

APPENDICE A

LA DETERMINAZIONE DELLE SOGLIE STRUTTURALI DI RISCHIO DELLE IMPRESE: UN'APPLICAZIONE DELLA METODOLOGIA ROC¹

La tassonomia delle imprese italiane sulla base del grado di rischio è ottenuta elaborando le informazioni provenienti dall'integrazione delle seguenti basi di dati:

- **Indagine COVID2** (novembre 2020), dalla quale si ricavano le informazioni sulla percezione di rischio, sulle difficoltà incontrate dalle imprese e sulle loro strategie di reazione;
- **Archivio Frame-SBS**, da cui si estraggono le variabili economiche e strutturali dell'universo delle unità produttive (anno 2018);
- **Archivio Racli**, che fornisce le informazioni sulla composizione del lavoro interno alle imprese (anno 2018);
- **Archivio Asia-gruppi**, che fornisce le informazioni sull'appartenenza e sul posizionamento delle imprese italiane all'interno di gruppi domestici e multinazionali (anno 2018);
- **Indagine Multiscopo del Censimento permanente sulle imprese (indagine MPS)**, da cui si estraggono dati sui profili relazionali e strategici delle unità produttive, oltre a quelle relative all'innovazione e alla digitalizzazione (anno 2018).

La variabile utilizzata per definire la condizione di rischio operativo delle imprese è tratta dalla indagine COVID2; dalle informazioni provenienti dalla base dati integrata (composta dagli archivi Frame-SBS, Racli, Asia-gruppi e indagine MPS) è stato costruito il seguente insieme di indicatori compositi:

- **Dimensione economica**, che include il numero di addetti, il volume d'affari, l'età dell'impresa e l'intensità di capitale (rapporto tra immobilizzazioni e addetti);
- **Competitività**, che include la produttività del lavoro (valore aggiunto per addetto), la profittabilità (rapporto tra margine operativo lordo e valore aggiunto), la competitività di costo (rapporto tra produttività del lavoro e retribuzione pro-capite);
- **Qualità delle risorse umane impiegate**, sintesi di indicatori riguardanti l'input e la gestione del lavoro dell'impresa (presenza di manager, quota di dipendenti a tempo indeterminato sul totale, quota di dipendenti con educazione terziaria sul totale, retribuzione media, presenza di investimenti in capitale umano);
- **Dimensione relazionale**, che sintetizza indicatori finalizzati a valutare la misura in cui l'impresa sia inclusa in una rete o comunque connessa al resto del sistema produttivo (numero di relazioni con altre imprese e loro tipologia, investimenti in internazionalizzazione, capacità di attivazione sul resto del sistema produttivo);
- **Tecnologia e digitalizzazione**, che sintetizza la presenza di investimenti in ricerca e sviluppo e digitalizzazione, la spesa per software e informatica, la presenza di attività di innovazione.

Ognuno degli indicatori è stato ottenuto utilizzando, dopo la normalizzazione che riporta alla stessa unità di misura variabili continue e discrete, un'analisi fattoriale che permette di sintetizzare le informazioni fornite dalle singole variabili.

¹ Appendice a cura di Federico Sallusti.

A partire da questi indicatori, la tassonomia è costruita in due stadi.

Nel primo, si definisce un indicatore composito che riassume le informazioni fornite dai cinque indicatori definiti sopra. Nel secondo, utilizzando la metodologia ROC (*Receiver Operating Characteristics*), si stimano i valori soglia di tale indicatore che consentono di discriminare in maniera efficiente le osservazioni sulla base del rischio operativo, misurato dall'indagine *COVID*. La metodologia ROC, largamente utilizzata in medicina, è stata già sperimentata per la stima di alcune componenti del sommerso economico all'interno delle nuove procedure sviluppate dall'Istat a partire dal 2014 e utilizzata per determinare le caratteristiche minime d'impresa per accedere ai mercati internazionali, nonché per l'analisi dell'internazionalizzazione delle imprese italiane².

