	5,4	0,198	0,282
Indeterminatezza	10,0	55,6	0,471	0,557
Weak debole	9,8	8,2	0,671	0,739
Weak forte	33,4	30,7	0,830	0,868
Totale	**11,0**	**100,0**	**0,583**	**0,653**

Fonte: Elaborazioni su dati Anpal (SAP, DID) e MLPS – Comunicazioni Obbligatorie

5. Conclusioni: Definizione delle Classi di profilazione

La finalità del nuovo sistema di Profilazione quantitativa è di fornire all'operatore del CPI una prima indicazione sul livello di occupabibilità dell'utente[5], sulla base del rischio di diventare disoccupato di lunga durata. A tal fine sono definite tre classi di rischio:

- 1 – basso (rischio)
- 2 – medio (rischio)
- 3 – alto (rischio)

In particolare, nella classe "1 – Basso" rientrano gli utenti per i quali la combinazione dei livelli di profiling dei due modelli y_0 e y_1 descrivono la situazione di "work-ready". Nella classe "2 – Medio" rientrano gli utenti per i quali la combinazione dei livelli di profiling dei due modelli y_0 e y_1 dà luogo ad indeterminatezza. Infine, nella classe "3 – Alto" rientrano gli utenti per i quali la combinazione dei livelli di profiling dei due modelli y_0 e y_1 dà luogo a situazioni di debolezza (debole o forte).

<table>
<tr><th colspan="2">Combinazione p(y0) e p(y1)</th><th>Livello di profilazione in classi</th></tr>
<tr><td rowspan="2">$P(y_0)<=0,27$</td><td>$p(y_1)<=0,36$</td><td>1. Basso</td></tr>
<tr><td>$p(y_1)> 0,36$</td><td rowspan="2">2. Medio</td></tr>
<tr><td>$0,27<P(y_0)<=0,71$</td><td>$p(y_1)<= 0,70$</td></tr>
<tr><td>$0,27<P(y_0)<=0,71$</td><td>$p(y_1)> 0,70$</td><td rowspan="2">3. Alto</td></tr>
<tr><td>$p(y_0)>0,71$</td><td></td></tr>
</table>

[5] Il nuovo sistema di profilazione quantitativa è soltanto uno dei complessivi strumenti di cui si compone la fase di assessment dell'utente. Non c'è dunque *automatismo* alcuno tra l'indice di profiling quantitativo e la definizione delle politiche e dei percorsi dell'utente.

APPENDICE

A1 - TAVOLE STASTICHE CAP. 2.4

Tavola 2.4: Quota di donne nella platea di riferimento. Differenze 2018-2019 per Regione

	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	0,525	0,522	-0,0034	0,00	-1,58	0,11
Valle d'Aosta	0,509	0,530	0,0206	0,01	1,76	0,08
Lombardia	0,504	0,519	0,0149	0,00	7,83	0,00
PA Bolzano	0,574	0,583	0,0088	0,00	1,82	0,07
Pa Trento	0,573	0,576	0,0032	0,00	0,70	0,49
Veneto	0,568	0,553	-0,0141	0,00	-6,88	0,00
Friuli Venezia Giulia	0,545	0,556	0,0105	0,00	2,80	0,01
Liguria	0,550	0,546	-0,0043	0,00	-1,28	0,20
Emilia Romagna	0,538	0,542	0,0048	0,00	2,22	0,03
Toscana	0,549	0,549	0,0005	0,00	0,25	0,80
Umbria	0,556	0,555	-0,0009	0,00	-0,20	0,84
Marche	0,530	0,538	0,0085	0,00	2,52	0,01
Lazio	0,494	0,499	0,0040	0,00	1,79	0,07
Abruzzo	0,473	0,479	0,0059	0,00	1,66	0,10
Molise	0,437	0,455	0,0179	0,01	2,40	0,02
Campania	0,455	0,478	0,0223	0,00	14,36	0,00
Puglia	0,467	0,485	0,0181	0,00	9,80	0,00
Basilicata	0,438	0,464	0,0264	0,00	5,59	0,00
Calabria	0,495	0,505	0,0097	0,00	4,07	0,00
Sicilia	0,462	0,479	0,0167	0,00	11,20	0,00
Sardegna	0,493	0,493	-0,0002	0,00	-0,08	0,93

Tavola 2.5: Età media nella platea di riferimento. Differenze 2018-2019 per Regione

Età	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	35,791	36,237	0,4463	0,06	7,88	0,00
Valle d'Aosta	36,546	36,960	0,4141	0,31	1,34	0,18
Lombardia	**32,099**	**34,359**	**2,2602**	**0,05**	**45,48**	**0,00**
PA Bolzano	37,334	36,972	-0,3628	0,12	-2,99	0,00
Pa Trento	37,225	36,871	-0,3543	0,12	-2,95	0,00
Veneto	37,086	37,103	0,0172	0,05	0,32	0,75
Friuli Venezia Giulia	37,015	37,087	0,0722	0,10	0,73	0,47
Liguria	38,542	38,456	-0,0857	0,09	-0,99	0,32
Emilia Romagna	36,061	36,781	0,7199	0,06	12,56	0,00
Toscana	36,801	37,042	0,2407	0,05	4,48	0,00
Umbria	37,442	37,035	-0,4068	0,12	-3,43	0,00
Marche	36,223	36,042	-0,1806	0,09	-2,05	0,04
Lazio	34,955	35,708	0,7536	0,06	12,96	0,00
Abruzzo	36,576	36,893	0,3166	0,09	3,41	0,00
Molise	37,865	37,295	-0,5706	0,19	-2,95	0,00
Campania	35,572	35,666	0,0940	0,04	2,31	0,02
Puglia	34,799	34,818	0,0183	0,05	0,38	0,70
Basilicata	37,382	38,044	0,6627	0,13	5,22	0,00
Calabria	37,057	37,288	0,2306	0,06	3,80	0,00
Sicilia	36,193	36,704	0,5106	0,04	13,13	0,00
Sardegna	36,885	37,105	0,2200	0,07	3,05	0,00

