



**Politecnico
di Torino**

Politecnico di Torino

Corso di Laurea Magistrale in Ingegneria Gestionale

A.a. 2025/2026

Sessione di Laurea Marzo 2026

La produttività dell'innovazione e il ruolo delle esternalità nel divario Nord-Sud italiano

Relatori:
Prof. Luigi Buzzacchi
Dott. Antonio De Marco

Candidata:
Greta Donadio

Indice dei contenuti

0.Introduzione	7
1.Innovazione, crescita economica e dimensione territoriale della conoscenza.....	8
1.1 L'evoluzione del concetto di innovazione.....	8
1.2 Innovazione e performance economica	9
1.3 Esternalità, spillover e dimensione territoriale dell'innovazione.....	11
2.Rassegna della letteratura	13
2.1 La letteratura sulla geografia dell'innovazione.....	13
2.2 La domanda di ricerca.....	15
3.Analisi di correlazione spaziale	16
3.1 Unità territoriale di riferimento.....	17
3.2 Orizzonte temporale	17
3.3 Dati occupazionali e armonizzazione territoriale.....	18
3.4 Criteri di attribuzione territoriale delle attività innovative	20
3.5 Strumenti di analisi	20
3.5.1. Moran's I.....	20
3.5.2. LISA.....	22
3.6 Brevetti EPO: analisi di distribuzione spaziale	23
3.6.1. Mappe coropletiche.....	23
3.6.2. Distribuzione della funzione EPO_weight_j	27
3.6.3. Moran's I globale e ordini di contiguità	30
3.6.4. LISA e significatività locale	31
3.7. Start-up innovative: analisi della distribuzione spaziale.....	34
3.7.1. Mappe coropletiche.....	35
3.7.2. Distribuzione della funzione $startup_weight_j$	38
3.7.3. Moran's I globale e ordini di contiguità	39
3.7.4. LISA e significatività locale	40
3.8. Brevetti UIBM: analisi della distribuzione spaziale.....	43
3.8.1. Mappe coropletiche.....	44
3.8.2. Distribuzione della funzione $UIBM_weight_j$	47
3.8.3. Moran's I globale e ordini di contiguità	48
3.8.4. LISA e significatività locale	50
4.Il modello di regressione OLS	52
4.1 Il modello empirico.....	52
4.2 Definizione e costruzione delle variabili di regressione.....	54

4.2.1. La specializzazione produttiva <i>PSij</i>	54
4.2.2. La diversità produttiva <i>PDj</i>	54
4.2.3. La specializzazione science-based <i>SBSij</i>	55
4.2.4. La diversità science-based <i>SBDij</i>	57
4.2.5. Le opportunità tecnologiche settoriali <i>TOi</i>	58
4.2.6. Le aree metropolitane <i>DMj</i>	61
4.2.7. I settori a più elevata propensione innovativa <i>DHTi</i>	63
4.2.8. La presenza di infrastrutture universitarie <i>UNIVj</i>	63
4.2.9. La presenza di infrastrutture universitarie <i>NORDj</i>	65
4.3 Conclusioni.....	66
5. Analisi dei risultati econometrici.....	67
5.1 Dataset di regressione.....	67
5.2 Modello di regressione OLS per i brevetti EPO.....	70
5.2.1 Le variabili della regressione: statistiche descrittive e correlazioni.....	70
5.2.2. Risultati della regressione per i brevetti EPO.....	72
5.3 Modello di regressione OLS per i brevetti UIBM	79
5.3.1 Le variabili della regressione: statistiche descrittive e correlazioni.....	79
5.3.2. Risultati della regressione per i brevetti UIBM	81
5.4 Risultati della regressione per le Start Up Innovative	87
5.3.1 Le variabili della regressione: statistiche descrittive e correlazioni.....	87
5.4.2. Risultati della regressione per le Start Up.....	89
6. Conclusioni	94
6.1 Limiti dello studio e sviluppi futuri.....	96
6.2 Implicazione di policy	97
Bibliografia.....	99

Indice delle figure

Figure 1: mappa dei SLL secondo l'aggiornamento del 2021	17
Figure 2: mappa coropletica (quantili) degli addetti per SLL secondo il censimento del 2011	18
Figure 3: boxplot della densità di addetti per km ² , il confronto Nord - Centro - Sud.....	19
Figure 4: mappa per quantili di EPO_weightj su 5 classi di uguale numerosità.....	24
Figure 5: mappa per percentili di EPO_weightj.....	24
Figure 6: mappa per quantili di num_EPOj, 5 classi di uguale dimensione	26
Figure 7: mappa per percentili di num_EPOj.....	27
Figure 8: mappa coropletica di EPO_weight, in nero gli SLL esclusi dall'analisi	28
Figure 9: distribuzione dell'indicatore EPO_weight a livello di SLL, istogramma in scala naturale	29