La costruzione dell'indicatore composito è effettuata a partire da un modello logistico in cui la variabile dipendente è una misura della condizione di piena sostenibilità operativa (dove il valore 1 corrisponde a una condizione di assenza di rischio operativo) ed è stimato sulla base della popolazione dei rispondenti all'indagine *COVID2* (circa 40mila imprese, rappresentative dell'universo delle circa 1 milione di imprese con almeno 3 addetti):

$$Prob(Proxy_i = 1 | DE_i, C_i, QL_i, R_i, D_i, S_i, T_i) = \Lambda(\alpha_1 DE_i + \alpha_2 C_i + \alpha_3 QL_i + \alpha_4 R_i + \alpha_5 D_i + \alpha_6 S_i + \alpha_7 T_i) \quad (1)$$

Dove *Proxy* è la variabile dummy che approssima la condizione di piena sostenibilità operativa (essendo uguale a 1 quando l'impresa ha dichiarato di non essere a rischio, uguale a 0 nel caso opposto) Λ è la distribuzione cumulata della funzione logistica, *DE* è la dimensione economica, *C* è la competitività, *QL* è l'indicatore di qualità delle risorse umane impiegate, *R* è la componente relazionale, *D* è l'indicatore di innovazione e digitalizzazione, ed *S* e *T* sono variabili (in forma di dummy) che controllano per l'appartenenza al settore economico (in termini di Ateco a 2 digit) e alla ripartizione geografica.

Per ogni impresa, l'indicatore composito Z_i , che corrisponde al profilo dell'impresa rispetto all'informazione fornita dagli indicatori di partenza, è definito come segue:

$$Z_i = \hat{\alpha}_1 DE_i + \hat{\alpha}_2 C_i + \hat{\alpha}_3 QL_i + \hat{\alpha}_4 R_i + \hat{\alpha}_5 D_i \quad (2)$$

dove i pesi $\hat{\alpha}$ sono i coefficienti stimati nell'equazione (1). Nel secondo stadio dell'analisi, l'indicatore composito è utilizzato come variabile esplicativa nella stima di un secondo modello logistico che anche in questo caso presenta, come variabile dipendente, la probabilità di una impresa di non essere a rischio operativo. Il modello stimato diviene dunque:

$$Prob(Proxy_i = 1 | Z_i) = \Lambda(\lambda Z_i) \quad (3)$$

Sulla base delle informazioni ottenute nella stima del secondo modello logistico è possibile derivare la curva ROC, cioè la rappresentazione grafica del rapporto tra *Sensitivity* (ovvero la probabilità di individuare veri positivi) e *Specificity* (ovvero la probabilità di individuare falsi positivi) al variare dell'indicatore (Fawcett, 2006). Sulla base dell'andamento di tale curva, la cui area sottesa rappresenta anche una misura di bontà di adattamento del modello logistico, è possibile definire il valore soglia dell'indicatore che consente di discriminare le osservazioni attraverso l'indice di Youden (1950):

$$Z^* = h * Sensitivity - (1 - h) * (1 - Specificity)$$

dove *h* è il peso relativo assegnato ai due parametri.

² Si vedano ad esempio Istat (2016c) e Costa, Sallusti, Vicarelli e Zurlo (2019).

L'indice di Youden consente dunque di individuare diversi valori della soglia a seconda della rilevanza assegnata alla capacità di individuare quanti più veri positivi (a scapito del fatto di includere anche un certo numero di falsi positivi) o di evitare l'inclusione di quanti più falsi positivi (a scapito dell'esclusione di una frazione di veri positivi).

In particolare, per $h=0,5$ si ha una selezione "neutrale", per valori di $h>0,5$ si tende a privilegiare l'esclusione dei falsi positivi (ovvero di assegna lo status di positivo solo in presenza di un segnale molto forte), mentre per valori di $h<0,5$ si tende a privilegiare l'inclusione dei veri positivi (ovvero si assegna lo status di negativo solo in presenza di un segnale forte).

In questo contesto, al variare del posizionamento della soglia tanto l'allocatione delle osservazioni lungo la curva ROC quanto il livello dell'indicatore per ogni osservazione rimangono invariati. Lo spostamento della soglia fornisce una classificazione complementare con qualsiasi altra ottenibile all'interno dello spazio classificatorio definito dal modello. In questo modo, è possibile ottenere una tassonomia composta costituita dall'interazione tra le diverse classificazioni definite attraverso lo spostamento della soglia.

La definizione del valore di h consente quindi di stabilire il punto della curva ROC nel quale si posiziona la soglia di classificazione delle osservazioni rispetto al fenomeno (nel caso in esame: il passaggio da impresa "a rischio" e impresa "non a rischio"). In questo modo, si otterrà un valore Z^* che sarà considerato come soglia. Ciascuna impresa sarà dunque classificata a seconda che il suo valore Z_i risulti maggiore o minore di Z^* .