Tavola 2.6: Età media per le donne nella platea di riferimento. Differenze 2018-2019 per Regione

Età Donne	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	36,652	36,926	0,2739	0,08	3,57	0,00
Valle d'Aosta	37,575	37,719	0,1445	0,42	0,34	0,73
Lombardia	32,772	34,893	2,1212	0,07	31,17	0,00
PA Bolzano	38,187	37,711	-0,4757	0,16	-2,98	0,00
Pa Trento	37,835	37,552	-0,2824	0,15	-1,83	0,07
Veneto	37,807	37,746	-0,0619	0,07	-0,88	0,38
Friuli Venezia Giulia	37,831	37,968	0,1374	0,13	1,06	0,29
Liguria	39,330	39,280	-0,0503	0,11	-0,44	0,66
Emilia Romagna	37,034	37,669	0,6349	0,08	8,24	0,00
Toscana	37,712	37,969	0,2574	0,07	3,63	0,00
Umbria	38,509	38,265	-0,2442	0,15	-1,58	0,11
Marche	37,233	36,975	-0,2578	0,12	-2,19	0,03
Lazio	35,178	36,062	0,8849	0,08	11,16	0,00
Abruzzo	36,848	37,229	0,3809	0,13	2,91	0,00
Molise	37,271	37,027	-0,2441	0,27	-0,89	0,37
Campania	35,331	35,691	0,3603	0,06	6,24	0,00
Puglia	34,640	34,710	0,0697	0,07	1,05	0,29
Basilicata	37,437	38,841	1,4034	0,19	7,57	0,00
Calabria	37,380	37,687	0,3075	0,08	3,72	0,00
Sicilia	36,065	36,910	0,8448	0,05	15,48	0,00
Sardegna	37,157	37,369	0,2118	0,10	2,12	0,03

Tavola 2.7: Età media per gli uomini nella platea di riferimento. Differenze 2018-2019 per Regione

Età Uomini	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	34,838	35,486	0,6475	0,08	7,75	0,00
Valle d'Aosta	35,479	36,105	0,6261	0,45	1,38	0,17
Lombardia	31,416	33,784	2,3680	0,07	32,67	0,00
PA Bolzano	36,187	35,939	-0,2473	0,18	-1,34	0,18
Pa Trento	36,408	35,945	-0,4631	0,19	-2,43	0,02
Veneto	36,139	36,307	0,1678	0,08	2,02	0,04
Friuli Venezia Giulia	36,037	35,985	-0,0517	0,15	-0,34	0,73
Liguria	37,576	37,464	-0,1118	0,13	-0,84	0,40
Emilia Romagna	34,931	35,729	0,7985	0,09	9,38	0,00
Toscana	35,694	35,912	0,2180	0,08	2,67	0,01
Umbria	36,103	35,498	-0,6050	0,18	-3,32	0,00
Marche	35,086	34,956	-0,1300	0,13	-0,98	0,32
Lazio	34,737	35,357	0,6195	0,08	7,29	0,00
Abruzzo	36,332	36,584	0,2516	0,13	1,92	0,05
Molise	38,327	37,518	-0,8085	0,27	-2,99	0,00
Campania	35,774	35,643	-0,1306	0,06	-2,28	0,02
Puglia	34,939	34,919	-0,0196	0,07	-0,28	0,78
Basilicata	37,339	37,354	0,0159	0,17	0,09	0,93
Calabria	36,741	36,881	0,1397	0,09	1,57	0,12
Sicilia	36,303	36,514	0,2110	0,06	3,83	0,00
Sardegna	36,620	36,849	0,2283	0,10	2,20	0,03

Tavola 2.8: Quota di disoccupati con titolo di studio basso. Differenze 2018-2019 per Regione

Titolo studio Basso	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	0,497	0,504	0,0069	0,00	3,24	0,00
Valle d'Aosta	0,583	0,534	-0,0482	0,01	-4,15	0,00
Lombardia	0,390	0,397	0,0068	0,00	3,64	0,00
PA Bolzano	0,533	0,520	-0,0131	0,00	-2,69	0,01
Pa Trento	0,462	0,467	0,0058	0,00	1,24	0,22
Veneto	0,443	0,454	0,0110	0,00	5,34	0,00
Friuli Venezia Giulia	0,431	0,446	0,0147	0,00	3,94	0,00
Liguria	0,455	0,446	-0,0092	0,00	-2,76	0,01
Emilia Romagna	0,496	0,486	-0,0107	0,00	-4,90	0,00
Toscana	0,468	0,465	-0,0035	0,00	-1,69	0,09
Umbria	0,421	0,439	0,0182	0,00	4,02	0,00
Marche	0,465	0,446	-0,0188	0,00	-5,58	0,00
Lazio	0,387	0,398	0,0110	0,00	5,00	0,00
Abruzzo	0,493	0,498	0,0051	0,00	1,43	0,15
Molise	0,428	0,459	0,0315	0,01	4,25	0,00
Campania	0,492	0,505	0,0127	0,00	8,14	0,00
Puglia	0,455	0,482	0,0268	0,00	14,52	0,00
Basilicata	0,476	0,504	0,0283	0,00	5,97	0,00
Calabria	0,422	0,436	0,0145	0,00	6,19	0,00
Sicilia	0,493	0,521	0,0282	0,00	18,83	0,00
Sardegna	0,516	0,521	0,0049	0,00	1,74	0,08