Figure 10: distribuzione dell'indicatore EPO_weight a livello di SLL, istogramma con asse orizzontale in scala logaritmica.....	29
Figure 11: Moran scatterplot di EPO_weight a livello di SLL	31
Figure 12: mappa LISA di EPO_weight a livello di SLL	32
Figure 13: mappa di significatività del Moran locale di EPO_weight a livello di SLL	33
Figure 14: mappa percentile di #startup a livello di SLL	35
Figure 15: mappa a quantili di #startup a livello di SLL.....	36
Figure 16: mappa a percentili di startup_weight j a livello di SLL	37
Figure 17: mappa a quantili di startup_weight j a livello di SLL	38
Figure 18: Moran scatterplot startup_weightj	40
Figure 19: mappa LISA di startup_weightj a livello di SLL.....	41
Figure 20: mappa di significatività del Moran locale di startup_weight a livello di SLL	42
Figure 21: mappa per percentili num_UIBM per SLL.....	44
Figure 22: mappa per quantili num_UIBM per SLL	45
Figure 23: mappa per percentili UIBM_weightj per SLL.....	46
Figure 24: mappa per quantili di UIBM_weightj per SLL.....	46
Figure 25: istogramma di $\log(1 + \text{UIBM_weightj})$	48
Figure 26: Moran Scatterplot per la funzione UIBM_weightj	49
Figure 27: mappa LISA di UIBM_weightj a livello di SLL.....	50
Figure 28: mappa di significatività del Moran locale di UIBM_weight a livello di SLL	51
Figure 29: mappatura degli SLL per i quali la $DM_j = 1$ dunque classificabili come aree metropolitane	62
Figure 30: mappatura della presenza universitaria nei SLL	64
Figure 31: statistiche descrittive delle variabili utilizzate nella regressione OLS dei brevetti EPO.....	70
Figure 32: matrice di correlazione tra le variabili del modello di regressione OLS per i brevetti EPO ..	71
Figure 33: regressione OLS per i brevetti EPO.....	73
Figure 34: grafico di confronto tra valori osservati e valori stimati, modello 2	74
Figure 35: grafico dei residui rispetto ai valori stimati per EPO, modello 2.....	75
Figure 36: coefficient plot con errori robusti HC1	76
Figure 37: statistiche descrittive delle variabili utilizzate nella regressione OLS dei brevetti UIBM	79
Figure 38: matrice di correlazione tra le variabili del modello di regressione OLS per i brevetti UIBM	80
Figure 39: esiti delle analisi di regressione OLS per i brevetti UIBM.....	81
Figure 40: grafico di confronto tra valori osservati e valori stimati, modello 2	82
Figure 41: grafico di confronto tra valori residui e valori stimati per UIBM, modello 2	83
Figure 42: statistiche descrittive delle variabili del modello di regressione OLS per le Start Up	87
Figure 43: matrice di correlazione tra le variabili della regressione OLS delle Start Up.....	88
Figure 44: analisi di regressione OLS su Start Up	89
Figure 45: grafico di confronto tra valori osservati e valori stimati per le Start Up, modello 2	90

Indice delle tabelle

Table 1: costruzione dei cluster regionali per l'analisi boxplot della densità di addetti per km^2	19
Table 2: riepilogo delle statistiche descrittive e quantili della distribuzione EPO_weight.....	28
Table 3: Correlazione spaziale Moran's I globale della produttività dell'innovazione per diversi ordini di contiguità	30
Table 4: classificazione dei cluster secondo l'analisi LISA.....	31

Table 5: i 9 UIBM SLL con il #startup più alto	36
Table 6: Correlazione spaziale Moran's I globale di startup_weightj per diversi ordini di contiguità	39
Table 7: classificazione dei cluster secondo l'analisi LISA.....	40
Table 8: riepilogo delle statistiche descrittive della funzione UIBM_weightj	48
Table 9: Correlazione spaziale Moran's I globale di UIBM_weightj per diversi ordini di contiguità	48
Table 10: classificazione dei cluster secondo l'analisi LISA.....	50
Table 11: classificazione dei settori industriali in cluster k	56
Table 12: distribuzione dei settori industriali (Ateco a 2 digit) nei cluster k	57
Table 13: IPC4 non mappati per assente corrispondenza con NACE Rev.2.....	60
Table 14: classificazione degli SLL nei due cluster Nord - Sud.....	65
Table 15: brevetti persi nella costruzione di P_ija causa della mappatura IPC4 - NACE Rev. 2	69
Table 16: numero di osservazioni (i,j) su cui sono definite le variabili dipendenti.....	69
Table 17: risultati dei test diagnostici condotti sul modello di regressione OLS (2) sui brevetti EPO....	74
Table 18: risultati dei test diagnostici condotti sul modello di regressione OLS sui brevetti UIBM	82
Table 19: risultati dei test diagnostici condotti sul modello di regressione OLS sulle Start Up, modello 2	90

0. Introduzione

La letteratura sulla geografia dell'innovazione parte dall'assunto che la produzione di conoscenza presenti una configurazione spaziale non casuale, ma strutturalmente radicata nei contesti territoriali (Feldman, 1994; Audretsch e Feldman, 1996). Numerosi contributi hanno mostrato come l'attività innovativa tenda a concentrarsi geograficamente in presenza di esternalità locali e spillover di conoscenza, evidenziando l'importanza della prossimità e della struttura produttiva territoriale (Jaffe et al., 1993; Boschma, 2005). Con riferimento al contesto italiano, Paci e Usai (2000) hanno documentato significativi fenomeni di autocorrelazione spaziale dell'innovazione a livello di Sistemi Locali del Lavoro, sottolineando il ruolo delle economie di specializzazione e di urbanizzazione nel plasmare la distribuzione territoriale dell'output innovativo.