In questa analisi, si è deciso di estendere la classificazione oltre la determinazione di due sole categorie ("a rischio" e "non a rischio"), definendo due ulteriori soglie (oltre a quella canonica con valutazione "neutrale" con $h=0,5$) in corrispondenza di un valore di h rispettivamente superiore e inferiore a 0,5. In questo modo, ciascuna delle due classi definita sulla base della soglia "neutrale" è stata ulteriormente suddivisa in due sotto-classi. In particolare, le imprese sotto-soglia ("a rischio") sono state ulteriormente classificate in "a rischio" e "fragili", mentre quelle sopra-soglia ("non a rischio") sono state ulteriormente classificate in "resilienti" e "solide".

Questa ulteriore classificazione si basa sulla necessità di individuare, all'interno di ognuna delle categorie originarie, una distanza dalla soglia che sia significativa nel distinguere un differente grado di, rispettivamente, "rischiosità" (le "fragili" sono meno a rischio di quelle "a rischio") e "solidità" (le "solide" sono più solide delle "resilienti").

A tal fine si sono derivate tutte le soglie Z^{h*} corrispondenti a cento valori possibili di h (da 0,01 a 0,99). Sulla base di queste, si sono scelti i due valori di h (rispettivamente superiore e inferiore a 0,5) che consentissero la massima discriminazione al di sopra e al di sotto della corrispondente soglia: in particolare, si sono scelti i valori di h in corrispondenza dei quali si sono osservati il numero più elevato di cambiamenti di status. In questo modo, per valori di h inferiori a è possibile distinguere, all'interno delle imprese "non a rischio", quelle "solide" da quelle "meno solide" (assegnando lo status di "non a rischio" solo alle imprese molto al di sopra dalla soglia "neutrale") e, per valori di h superiori a 0,5, tra le imprese "a rischio" quelle "a rischio" da quelle "fragili" (assegnando lo status di "a rischio" solo alle imprese molto al di sotto della soglia "neutrale").

Individuati in questo modo i due valori di h (0,41 e 0,58 in questa analisi), combinando il posizionamento delle imprese rispetto alle tre soglie, si sono definite le seguenti quattro categorie:

- **A rischio:** imprese che risultano al di sotto della soglia definita da $h=0,41$;

- **Fragili:** imprese che risultano al di sotto della soglia definita da $h=0,5$ ma al di sopra di quella definita da $h=0,41$;
- **Resilienti:** imprese che risultano al di sopra della soglia definita da $h=0,5$ e al di sotto di quella definita da $h=0,58$;
- **Solide:** imprese che risultano al di sopra della soglia definita da $h=0,58$.

APPENDICE B

LA COSTRUZIONE DELL'INDICE DI RISCHIO TERRITORIALE

L'indice sintetico (Indice IRT) è ottenuto come media aritmetica di un dato numero (V) di indicatori elementari, ponderata con i contributi che questi ultimi forniscono alla determinazione di fattori ricavati dall'analisi in componenti principali (ACP) e considerati nell'analisi del fenomeno oggetto di studio (due o più).

Il processo di calcolo dell'indice prevede i seguenti passaggi:

- 1) *individuazione degli indicatori elementari per la rappresentazione del fenomeno da analizzare;*
- 2) *ACP effettuata utilizzando gli indicatori di cui al punto 1;*
- 3) *standardizzazione degli indicatori elementari di cui al punto 1;*
- 4) *media aritmetica ponderata degli indicatori standardizzati di cui al punto precedente, calcolata utilizzando come pesi le coordinate degli assi fattoriali e la varianza delle componenti ricavate dall'ACP.*

Formalmente, nel caso delle tre componenti principali considerate nel capitolo 4 del presente Rapporto (che complessivamente sintetizzano oltre l'82 per cento della varianza totale), l'indice CP (punto 4) è calcolato come segue:

$$\text{Indice } CP_i = \frac{\lambda_I \sum_{v=1}^V z_{vi} a_{Iv} + \lambda_{II} \sum_{v=1}^V z_{vi} a_{IIv} + \lambda_{III} \sum_{v=1}^V z_{vi} a_{IIIv}}{\lambda_I + \lambda_{II} + \lambda_{III}}$$

Dove z sono gli indicatori standardizzati per ciascuna variabile v e unità territoriale (SI) i , a rappresenta le coordinate degli assi fattoriali relative alle tre componenti principali (I, II e III) e λ esprime la varianza delle componenti principali. Il metodo presenta il vantaggio di basarsi su fattori incorrelati. Il peso attribuito alle variabili nel calcolo dell'indice di sintesi è dato dalle corrispondenti quote di varianza riprodotta dai fattori.