Tavola 2.9: Quota di disoccupati con titolo di studio medio. Differenze 2018-2019 per Regione

Titolo studio Medio	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	0,383	0,380	-0,0032	0,00	-1,52	0,13
Valle d'Aosta	0,324	0,359	0,0349	0,01	3,14	0,00
Lombardia	0,425	0,428	0,0026	0,00	1,36	0,17
PA Bolzano	0,410	0,420	0,0108	0,00	2,25	0,02
Pa Trento	0,404	0,397	-0,0077	0,00	-1,66	0,10
Veneto	0,396	0,401	0,0042	0,00	2,09	0,04
Friuli Venezia Giulia	0,424	0,410	-0,0132	0,00	-3,55	0,00
Liguria	0,411	0,425	0,0136	0,00	4,11	0,00
Emilia Romagna	0,366	0,367	0,0012	0,00	0,59	0,56
Toscana	0,380	0,384	0,0048	0,00	2,41	0,02
Umbria	0,433	0,415	-0,0178	0,00	-3,95	0,00
Marche	0,402	0,403	0,0010	0,00	0,30	0,77
Lazio	0,449	0,445	-0,0045	0,00	-2,03	0,04
Abruzzo	0,378	0,383	0,0050	0,00	1,46	0,15
Molise	0,428	0,413	-0,0150	0,01	-2,04	0,04
Campania	0,398	0,389	-0,0094	0,00	-6,14	0,00
Puglia	0,440	0,417	-0,0231	0,00	-12,64	0,00
Basilicata	0,410	0,387	-0,0235	0,00	-5,06	0,00
Calabria	0,443	0,422	-0,0206	0,00	-8,76	0,00
Sicilia	0,400	0,373	-0,0266	0,00	-18,25	0,00
Sardegna	0,362	0,365	0,0035	0,00	1,28	0,20

Tavola 2.10: Quota di disoccupati con titolo di studio alto. Differenze 2018-2019 per Regione

Titolo studio Alto	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	0,120	0,116	-0,0038	0,00	-2,74	0,01
Valle d'Aosta	0,093	0,106	0,0133	0,01	1,91	0,06
Lombardia	0,185	0,175	-0,0093	0,00	-6,39	0,00
PA Bolzano	0,057	0,060	0,0023	0,00	1,00	0,32
Pa Trento	0,134	0,136	0,0019	0,00	0,58	0,56
Veneto	0,161	0,146	-0,0152	0,00	-10,21	0,00
Friuli Venezia Giulia	0,145	0,143	-0,0015	0,00	-0,59	0,56
Liguria	0,134	0,129	-0,0044	0,00	-1,94	0,05
Emilia Romagna	0,138	0,148	0,0095	0,00	6,18	0,00
Toscana	0,152	0,151	-0,0013	0,00	-0,91	0,36
Umbria	0,146	0,146	-0,0004	0,00	-0,12	0,91
Marche	0,133	0,151	0,0178	0,00	7,54	0,00
Lazio	0,164	0,157	-0,0065	0,00	-3,91	0,00
Abruzzo	0,129	0,119	-0,0101	0,00	-4,31	0,00
Molise	0,144	0,127	-0,0165	0,01	-3,22	0,00
Campania	0,110	0,107	-0,0033	0,00	-3,43	0,00
Puglia	0,105	0,101	-0,0036	0,00	-3,24	0,00
Basilicata	0,114	0,109	-0,0048	0,00	-1,61	0,11
Calabria	0,135	0,142	0,0061	0,00	3,70	0,00
Sicilia	0,107	0,106	-0,0016	0,00	-1,69	0,09
Sardegna	0,123	0,114	-0,0083	0,00	-4,59	0,00

Tavola 2.11: Quota di disoccupati con esperienze lavorative nell'anno precedente la DID. Differenze 2018-2019 per Regione

Lavorato ultimo anno	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	0,694	0,700	0,0066	0,00	3,36	0,00
Valle d'Aosta	0,685	0,708	0,0225	0,01	2,09	0,04
Lombardia	**0,431**	**0,596**	**0,1652**	**0,00**	**87,91**	**0,00**
PA Bolzano	0,904	0,924	0,0199	0,00	7,26	0,00
Pa Trento	0,809	0,837	0,0280	0,00	7,80	0,00
Veneto	0,751	0,776	0,0247	0,00	14,11	0,00
Friuli Venezia Giulia	0,702	0,728	0,0257	0,00	7,56	0,00
Liguria	0,775	0,736	-0,0391	0,00	-13,60	0,00
Emilia Romagna	0,680	0,731	0,0509	0,00	25,52	0,00
Toscana	0,734	0,752	0,0185	0,00	10,24	0,00
Umbria	0,682	0,674	-0,0083	0,00	-1,94	0,05
Marche	0,612	0,610	-0,0021	0,00	-0,64	0,52
Lazio	0,603	0,621	0,0186	0,00	8,48	0,00
Abruzzo	0,689	0,694	0,0056	0,00	1,69	0,09
Molise	0,612	0,601	-0,0106	0,01	-1,45	0,15
Campania	0,639	0,598	-0,0401	0,00	-26,47	0,00
Puglia	0,598	0,576	-0,0221	0,00	-12,15	0,00
Basilicata	0,723	0,738	0,0145	0,00	3,45	0,00
Calabria	0,614	0,571	-0,0433	0,00	-18,58	0,00
Sicilia	0,550	0,532	-0,0184	0,00	-12,33	0,00
Sardegna	0,729	0,721	-0,0079	0,00	-3,16	0,00