In parallelo, la letteratura ha evidenziato come le traiettorie innovative dei territori assumano carattere cumulativo e storicamente condizionato, nel senso che la nuova conoscenza si sviluppa più facilmente laddove esistano competenze e specializzazioni affini (Boschma e Frenken, 2011; Balland, 2016).

In questo quadro teorico, l'innovazione non è il risultato di shock isolati, ma riflette la combinazione tra dotazioni territoriali, capacità locali e meccanismi di diffusione della conoscenza.

Lo scopo di questa tesi è analizzare le variabili che incidono sulla produttività dell'innovazione nei Sistemi Locali del Lavoro italiani e verificare se siano presenti delle differenze strutturali riconducibili al persistente divario tra Nord e Sud.

La ricerca adotta una prospettiva territoriale sub-regionale al fine di misurare la performance innovativa dei sistemi locali e di esplorare il legame tra tale performance e un insieme di caratteristiche strutturali e knowledge-based. Il contributo del lavoro si inserisce nel dibattito sugli spillover di conoscenza e sulle determinanti territoriali dell'innovazione, proponendo una lettura del dualismo territoriale italiano fondata sulla diversa capacità dei sistemi locali di generare e valorizzare conoscenza.

1. Innovazione, crescita economica e dimensione territoriale della conoscenza

CAPITOLO 1

L'innovazione occupa una posizione centrale nell'analisi economica contemporanea, in quanto rappresenta uno dei principali fattori capaci di spiegare le differenze nei livelli di produttività, competitività e sviluppo tra imprese, settori e territori. Essa non coincide semplicemente con l'introduzione di una nuova tecnologia, ma rinvia a un processo più ampio di produzione, diffusione e valorizzazione economica della conoscenza. Per questa ragione, lo studio dell'innovazione richiede un inquadramento teorico che tenga insieme più dimensioni: la sua evoluzione concettuale nel pensiero economico, il suo ruolo nella crescita e nella performance dei sistemi produttivi, e infine la sua natura territoriale, legata ai meccanismi di esternalità e spillover di conoscenza.

Il presente capitolo intende fornire tale quadro interpretativo. Si ricostruisce l'evoluzione del concetto di innovazione e si analizza il rapporto tra innovazione e performance economica, chiarendo perché l'innovazione costituisca una determinante fondamentale della produttività e della crescita di lungo periodo. Infine, l'attenzione si concentra sulla dimensione territoriale dell'innovazione, con particolare riferimento alle esternalità di conoscenza, agli spillover e ai meccanismi attraverso cui la prossimità geografica, la specializzazione produttiva e la diversificazione economica influenzano la capacità innovativa dei sistemi locali.

1.1 L'evoluzione del concetto di innovazione

L'importanza dell'innovazione per la crescita economica è oggi ampiamente riconosciuta, ma il significato attribuito a questo concetto non è stato sempre lo stesso. Soltanto nel corso del Novecento l'innovazione ha assunto una collocazione pienamente centrale nell'analisi economica, fino a essere considerata una variabile strutturale dello sviluppo e non più un semplice elemento esterno o residuale. La nozione contemporanea di innovazione è dunque il risultato di una lunga evoluzione teorica, che prende avvio già nel pensiero classico.

Tra i primi autori a cogliere, seppure indirettamente, il nesso tra progresso tecnico e sviluppo economico vi è Adam Smith. Nella sua riflessione, l'attenzione non si concentra sull'innovazione nel significato moderno del termine, bensì sui meccanismi che rendono possibile l'aumento della produttività all'interno del sistema economico. In questa prospettiva, un ruolo centrale è attribuito alla divisione del lavoro, intesa come principio fondamentale dell'efficienza produttiva. Smith osserva infatti che la specializzazione delle mansioni, resa possibile dall'ampliamento del mercato, accresce la produttività del lavoro e crea le condizioni favorevoli all'introduzione di miglioramenti tecnici e organizzativi. Il progresso tecnico risulta quindi strettamente connesso alla divisione del lavoro, all'apprendimento attraverso l'esperienza e alla crescente articolazione del sistema produttivo. Pur non essendo ancora concepita come un processo autonomo, l'innovazione appare già legata alla trasformazione della struttura economica e all'espansione del mercato.

Una prospettiva in parte diversa si ritrova in Marx, il quale attribuisce al cambiamento tecnologico un ruolo molto più esplicito nelle dinamiche del capitalismo. Nella sua analisi, l'introduzione delle macchine non rappresenta soltanto un miglioramento tecnico, ma una trasformazione profonda dei rapporti di produzione, dell'organizzazione del lavoro e dei rapporti sociali. L'innovazione viene letta

come un processo inserito nella logica competitiva del capitalismo, in cui la ricerca di produttività e profitto spinge le imprese a introdurre nuove tecnologie e nuove modalità organizzative. In questa prospettiva, il progresso tecnico non è né neutrale né puramente individuale, esso è il prodotto di relazioni economiche e sociali e produce a sua volta effetti strutturali sull'intero sistema produttivo.