Per favorire l'interpretazione, dato che l'indice sintetico ottenuto può assumere valori sia positivi sia negativi, lo stesso è stato riproporzionato mediante procedure min-max facendolo dunque variare tra 0 ed 1, ovvero:

$$\text{Indice } CP_i^s = \frac{\text{Indice } CP_i - \text{Indice } CP_{\min}}{\text{Indice } CP_{\max} - \text{Indice } CP_{\min}}$$

APPENDICE C¹

VALUTAZIONE DELL'IMPATTO DELLA CHIUSURA DELLE ATTIVITÀ ECONOMICHE SUL FATTURATO E SUL VALORE AGGIUNTO DELLE IMPRESE A LIVELLO TERRITORIALE

L'esercizio è stato articolato in 3 fasi: (1) la proiezione al 2019 dei dati del Frame Sbs territoriale riferito al 2018 (scenario base); (2) la costruzione di tre diversi scenari di chiusura e di limitazione effettiva dello svolgimento delle attività economiche nel corso dell'anno e la stima delle variazioni di fatturato e del valore aggiunto; (3) il calcolo delle variazioni congiunturali delle variabili di interesse, a livello di divisione di attività economica (Ateco-2 digit), osservati nel corso del 2020.

1) *Proiezione dei dati del Frame Sbs territoriale al 2019*

Il primo passo ha riguardato la stima dei valori economici per il 2019 delle variabili contenute nel registro Frame Sbs territoriale, riferito all'anno 2018 (4,7 milioni di record; si veda Istat 2020, Risultati economici delle imprese e delle multinazionali a livello territoriale - <https://www.istat.it/it/archivio/252159>). La stima è stata effettuata integrando le seguenti basi di microdati: *a)* la versione anticipata del Registro Asia relativa al 2019, con circa 1,5 milioni di unità produttive per ciascuna delle quali sono presenti i valori relativi al numero degli addetti e (ove disponibile) il volume d'affari; *b)* i bilanci delle società di capitali, da cui sono stati tratti i valori di fatturato e di valore aggiunto per circa 700mila imprese).

In base a queste informazioni preliminari è stata calcolata anche una variazione delle variabili di interesse per strato (individuato dalla combinazione delle seguenti caratteristiche: Ateco a 3 digit, Regione, 5 classi di addetti), utilizzata a supporto dell'imputazione dei dati mancanti. Questo calcolo ha interessato tutte le imprese non incluse nella stima anticipata del Registro Asia 2019, ovvero circa 3 milioni di imprese non società di capitali e senza dipendenti, pari al 67 per cento del totale delle unità produttive e a un terzo del valore aggiunto complessivo.

2) *Costruzione degli scenari e stima delle perdite di fatturato*

Per il 2020 sono stati ipotizzati tre scenari di limitazione/chiusura delle attività in base ai provvedimenti entrati in vigore fino al 14 Aprile (lock1) e al 4 Maggio (lock2)². Inoltre sono state considerate chiusure ulteriori, differenziate a livello regionale in base ai livelli di rischio (giallo, arancione e rosso).

La riduzione delle attività è stata calcolata in base a tre ipotesi relative ai mesi di chiusura, una "soft" e due meno ottimistiche ("medium" e "hard"). L'ipotesi soft è relativa a un *lockdown* di un mese (ottimistica rispetto ai tempi effettivi del primo provvedimento). L'ipotesi hard introduce tre mesi in più di chiusura effettiva rispetto alla ipotesi soft, per tenere conto delle difficoltà di ripresa delle attività e del rispetto delle nuove regolamentazioni (si pensi alle limitazioni alla capienza dei ristoranti o agli ingressi contingentati nei locali aperti al pubblico). Si tratta quindi di assunzioni volte a indicare i limiti superiore e inferiore del calo di fatturato.

Nel dettaglio, le ipotesi sono le seguenti:

- 1) *Per le attività economiche interessate dal provvedimento di aprile 2020:*
 - a) lock1 soft: 1 mese;
 - b) lock1 medium: 2,5 mesi;
 - c) lock1 hard: 4 mesi.

¹ Appendice a cura di Filippo Oropallo.

² Si veda il comunicato Istat con la lista delle attività economiche soggette a chiusura: www.istat.it/it/archivio/241495.

Per le attività economiche interessate dal provvedimento di maggio 2020:

- a) lock2 soft: 2 mese;
- b) lock2 medium: 3,5 mesi;
- c) lock2 hard: 5 mesi.