Tavola 2.12: Quota di disoccupati in nuclei famigliari con figli. Differenze 2018-2019 per Regione

Presenza figli	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	0,293	0,295	0,0014	0,00	0,73	0,47
Valle d'Aosta	0,313	0,305	-0,0078	0,01	-0,72	0,47
Lombardia	**0,247**	**0,284**	**0,0367**	**0,00**	**21,70**	**0,00**
PA Bolzano	0,325	0,324	-0,0010	0,00	-0,22	0,82
Pa Trento	0,326	0,332	0,0054	0,00	1,22	0,22
Veneto	0,356	0,349	-0,0071	0,00	-3,59	0,00
Friuli Venezia Giulia	0,288	0,295	0,0063	0,00	1,85	0,06
Liguria	0,356	0,367	0,0115	0,00	3,57	0,00
Emilia Romagna	**0,285**	**0,320**	**0,0353**	**0,00**	**17,52**	**0,00**
Toscana	0,279	0,274	-0,0052	0,00	-2,79	0,01
Umbria	0,320	0,334	0,0140	0,00	3,28	0,00
Marche	0,338	0,360	0,0214	0,00	6,65	0,00
Lazio	**0,323**	**0,373**	**0,0504**	**0,00**	**23,49**	**0,00**
Abruzzo	0,240	0,252	0,0123	0,00	4,01	0,00
Molise	0,413	0,429	0,0163	0,01	2,21	0,03
Campania	0,250	0,245	-0,0052	0,00	-3,86	0,00
Puglia	0,347	0,347	0,0001	0,00	0,03	0,98
Basilicata	0,217	0,223	0,0054	0,00	1,38	0,17
Calabria	**0,342**	**0,369**	**0,0268**	**0,00**	**11,79**	**0,00**
Sicilia	**0,269**	**0,291**	**0,0224**	**0,00**	**16,65**	**0,00**
Sardegna	0,325	0,333	0,0084	0,00	3,18	0,00

Tavola 2.13: Quota di disoccupati donne in nuclei famigliari con figli. Differenze 2018-2019 per Regione

Presenza figli Donne	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	0,342	0,342	0,0001	0,00	0,03	0,98
Valle d'Aosta	0,383	0,376	-0,0068	0,02	-0,43	0,67
Lombardia	**0,322**	**0,343**	**0,0208**	**0,00**	**8,26**	**0,00**
PA Bolzano	0,375	0,382	0,0070	0,01	1,12	0,26
Pa Trento	0,382	0,387	0,0045	0,01	0,75	0,45
Veneto	0,417	0,411	-0,0057	0,00	-2,11	0,03
Friuli Venezia Giulia	0,348	0,360	0,0117	0,00	2,41	0,02
Liguria	0,418	0,442	0,0238	0,00	5,34	0,00
Emilia Romagna	**0,331**	**0,373**	**0,0426**	**0,00**	**14,93**	**0,00**
Toscana	0,308	0,303	-0,0049	0,00	-1,89	0,06
Umbria	0,362	0,386	0,0236	0,01	4,00	0,00
Marche	**0,384**	**0,417**	**0,0331**	**0,00**	**7,30**	**0,00**
Lazio	**0,363**	**0,422**	**0,0585**	**0,00**	**18,77**	**0,00**
Abruzzo	0,239	0,254	0,0154	0,00	3,45	0,00
Molise	0,456	0,467	0,0116	0,01	1,04	0,30
Campania	0,221	0,230	0,0091	0,00	4,76	0,00
Puglia	0,377	0,383	0,0060	0,00	2,31	0,02
Basilicata	0,204	0,216	0,0120	0,01	2,08	0,04
Calabria	**0,366**	**0,409**	**0,0438**	**0,00**	**13,39**	**0,00**
Sicilia	**0,263**	**0,292**	**0,0292**	**0,00**	**14,97**	**0,00**
Sardegna	0,361	0,374	0,0132	0,00	3,42	0,00

Tavola 2.14: Quota di disoccupati uomini in nuclei famigliari con figli. Differenze 2018-2019 per Regione

Presenza figli Uomini	2018	2019	Diff	SE	t	P(|t|>t)
Piemonte	0,240	0,243	0,0036	0,00	1,36	0,18
Valle d'Aosta	0,241	0,226	-0,0150	0,01	-1,05	0,30
Lombardia	**0,171**	**0,220**	**0,0491**	**0,00**	**22,56**	**0,00**
PA Bolzano	0,258	0,243	-0,0146	0,01	-2,25	0,02
Pa Trento	0,251	0,256	0,0056	0,01	0,89	0,38
Veneto	0,276	0,272	-0,0043	0,00	-1,55	0,12
Friuli Venezia Giulia	0,216	0,213	-0,0035	0,00	-0,76	0,45
Liguria	0,280	0,278	-0,0021	0,00	-0,47	0,64
Emilia Romagna	**0,232**	**0,258**	**0,0257**	**0,00**	**9,24**	**0,00**
Toscana	0,244	0,238	-0,0056	0,00	-2,13	0,03
Umbria	0,266	0,269	0,0022	0,01	0,37	0,71
Marche	0,287	0,293	0,0060	0,00	1,34	0,18
Lazio	**0,283**	**0,325**	**0,0416**	**0,00**	**14,26**	**0,00**
Abruzzo	0,241	0,250	0,0095	0,00	2,25	0,02
Molise	0,380	0,397	0,0177	0,01	1,81	0,07
Campania	0,274	0,258	-0,0160	0,00	-8,47	0,00
Puglia	0,321	0,313	-0,0075	0,00	-3,17	0,00
Basilicata	0,227	0,228	0,0008	0,01	0,15	0,88
Calabria	0,319	0,328	0,0086	0,00	2,73	0,01
Sicilia	**0,274**	**0,291**	**0,0164**	**0,00**	**8,87**	**0,00**
Sardegna	0,290	0,294	0,0038	0,00	1,06	0,29