Fondamentale è il contributo di Usher che per primo interpreta l'innovazione come un processo cumulativo. Nel suo approccio, essa non nasce come evento isolato, ma come risultato di una sequenza di percezione dei problemi, elaborazione di soluzioni, sperimentazione e perfezionamento. Questa visione processuale dell'innovazione rappresenta un punto di svolta, perché consente di superare l'idea di invenzione come atto puntuale e di leggere invece il cambiamento tecnico come una dinamica graduale e cumulativa.

È con Joseph Schumpeter che l'innovazione assume per la prima volta una collocazione pienamente centrale nell'analisi economica. Generalmente considerato il fondatore dell'economia dell'innovazione, Schumpeter attribuisce infatti all'innovazione un ruolo decisivo nella dinamica capitalistica, definendola come il motore del cambiamento economico. La sua impostazione è nota per il concetto di "distruzione creatrice", secondo cui lo sviluppo non procede in modo lineare, ma attraverso la continua sostituzione di tecnologie, prodotti e assetti organizzativi esistenti con nuovi assetti più efficienti. In questa prospettiva, l'innovazione non si limita a migliorare l'esistente, ma riconfigura i mercati, altera gli equilibri competitivi e apre nuove traiettorie di sviluppo. L'imprenditore innovatore assume quindi una funzione fondamentale, in quanto soggetto capace di introdurre nuove combinazioni produttive e di trasformare l'invenzione in applicazione economica.

E' proprio con Schumpeter che si consolida anche la distinzione tra invenzione e innovazione. L'invenzione rinvia alla produzione di una nuova idea, di una nuova conoscenza o di una nuova soluzione tecnica. L'innovazione, invece, consiste nella sua effettiva introduzione nel processo produttivo o nel mercato. In altre parole, non ogni invenzione diventa innovazione, perché ciò accada è necessario che la novità venga incorporata nell'attività economica e si traduca in un cambiamento effettivo di prodotti, processi o organizzazione. Questa distinzione è particolarmente importante, perché chiarisce che l'innovazione non coincide semplicemente con la creatività scientifica o tecnica, ma con la sua valorizzazione economica.

Nel secondo dopoguerra, e soprattutto con la teoria della crescita endogena, il ruolo dell'innovazione viene ulteriormente rafforzato. Se nella tradizione neoclassica il progresso tecnologico tendeva a essere trattato come fattore esogeno, con Romer e Lucas esso diventa il risultato di investimenti intenzionali in ricerca, capitale umano e produzione di conoscenza. In questa impostazione, la conoscenza è un bene peculiare, caratterizzato da non rivalità e da escludibilità imperfetta: il suo utilizzo da parte di un soggetto non ne riduce la disponibilità per altri, e i suoi benefici tendono a diffondersi oltre i confini del soggetto che l'ha prodotta. Proprio per questa ragione, l'innovazione è considerata una fonte di rendimenti crescenti e di crescita cumulativa.

Alla luce di tale evoluzione teorica, l'innovazione può essere definita come un processo economico e sociale di produzione, applicazione e diffusione di nuova conoscenza. Essa non coincide con il solo progresso tecnico in senso stretto, ma comprende l'insieme dei meccanismi attraverso cui idee, competenze e soluzioni organizzative si trasformano in cambiamento economico.

1.2 Innovazione e performance economica

L'innovazione è comunemente considerata una delle principali determinanti della performance economica, poiché incide sulla produttività dei fattori, sulla competitività delle imprese e sulla capacità di un sistema economico di sostenere la crescita nel lungo periodo. In termini generali, innovare

significa produrre di più o meglio a parità di risorse impiegate, oppure generare nuovi beni, nuovi processi e nuove modalità organizzative che ampliano le possibilità di sviluppo di un'economia. Per questa ragione, il legame tra innovazione e crescita rappresenta uno dei temi centrali della letteratura economica contemporanea.

Il primo canale attraverso cui l'innovazione influenza la performance economica è l'aumento della produttività. L'introduzione di nuove tecnologie, di processi produttivi più efficienti o di nuove modalità organizzative consente infatti di ridurre i costi, migliorare la qualità dell'output e aumentare l'efficienza nell'impiego degli input. In questo senso, la crescita economica non dipende soltanto dall'accumulazione di capitale e lavoro, ma dalla capacità di combinare tali fattori in modo più efficace grazie alla conoscenza. L'innovazione, dunque, non rappresenta soltanto una componente del cambiamento economico, ma uno dei suoi principali motori.

Un secondo aspetto riguarda il fatto che l'innovazione non produce solo incrementi di efficienza, ma anche trasformazioni strutturali. L'emergere di nuove attività, nuovi settori e nuove opportunità imprenditoriali modifica la composizione del sistema produttivo, spostando risorse verso impieghi a maggiore contenuto tecnologico e a più elevato valore aggiunto. Da questo punto di vista, innovazione e crescita non sono legate solo in termini quantitativi, ma anche qualitativi: l'innovazione contribuisce a ridefinire la struttura dell'economia, la specializzazione produttiva e la posizione competitiva di un Paese nei mercati internazionali.