Per le imprese localizzate in regioni con allerta arancione si prevede inoltre mezzo mese in più di chiusura, mentre per quelle in regioni in allerta rossa si conta un ulteriore mese di chiusura.

Tabella 1 - Regioni, numero di unità locali e grado di allerta nei mesi ottobre-dicembre

Regione	Unità locali		Grado di allerta
	N.	%	
Piemonte	347323	7,4	Rossa
Valle d'Aosta	11879	0,25	Rossa
Lombardia	872080	18,58	Rossa
Provincia autonoma di Bolzano	47505	1,01	Rossa
Provincia autonoma di Trento	44273	0,94	Gialla
Veneto	420283	8,95	Gialla
Friuli-Venezia Giulia	89738	1,91	Gialla
Liguria	132101	2,81	Arancione
Emilia Romagna	390527	8,32	Gialla
Toscana	344164	7,33	Arancione
Umbria	70596	1,5	Arancione
Marche	134700	2,87	Gialla
Lazio	467859	9,97	Gialla
Abruzzo	104494	2,23	Arancione
Molise	22227	0,47	Gialla
Campania	371988	7,92	Gialla
Puglia	270989	5,77	Arancione
Basilicata	37732	0,8	Arancione
Calabria	114941	2,45	Rossa
Sicilia	287041	6,11	Arancione
Sardegna	111823	2,38	Gialla
Totale	4694263	100	

In base a questa riduzione ipotizzata ($\Delta\%c_s$) per ogni scenario (s) è stata calcolata la riduzione dei ricavi nell'esercizio 2020 sulla base della seguente relazione:

$$\Delta R_{p,a,cl,i}^{20,s} = ru_{p,a,cl}^{19} \text{ addetti}_{p,a,cl,i}^{19} \Delta\%c_s \quad [1]$$

Dove p = Provincia, a = Ateco-3 cifre (gruppo), cl = Classe di addetti e i = Unità locale.

La riduzione è dunque calcolata a partire da una misura della produttività di strato, in termini di fatturato per addetto ($ru_{p,a,cl}^{19}$)³; questa viene riportata a livello d'impresa moltiplicandola per il numero di addetti dell'unità locale ($\text{addetti}_{p,a,cl,i}^{19}$), per poi ottenere la variazione del fatturato d'impresa applicando la riduzione ipotizzata nello scenario s di riferimento ($\Delta\%c_s$).

3 L'indicatore è pari al valore mediano di strato (provincia di appartenenza dell'unità locale, gruppo Ateco-3 digit e classe di addetti dell'unità locale) del fatturato per addetto. L'indicatore è stato imputato a tutte le unità locali dello strato di appartenenza al fine di avere una stima del rendimento per addetto in quel particolare contesto di mercato.

Successivamente la riduzione di fatturato è stata utilizzata nel calcolo del valore aggiunto, ipotizzando che: *i*) i costi variabili si riducano in proporzione al fatturato; *ii*) i costi fissi rimangano per metà costanti, mentre l'altra metà, al pari dei costi variabili, si riduce in proporzione al fatturato.

3) *Calibrazione finale*

Dal fatturato d'impresa così calcolato si è ottenuto un totale settoriale che è stato successivamente riproporzionato a livello di divisione economica (Ateco-2 digit) sulla base delle variazioni registrate dai dati mensili della fatturazione elettronica. In particolare, a partire da questi ultimi, sono state calcolate le variazioni congiunturali mensili relative al periodo gennaio-settembre 2020 (ultimo mese disponibile al momento della preparazione dell'esercizio); le variazioni annuali sono state infine calcolate, anche in questo caso, seguendo tre scenari – *ottimistico*, *pessimistico*, *intermedio* –, per tenere conto degli effetti della seconda ondata della pandemia:

- *Ottimistico* - la variazione media mensile osservata fino a settembre (media geometrica delle variazioni) viene applicata anche al quarto trimestre dell'anno (ottobre-dicembre);
- *Pessimistico* - al quarto trimestre viene applicata la variazione di fatturato media mensile osservata fino a giugno (media geometrica delle variazioni);
- *Intermedio* - media geometrica degli scenari ottimistico e pessimistico.

In particolare, i risultati dell'esercizio svolto nel paragrafo 4.2 si basano sullo scenario peggiore possibile: il contemporaneo verificarsi dell'ipotesi "lock1 *hard*", dell'ipotesi "lock2 *hard*" e dello scenario pessimistico relativo alla perdita di fatturato nel quarto trimestre.