A2 - STIME DEL MODELLO DI PROFILING

Tavola 1 – Stima modello logit y=y0 **(*)**

Logistic regression Number of obs = 3368727

LR chi2(177)= 567851.35

Prob>chi2=0

Log likelihood= -2023606.7

Pseudo R2 = 0.1230

Y0	X	Coeff.	P>z	Std. Err.
	donna	0.35681	***	0.00537
	età	-0.07314	***	0.00154
	età al quadrato	0.00125	***	0.00002
Interazione donna classe di età	25 a 29 anni	0.06064	***	0.00903
	30 a 39 anni	0.30868	***	0.01211
	40 a 49 anni	0.48346	***	0.01584
	over 50 anni	0.45142	***	0.01835
	donna - 25 a 29 anni	0.08254	***	0.00812
	donna - 30 a 39 anni	0.12333	***	0.00769
	donna - 40 a 49 anni	-0.09954	***	0.00810
	donna - over 50 anni	-0.12999	***	0.00798
Durata presenza in Italia (rif. Cittadino italiano)	Nato In Italia	-0.11087	***	0.00907
	Fino a 12 mesi	-0.02722	***	0.00700
	Da 1 a 2 anni	-0.07374	***	0.01121
	Oltre 2 anni	0.07119	***	0.00470
Interazione classe età e precedente esperienza di lavoro 12 mesi pre DID	Occupato data DID	0.20041	***	0.00533
	Prec. Esperienza - inferiore 24 anni	-0.32165	***	0.00724
	Prec. Esperienza - 25 a 29 anni	0.08490	***	0.00837
	Prec. Esperienza - 30 a 39 anni	-0.12505	***	0.00808
	Prec. Esperienza - 40 a 49 anni	-0.49783	***	0.00850
	Prec. Esperienza - over 50	-0.69592	***	0.00882
Titolo di studio (cat. Rif. Fino licenza elementare)	Licenza media	-0.07299	***	0.00471
	Qualifica professionale	-0.25288	***	0.00710
	Istituto professionale	-0.28140	***	0.00663
	Istituto tecnico	-0.18764	***	0.00563
	Liceo	-0.02406	***	0.00705
	Diploma: altro	-0.17521	***	0.00639
Diploma o triennale	Scienze umanistiche	-0.35588	***	0.01217
	Scienze sociali	-0.20022	***	0.01106
	Scienze della salute	-0.51349	***	0.01571
	Ingegneria, informatica e trasporti	-0.54653	***	0.01927
	Altro diploma o triennale	-0.23862	***	0.01337
Laurea magistrale, specialistica, vecchio ordinamento	Scienze umanistiche	-0.55807	***	0.01115
	Scienze sociali	-0.25735	***	0.00893
	Scienze della salute	-0.66594	***	0.01648
	Ingegneria, informatica e trasporti	-0.74078	***	0.01574
	Scienze naturali	-0.57046	***	0.02125
	Architettura	-0.04156	**	0.02023

	Altra laurea magistrale, specialistica, vecchio ordinamento	-0.34771	***	0.01081
Possedere una patente		-0.18909	***	0.00293
Condizione prevalente anno precedente (dichiarata) cat. Rif. Altro Inattivo	Occupato	-0.00667		0.00474
	In cerca di nuova occ.	-0.33050	***	0.00486
	In cerca di prima occ.	0.06591	***	0.00588
	Studente	-0.14136	***	0.00587
	Ha svolto tirocinio nei 12 mesi precedenti DID	-0.13303	***	0.00800
Professione prevalente 24 mesi precedenti DID (cat. Non ha lavorato nel periodo)	Bassa qualifica	-1.15914	***	0.01089
	Media qualifica	-1.12805	***	0.01073
	Alta qualifica	-1.26750	***	0.01150
Interazione donna e Settore prevalente 24 mesi precedenti DID (cat. Non ha lavorato nel periodo)	Agricoltura	-0.82309	***	0.01254
	Industria in senso stretto	0.02307	**	0.00917
	Costruzioni	-0.21646	***	0.00934
	Commercio	0.12146	***	0.00964
	Alloggio e Ristorazione	-0.45065	***	0.00934
	Trasporto Immagazzinaggio e Altri servizi di mercato	-0.14047	***	0.00874
	P.A., Istruzione e Sanità	-0.35793	***	0.01267
	donna - Agricoltura	-0.19668	***	0.01685
	donna - Industria in senso stretto	-0.08779	***	0.00968
	donna - Costruzioni	0.61878	***	0.02061
	donna - Commercio	-0.13387	***	0.00910
	donna - Alloggio e Ristorazione	-0.05079	***	0.00843
	donna - Trasporto Immagazzinaggio e Altri servizi di mercato	0.01362	*	0.00790
	donna - P.A., Istruzione e Sanità	-0.54571	***	0.01265
	donna - Altri servizi pubblici, sociali e personali	0.06949	***	0.01015
Numero datori cambiati 24 mesi precedenti DID (cat. Non ha lavorato nel periodo)	1 datore di lavoro	0.86323	***	0.00433
	2 datori di lavoro	0.49742	***	0.00467
	Più di 2 datori di lavoro	0.00000		0.00000
Impegno familiare	Famiglia con figli	-0.05840	***	0.00459
	Donna con figli	0.37329	***	0.00578
Numero componenti famiglia (cat. Rif. Unico componente)	2 componenti	0.01955	***	0.00436
	3 componenti	-0.06298	***	0.00434
	4 componenti	-0.13755	***	0.00434
	5 componenti	-0.10155	***	0.00534
	Più di 5 componenti	0.00871		0.00747
Costante		1.65022	***	0.02659