Per analizzare empiricamente questo rapporto, la letteratura distingue generalmente tra input e output del processo innovativo. Tra gli input rientrano, ad esempio, gli investimenti in ricerca e sviluppo, il capitale umano qualificato, la presenza di università, laboratori, infrastrutture scientifiche e reti di trasferimento tecnologico. Tra gli output figurano invece i brevetti, le nuove imprese innovative, i nuovi prodotti e i nuovi processi. Questa distinzione è essenziale, poiché consente di chiarire che la disponibilità di risorse innovative non si traduce automaticamente in risultati innovativi. La relazione tra input e output non è mai meccanica: territori o paesi dotati di risorse simili possono conseguire performance molto differenti, a seconda della qualità delle istituzioni, della struttura produttiva, della capacità di assorbimento tecnologico e dell'efficacia dei meccanismi di diffusione della conoscenza.

La letteratura empirica conferma ampiamente l'esistenza di una relazione positiva tra investimenti in ricerca e sviluppo e performance economica. I paesi caratterizzati da una maggiore intensità di R&S tendono in genere a mostrare livelli più elevati di produttività, competitività e capacità innovativa. Al tempo stesso, diversi studi hanno evidenziato che la composizione degli investimenti è rilevante: gli investimenti privati in ricerca e sviluppo appaiono spesso associati in modo più diretto agli output innovativi e ai risultati economici, pur restando fondamentale il ruolo del settore pubblico nella formazione del capitale umano, nel sostegno alla ricerca di base e nella costruzione delle infrastrutture della conoscenza. Ne deriva che la performance innovativa di un sistema economico dipende non soltanto dal volume complessivo degli investimenti, ma anche dalla loro struttura e dalla qualità dell'ecosistema in cui essi si collocano.

In questo quadro, l'innovazione può essere considerata una misura indiretta della capacità di crescita di un sistema economico. Essa segnala, infatti, la presenza di processi di apprendimento, accumulazione di conoscenza e adattamento competitivo. Un sistema che innova è un sistema che non si limita a riprodurre tecniche esistenti, ma è in grado di modificare continuamente la propria base produttiva, di generare nuove opportunità e di reagire in modo dinamico ai cambiamenti del contesto economico. Per tale ragione, gli indicatori dell'innovazione sono spesso utilizzati come variabili chiave nell'analisi della performance economica e delle differenze di sviluppo tra territori.

Tuttavia, la relazione tra innovazione e crescita non può essere compresa pienamente se non si considera il carattere diffusivo della conoscenza. Una parte rilevante degli effetti economici

dell'innovazione, infatti, non si esaurisce nei benefici privati ottenuti dall'impresa o dall'istituzione che l'ha prodotta, ma si trasmette anche ad altri soggetti. È proprio questo elemento a introdurre il tema delle esternalità e degli spillover di conoscenza, che costituiscono il ponte teorico tra economia dell'innovazione ed economia territoriale.

1.3 Esternalità, spillover e dimensione territoriale dell'innovazione

Nel linguaggio economico, le esternalità sono gli effetti che l'attività di un agente economico esercita su altri soggetti senza che tali effetti siano pienamente riflessi nei prezzi di mercato. Nel caso dell'innovazione, le esternalità assumono prevalentemente una forma positiva: la ricerca svolta da un'impresa, l'attività scientifica di un'università, l'esperienza accumulata da lavoratori altamente qualificati o l'adozione di nuove tecnologie possono generare benefici che si diffondono anche ad altri attori economici. In altri termini, la conoscenza prodotta da un soggetto tende a eccedere i confini della sua appropriabilità privata e a produrre vantaggi diffusi.

Una forma specifica di queste esternalità è costituita dagli spillover di conoscenza. Con tale espressione si indicano i processi attraverso cui idee, informazioni tecniche, competenze, pratiche produttive e capacità organizzative si trasferiscono, intenzionalmente o meno, da un'impresa, da un settore o da un territorio ad altri soggetti. Gli spillover possono operare attraverso diversi canali: la mobilità del lavoro qualificato, i rapporti di fornitura, l'imitazione, le collaborazioni tra imprese e università, le reti professionali, i contatti informali e, più in generale, tutti i processi di apprendimento collettivo che caratterizzano i sistemi produttivi. Grazie agli spillover, chi innova non produce soltanto un vantaggio privato, ma contribuisce anche ad ampliare la base di conoscenza disponibile per altri attori, riducendo i costi della ricerca successiva e favorendo nuove innovazioni.

Questo fenomeno è particolarmente importante perché rende l'innovazione un processo cumulativo. Le nuove conoscenze non sorgono nel vuoto, ma si sviluppano a partire da conoscenze preesistenti, che fungono da base per ulteriori miglioramenti, adattamenti e ricombinazioni. Da qui deriva un meccanismo circolare: l'innovazione genera nuova conoscenza, la nuova conoscenza si diffonde almeno in parte attraverso spillover, e tale diffusione crea condizioni favorevoli per ulteriori attività innovative. Per questo motivo, la letteratura economica riconosce negli spillover uno dei principali canali attraverso cui l'innovazione alimenta la crescita economica.

La diffusione della conoscenza, tuttavia, non è uniforme né priva di attriti. Una parte della conoscenza può essere codificata e trasferita relativamente facilmente, un'altra parte è di natura tacita, cioè incorporata nelle competenze individuali, nelle routine organizzative, nelle relazioni di fiducia e nei contesti produttivi locali. Proprio per questa ragione, gli spillover tendono a essere tanto più intensi quanto maggiore è la prossimità tra i soggetti coinvolti. La vicinanza geografica facilita infatti le interazioni faccia a faccia, l'osservazione reciproca, la mobilità del lavoro e l'apprendimento informale. In questo senso, la produzione e la diffusione della conoscenza dipendono in misura significativa dal contesto territoriale.

Su questa base si sviluppa il tema della geografia dell'innovazione. Se gli spillover sono localizzati, allora i territori non sono semplici contenitori neutri dell'attività economica, ma ambienti che possono favorire o ostacolare la capacità innovativa. La concentrazione spaziale di imprese, competenze, infrastrutture e istituzioni di ricerca può infatti generare economie di agglomerazione, cioè vantaggi derivanti dalla prossimità tra attività economiche. In questo ambito, la letteratura distingue tradizionalmente tra esternalità di specializzazione ed esternalità di diversità.

Le prime, riconducibili alla tradizione di Marshall, Arrow e Romer, derivano dalla concentrazione di imprese appartenenti allo stesso settore. In contesti di elevata specializzazione produttiva, le imprese

possono beneficiare di un mercato del lavoro più specializzato, di fornitori dedicati, di conoscenze tecniche settoriali e di una più intensa circolazione di informazioni all'interno dello stesso ambito produttivo. Le seconde, invece, rinviano al contributo di Jacobs e sottolineano il ruolo della varietà economica. In territori caratterizzati dalla compresenza di settori differenti, la diversità delle competenze e delle basi di conoscenza aumenta le occasioni di contaminazione, ricombinazione e apprendimento trasversale, favorendo soprattutto l'emergere di innovazioni nuove o più complesse.

La contrapposizione tra specializzazione e diversificazione non va però letta in termini assoluti. La letteratura empirica mostra infatti che entrambi i meccanismi possono essere rilevanti, a seconda del settore considerato, del livello tecnologico delle attività osservate e delle caratteristiche del contesto locale. In alcuni casi, l'innovazione si sviluppa più facilmente in ambienti fortemente specializzati, nei quali prevalgono economie di localizzazione; in altri, essa trae maggiore impulso da contesti urbani e metropolitani più diversificati, in cui la varietà produttiva favorisce la fertilizzazione incrociata tra conoscenze differenti. La dimensione territoriale dell'innovazione, dunque, non dipende da un unico principio, ma dal modo in cui diversi meccanismi di agglomerazione interagiscono tra loro.

Questo quadro teorico risulta particolarmente rilevante nel caso italiano, caratterizzato da persistenti squilibri territoriali nella struttura produttiva, nella dotazione di risorse innovative e nella capacità di trasformare la conoscenza in risultati economici. Le differenze tra aree del Paese non si manifestano soltanto nei livelli di reddito o occupazione, ma anche nella diversa possibilità di attivare esternalità, generare spillover e costruire ecosistemi innovativi dinamici. Per tale ragione, lo studio dell'innovazione richiede un approccio che tenga conto simultaneamente della dimensione economica e di quella territoriale.

Alla luce di queste considerazioni, l'innovazione può essere interpretata come un fenomeno endogeno, cumulativo e territorialmente radicato. Essa dipende dalla produzione di conoscenza, ma anche dalla capacità di assorbirla, diffonderla e valorizzarla all'interno di sistemi locali caratterizzati da differenti combinazioni di specializzazione, diversità, densità relazionale e prossimità geografica. È proprio su questo terreno teorico che si colloca l'analisi sviluppata nella quale l'attenzione si concentrerà sugli output innovativi e sulle determinanti territoriali della loro distribuzione.

2. Rassegna della letteratura

CAPITOLO 2

Il presente capitolo inquadra il lavoro all'interno della letteratura sulla geografia dell'innovazione, ricostruendo i principali contributi teorici ed empirici relativi alla distribuzione spaziale dell'attività innovativa, agli spillover di conoscenza e al ruolo delle esternalità locali. A partire da questa rassegna, si mette in evidenza il vuoto analitico che la tesi intende colmare, con particolare riferimento al caso del Mezzogiorno italiano e chiarisce il contributo specifico del lavoro.

2.1 La letteratura sulla geografia dell'innovazione

La letteratura sulla geografia dell'innovazione parte da una constatazione ormai classica: l'attività innovativa non è distribuita uniformemente nello spazio, ma tende a concentrarsi in luoghi specifici, dove si addensano imprese, competenze, funzioni di ricerca e occasioni di interazione. In questo quadro, la dimensione territoriale non rappresenta un semplice sfondo dell'innovazione, ma una sua determinante essenziale. Audretsch e Feldman (1996) sintetizzano con chiarezza questo punto quando osservano che esiste una considerevole evidenza del fatto che gli spillover di R&S esistano e siano geograficamente delimitati. L'intensità di tali esternalità e i costi della loro trasmissione non sono uguali tra industrie e ciò contribuisce a spiegare perché l'innovazione tenda a concentrarsi in alcuni settori e in alcuni luoghi più che in altri.