(*) non si riportano i coefficienti provinciali

Figura 1 - Curva ROC, grafico sensitivity-specificity. Modello y0

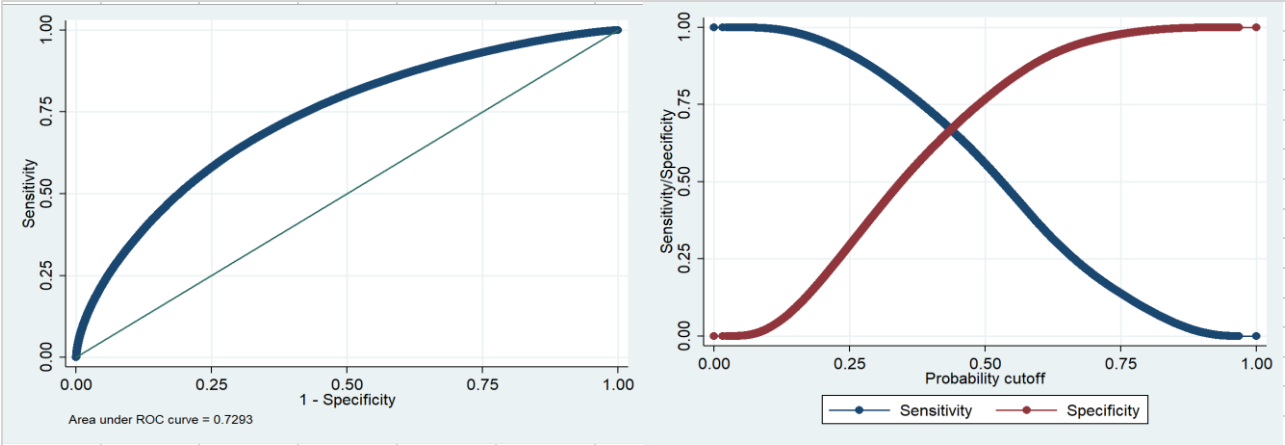


Tavola 2 – Stima modello logit y=y1 **(*)**

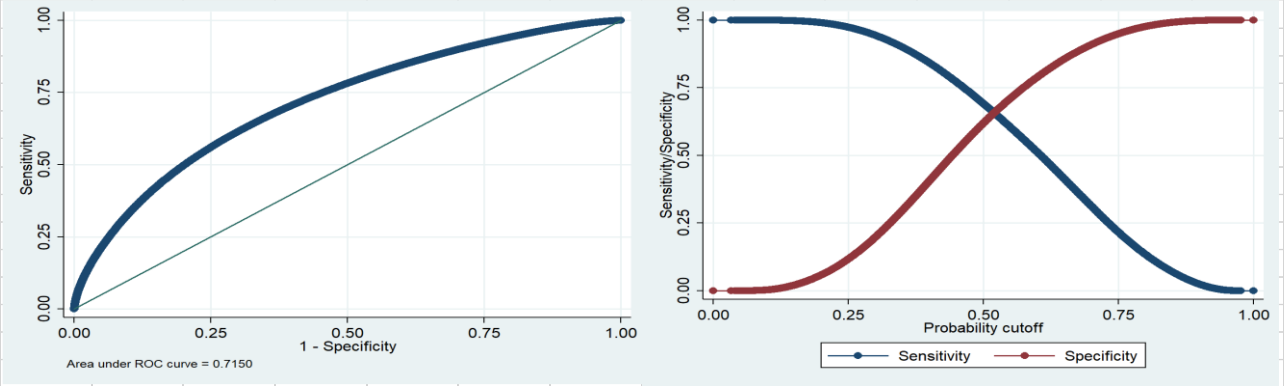
Logistic regression Number of obs = 3368727
LR chi2(177)= 504774.02
Prob>chi2=0
Log likelihood= -2077162.7
Pseudo R2 = 0.1083

Y1	X	Coeff.	P>z	Std. Err.
	donna	0.33548	***	0.00558
	età	-0.11543	***	0.00154
	età al quadrato	0.00162	***	0.00002
Interazione donna classe di età	25 a 29 anni	0.10157	***	0.00895
	30 a 39 anni	0.51155	***	0.01203
	40 a 49 anni	0.79982	***	0.01572
	over 50 anni	0.80380	***	0.01827
	donna - 25 a 29 anni	0.05075	***	0.00789
	donna - 30 a 39 anni	0.08164	***	0.00752
	donna - 40 a 49 anni	-0.13385	***	0.00795
	donna - over 50 anni	-0.17954	***	0.00794
Durata presenza in Italia (rif. Cittadino italiano)	Nato In Italia	-0.04788	***	0.00902
	Fino a 12 mesi	0.05647	***	0.00693
	Da 1 a 2 anni	-0.06590	***	0.01100
	Oltre 2 anni	0.05690	***	0.00462
Interazione classe età e precedente esperienza di lavoro 12 mesi pre DID	Occupato data DID	0.13278	***	0.00511
	Prec. Esperienza - inferiore 24 anni	-0.22344	***	0.00702
	Prec. Esperienza - 25 a 29 anni	0.03269	***	0.00815
	Prec. Esperienza - 30 a 39 anni	-0.22839	***	0.00804
	Prec. Esperienza - 40 a 49 anni	-0.59969	***	0.00858
	Prec. Esperienza - over 50	-0.81599	***	0.00908
Titolo di studio (cat. Rif. Fino licenza elementare)	Licenza media	-0.06629	***	0.00473
	Qualifica professionale	-0.28030	***	0.00694
	Istituto professionale	-0.31526	***	0.00652
	Istituto tecnico	-0.24783	***	0.00561
	Liceo	-0.05439	***	0.00705
Diploma o triennale	Diploma: altro	-0.20784	***	0.00637
	Scienze umanistiche	-0.38502	***	0.01173

Laurea magistrale, specialistica, vecchio ordinamento	Scienze sociali	-0.30349	***	0.01087
	Scienze della salute	-0.59178	***	0.01512
	Ingegneria, informatica e trasporti	-0.74215	***	0.01872
	Altro diploma o triennale	-0.30420	***	0.01306
	Scienze umanistiche	-0.60221	***	0.01066
	Scienze sociali	-0.37196	***	0.00881
	Scienze della salute	-0.78932	***	0.01595
	Ingegneria, informatica e trasporti	-0.94553	***	0.01531
	Scienze naturali	-0.68047	***	0.02040
	Architettura	-0.18987	***	0.02008
	Altra laurea magistrale, specialistica, vecchio ordinamento	-0.46654	***	0.01060
Possedere una patente		-0.17072	***	0.00289
Condizione prevalente anno precedente (dichiarata) cat. Rif. Altro Inattivo	Occupato	-0.08474	***	0.00473
	In cerca di nuova occ.	-0.34093	***	0.00487
	In cerca di prima occ.	0.04654	***	0.00611
	Studente	-0.14670	***	0.00595
Ha svolto tirocinio nei 12 mesi precedenti DID		-0.19066	***	0.00764
Professione prevalente 24 mesi precedenti DID (cat. Non ha lavorato nel periodo)	Bassa qualifica	-0.80958	***	0.01050
	Media qualifica	-0.79852	***	0.01034
	Alta qualifica	-0.96198	***	0.01105
Interazione donna e Settore prevalente 24 mesi precedenti DID (cat. Non ha lavorato nel periodo)	Agricoltura	-0.76503	***	0.01132
	Industria in senso stretto	-0.08283	***	0.00879
	Costruzioni	-0.22451	***	0.00891
	Commercio	0.01988	**	0.00926
	Alloggio e Ristorazione	-0.38577	***	0.00872
	Trasporto Immagazzinaggio e Altri servizi di mercato	-0.17384	***	0.00831
	P.A., Istruzione e Sanità	-0.49687	***	0.01192
	donna - Agricoltura	-0.21871	***	0.01530
	donna - Industria in senso stretto	-0.02847	***	0.00967
	donna - Costruzioni	0.58425	***	0.02115
	donna - Commercio	-0.07447	***	0.00908
	donna - Alloggio e Ristorazione	-0.07571	***	0.00803
	donna - Trasporto Immagazzinaggio e Altri servizi di mercato	0.03109	***	0.00783
	donna - P.A., Istruzione e Sanità	-0.51840	***	0.01202
	donna - Altri servizi pubblici, sociali e personali	0.00643		0.01003
Numero datori cambiati 24 mesi precedenti DID (cat. Non ha lavorato nel periodo)	1 datore di lavoro	0.63592	***	0.00390
	2 datori di lavoro	0.33012	***	0.00421
	Più di 2 datori di lavoro	0.00000		0.00000
Impegno familiare	Famiglia con figli	-0.09423	***	0.00443
	Donna con figli	0.38334	***	0.00568
Numero componenti famiglia (cat. Rif. Unico componente)	2 componenti	-0.00100		0.00429
	3 componenti	-0.08281	***	0.00425
	4 componenti	-0.16043	***	0.00424
	5 componenti	-0.11227	***	0.00524
	Più di 5 componenti	0.00811		0.00743
Costante		2.87448	***	0.02662

(*) non si riportano i coefficienti provinciali

Figura 2 - Curva ROC, grafico sensitivity-specificity. Modello y1



A3 - VARIABILI DI INPUT DELL'ALGORITMO DI CALCOLO DELLA CLASSE DI PROFILING IMPLEMENTATO NEL SISTEMA INFORMATIVO UNITARIO DELLE POLITICHE ATTIVE DEL LAVORO (SIU)

Variabile	Dato richiesto in cooperazione applicativa	Tipo di Input	Origine dato	Note
Genere	NO	Proposto da sistema. Non modificabile.	Recuperato da SAP.	
Età (in anni compiuti)	NO	Proposto da sistema. Non modificabile.	Calcolato da data nascita presente in SAP.	Anni compiuti al momento del calcolo profiling.
Durata presenza in Italia	SI	Selezione da una lista di valori.	Input dell'utente.	
Occupato al momento della data calcolo profiling	NO	Calcolato da sistema non modificabile.	Si, se nella tabella Rapporti di lavoro (RL)totale esiste un rapporto attivo alla data calcolo con tipo di contratto diverso da tirocinio e altre tipologie contrattuali atipiche.	
Ha lavorato nei 12 mesi precedenti data calcolo profiling (esclusi eventuali rapporti di lavoro in corso)	NO	Calcolato da sistema.	Si, se nella tabella RL totale esiste un rapporto concluso nei 12 mesi precedenti alla data calcolo (incluso il giorno del calcolo) tipo di contratto diverso da tirocinio e altre tipologie contrattuali atipiche.	
Titolo di studio più elevato conseguito	SI	Selezione da lista titoli a 4 livelli.	Input dell'utente	Dal titolo di studio inserito viene ricavato il codice del titolo di studio utilizzato nella stima del modello (vedi tavv. 1 e 2 in Allegato)
Possedere una patente	SI	Selezione Si/No	Input dell'utente	
Condizione professionale anno precedente	SI	Selezione da una lista di valori.	Input dell'utente	
Ha svolto tirocinio nei 12 mesi precedenti il calcolo profiling	NO	Calcolato da sistema non modificabile.	Sì, se nella tabella RL totale esiste un rapporto di tipo tirocinio concluso o iniziato nei 12 mesi precedenti alla data calcolo.	
Qualifica prevalente 24 mesi precedenti il calcolo profiling	NO	Calcolato da sistema non modificabile.	Qualifica presente nel RL prevalente nei 24 mesi precedenti ricondotti alla classificazione LOW (COD_L1=8 o 10), MEDIUM (COD_L1=4 o 5 o 6 o 7 o 9), HIGH (COD_L1=1,2,3).	Nel caso di UNISOMM la qualifica deve essere quella dell'ultima missione.