Nel caso italiano, il contributo di riferimento è quello prodotto da Paci e Usai (2000) che studiano la distribuzione spaziale dell'innovazione utilizzando i Sistemi Locali del Lavoro come unità territoriale d'analisi. Il loro lavoro mostra che la geografia dell'innovazione in Italia è fortemente polarizzata e che l'attività brevettuale presenta chiari segni di correlazione spaziale, gli autori evidenziano il carattere dualistico del sistema innovativo italiano: l'innovazione si concentra prevalentemente nel Centro-Nord, mentre il Mezzogiorno resta ampiamente marginale. Nel loro database, oltre l'80% dell'attività brevettuale è localizzata nel Nord, mentre il Sud pesa solo in misura molto ridotta. La stessa mappa dell'innovazione restituisce un'Italia in cui i cluster più forti si addensano attorno ai principali poli settentrionali, mentre nel Mezzogiorno emergono soltanto alcuni nuclei più deboli e isolati.

Da questo primo risultato discende una conseguenza teorica rilevante: il divario territoriale non può essere letto semplicemente come differenza quantitativa nel numero di innovazioni, ma come differenza nella struttura territoriale dei processi innovativi. Paci e Usai insistono infatti sul fatto che vi siano chiari segni di correlazione spaziale nella distribuzione dell'innovazione e che gli spillover tecnologici abbiano una natura localizzata. Il loro modello mostra che l'attività innovativa di un sistema locale è positivamente influenzata dal livello di innovatività dei sistemi contigui, ma che tale influenza tende a esaurirsi con la distanza. Non è sufficiente che esistano alcuni centri innovativi nel Mezzogiorno, se questi non sono immersi in un contesto capace di generare diffusione cumulativa della conoscenza.

Su questo aspetto si collega la letteratura sugli spillover di conoscenza, che ha mostrato come la trasmissione della conoscenza, soprattutto quella tacita, sia ancora fortemente dipendente dalla prossimità. Paci e Usai richiamano esplicitamente l'idea che sia difficile e costoso trasmettere la conoscenza e che proprio per questo la prossimità favorisca processi locali di condivisione delle informazioni e di apprendimento collettivo.

Un ulteriore sviluppo della stessa linea interpretativa è fornito da Figueiredo, Guimarães e Woodward (2015), che mostrano che gli spillover di conoscenza sono influenzati da due forze simultanee: la

localizzazione industriale e il distance decay. I loro risultati indicano che gli spillover sono positivamente associati alla localizzazione delle industrie e che l'agglomerazione settoriale può persino compensare, almeno in parte, l'effetto negativo della distanza. Questa evidenza è particolarmente utile per leggere il caso del Mezzogiorno: se la base industriale locale è poco densa e poco concentrata nei settori più innovativi, il problema non è soltanto l'assenza di innovazione in senso stretto, ma anche la debolezza del contesto che dovrebbe amplificare e trattenere gli spillover.

Il dibattito classico sulle esternalità locali si articola poi attorno alla contrapposizione tra esternalità di Marshall e Jacobs. Nella formulazione di Marshall, la concentrazione geografica di imprese appartenenti allo stesso settore favorisce l'innovazione grazie a mercati del lavoro specializzati, fornitori dedicati, imitazione, circolazione di competenze e riduzione dei costi di coordinamento. Nella prospettiva di Jacobs, al contrario, la fonte dell'innovazione è la varietà urbana e settoriale, perché la ricombinazione tra conoscenze diverse genera nuove idee. Paci e Usai (2000) mostrano che, nel caso italiano, le due forme di esternalità non si escludono a vicenda. La loro analisi econometrica conclude che specializzazione e diversificazione sono entrambe rilevanti e che l'attività innovativa di un settore locale è positivamente influenzata sia dalla specializzazione produttiva del settore stesso sia dal grado di diversità del sistema locale. Gli autori osservano che le esternalità jacobsoniane risultano più forti nei settori high-tech e nelle aree metropolitane.

Beaudry e Schiffauerova (2009), in una rassegna molto ampia del dibattito Marshall versus Jacobs, mostrano che la mancanza di risultati univoci dipende in larga misura da questioni di misurazione e di metodo. In particolare, il livello di aggregazione industriale e geografica, così come la scelta degli indicatori di performance, di specializzazione e di diversità, sono tra le cause principali dell'irrisolutezza del dibattito. Ad esempio, la classificazione industriale a 3 digit rappresenta spesso un livello in cui gli effetti MAR e Jacobs diventano difficilmente distinguibili. E' per questo che nel seguente lavoro di tesi è stata adottata una classificazione industriale a 2 digit.