Variabile	Dato richiesto in cooperazione applicativa	Tipo di Input	Origine dato	Note
Settore prevalente 24 mesi precedenti il calcolo profiling	NO	Calcolato da sistema non modificabile.	Settore presente nel Rapporto di lavoro prevalente nei 24 mesi precedenti.	Nel caso di UNISOMM il settore sarà definito con un valore di default.
Numero datori di lavoro distinti nei 24 mesi precedenti il calcolo profiling	NO	Calcolato da sistema non modificabile.	Conteggio dei CF distinti dei datori di lavoro indicati nei RL degli ultimi 24 mesi.	Rapporti di lavoro attivi negli ultimi 24 mesi.
Famiglia con figli a carico	SI	Selezione da una lista di valori.	Input dell'utente.	
Numero componenti nucleo familiare	SI	Selezione da una lista di valori.	Input dell'utente.	
Provincia domicilio	NO	Calcolato da sistema non modificabile.	Comune domicilio da SAP se presente, altrimenti da anagrafica MyAnpal.	

BIBLIOGRAFIA

AlgorithmWatch (2019). Automating society: Taking stock of a automated decision making in the EU. Berlin, AlgorithmWatch in cooperation with Bertelsmann Stiftung.

Altman, DG, Bland, JM. Diagnostic tests 2: Predictive values. BMJ. 1994 Jul 9;309(6947):102. doi: 10.1136/bmj.309.6947.102. PMID: 8038641; PMCID: PMC2540558.

Altman, DG, Bland, JM. Diagnostic tests. 1: Sensitivity and specificity. BMJ. 1994 Jun 11;308(6943):1552. doi: 10.1136/bmj.308.6943.1552. PMID: 8019315; PMCID: PMC2540489

Barnes, S-A., Wright, S., Irving, P., Deganis, I. (2015), "Identification of latest trends and current developments in methods to profile jobseekers in European public employment services". Final report, Brussels: Directorate-General for Employment, Social Affairs and Inclusion. European Commission.

Barnes, S.-A., Wright, S., Irving, P. and Deganis, I. (2015), Identification of latest trends and current developments in methods to profile jobseekers in European public employment services: final report.

Black, D.A., Smith, J.A., Plesca, M. and Shannon, S. (2003), Profiling UI claimants to allocate reemployment services: Evidence and Recommendations for States. Final Report to United States Department of Labor.

Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001)

Centra, M., De Minicis, M., Marocco, M., Gualtieri, V., 2016, Profiling e garanzia giovani, in Canal T. (a cura di), L'Italia fra Jobs Act ed Europa 2020, Isfol, I libri del Fondo Sociale Europeo, Spoleto (PG), Del Gallo Editori

Desiere, S., Langenbucher, K. and Struyven, L. (2019), Statistical profiling in public employment services. OECD Working Paper.

Eberts, R.W., O'Leary, C.J. and Wandner, S.A. (2002), Targeting employment services, WE Upjohn Institute.

Eubanks, V. (2018), Automating inequality: How high-tech tools profile, police, and punish the poor, St. Martin's Press.

Henman, P. (2004). 'Targeted! Population segmentation, electronic surveillance and governing the unemployed in Australia.' International Sociology, 19(2), 173–191.

Kleinberg, J., Mullainathan, S. and Raghavan, M. (2016), 'Inherent trade-offs in the fair determination of risk scores.' arXiv preprint arXiv:1609.05807.

Loxha, A. and Morgandi, M. (2014), Profiling the unemployed: a review of OECD experiences and implications for emerging economies. Social Protection and labor discussion paper. World Bank Group, Washington, DC.

Martin, K. (2018), 'Ethical implications and accountability of algorithms.' Journal of Business Ethics, 1–16.

McCullagh P. & Nelder (1989), Generalized linear models 2nd edition, Chapman and Hall, London.

OECD (2018), "Profiling tools for early identification of jobseekers who need extra support." Policy Brief on Activation Policies, OECD Publishing, Paris

OECD (1998), Early identification of jobseekers at risk of long-term unemployment: the role of profiling, OECD.

Pope, D.G. and Sydnor, J.R. (2011), 'Implementing anti-discrimination policies in statistical profiling models.' American Economic Journal: Economic Policy, 3, 3, 206–231.

Schwab, S. (1986), 'Is statistical discrimination efficient?' The American Economic Review, 76, 1, 228–234.

Weber, T., (2011), "Profiling systems for effective labour market integration". Thematic Synthesis Paper. European Commission. Directorate-General for Employment, Social Affairs and Inclusion
Wijnhoven, M. and Havinga, H. (2014), 'The Work Profiler: A digital instrument for selection and diagnosis of the unemployed.' Local Economy, 29, 6-7, 740–749.