Carlino, Chatterjee e Hunt (2007) mostrano che la densità occupazionale delle aree metropolitane è positivamente associata all'intensità brevettuale: a parità di altre condizioni, una città con densità occupazionale doppia presenta un'intensità brevettuale circa del 20% più elevata. Gli autori trovano anche l'esistenza di rendimenti decrescenti oltre una certa soglia di densità. Questo risultato è molto rilevante per interpretare il dualismo italiano: se la densità urbana e occupazionale facilita matching, condivisione di input e spillover, allora parte del divario innovativo del Mezzogiorno potrebbe dipendere non solo da una minore quantità di input innovativi, ma anche da una minore capacità del territorio di trasformare tali input in output innovativi. È qui che il tema della produttività dell'innovazione diventa teoricamente fondato, non basta chiedersi dove ci sia più innovazione, bisogna chiedersi dove il contesto renda più efficiente il processo innovativo.

Su questo stesso piano si colloca il contributo critico di Boschma (2005), che ridimensiona l'idea, troppo semplice, secondo cui la prossimità geografica basterebbe da sola a spiegare l'innovazione. Boschma sostiene che la prossimità geografica non è né condizione necessaria né condizione sufficiente per l'apprendimento interattivo e per l'innovazione. Essa può al più svolgere un ruolo complementare, ma funziona davvero solo se accompagnata da altre forme di prossimità: cognitiva, organizzativa, sociale e istituzionale. Troppa prossimità, soprattutto in contesti altamente specializzati e chiusi, può irrigidire le traiettorie di apprendimento e ostacolare l'innovazione radicale (lock-in).

Castellano, Musella e Punzo (2023) propongono un quadro integrato in cui l'innovazione territoriale dipende congiuntamente da esternalità di agglomerazione, complementarità delle competenze e mobilità dei lavoratori. Il loro risultato principale è che, nel caso italiano, la specializzazione gioca un ruolo più decisivo della diversificazione nel favorire l'innovazione. L'interazione tra skills complementari e specializzazione incide in modo forte sulle attività innovative. I risultati mostrano che il Mezzogiorno incontra difficoltà nel trattenere lavoratori con competenze più strettamente connesse

ai settori della conoscenza, mentre una parte significativa della forza lavoro qualificata si sposta verso il Nord. Gli autori osservano esplicitamente che nel Sud la forza lavoro è relativamente più composta da competenze “unlinked”, meno capaci di alimentare nuova conoscenza, e che oltre due milioni di persone si sono spostate dal Sud al Nord tra il 2002 e il 2017, con una quota molto rilevante di giovani e laureati.

2.2 La domanda di ricerca

Alla luce di questa letteratura, il vuoto teorico ed empirico può essere formulato in modo più preciso. La letteratura ha chiarito che l'innovazione è spazialmente concentrata, che gli spillover sono localizzati, che specializzazione e diversità possono co-esistere e che le traiettorie innovative seguono logiche cumulative e path-dependent. Meno chiaro è invece come questi meccanismi operino nei territori strutturalmente deboli e, in particolare, se il ritardo innovativo del Mezzogiorno rifletta soltanto una minore dotazione di input o anche una minore capacità sistemica di trasformarli in innovazione.

Paci e Usai mostrano il dualismo, Boschma spiega perché la sola prossimità non basta, Balland suggerisce che lo sviluppo segue traiettorie legate alle capacità esistenti e Castellano, Musella e Punzo mostrano che nel caso italiano contano insieme struttura produttiva, competenze e mobilità. Quello che ancora manca è un'analisi che metta questi livelli insieme nel caso dei Sistemi Locali del Lavoro, verificando se le radici del divario siano da ricondurre a una minore produttività territoriale dell'innovazione o, più precisamente, a una combinazione di scarsa massa critica, debolezza delle esternalità locali, fragilità della struttura cognitiva e perdita di lavoro qualificato verso le aree forti del Paese.

In questa prospettiva, il presente lavoro non si limita a verificare se il Centro – Sud innovi meno del Nord, dato già ampiamente documentato dalla letteratura, ma intende indagare se il divario osservato rifletta una minore produttività territoriale dell'innovazione oppure una configurazione strutturale più complessa.

3. Analisi di correlazione spaziale

CAPITOLO 3

L'innovazione è un fenomeno complesso e multidimensionale. Al fine di cogliere dimensioni differenti ma complementari della capacità territoriale di generazione e valorizzazione di innovazione, il presente lavoro adotta una strategia di misurazione plurima, costruendo tre distinti indicatori di produttività dell'innovazione, ciascuno basato rispettivamente su:

- (i) brevetti UIBM
- (ii) brevetti EPO
- (iii) numero di start-up innovative

Tali grandezze non sono considerate in valore assoluto, bensì normalizzate rispetto al numero di addetti, per ottenere misure comparabili nello spazio e depurate dalla mera dimensione economica dei territori.

I brevetti depositati presso l'Ufficio Italiano Brevetti e Marchi (UIBM) rappresentano la forma più ampia e diffusa di tutela della proprietà industriale a livello nazionale. Essi intercettano una gamma eterogenea di invenzioni, frequentemente associate a innovazioni incrementali, miglioramenti di processo o adattamenti tecnologici contestuali, tipici del tessuto produttivo delle piccole e medie imprese. Rispetto ai brevetti europei, presentano costi inferiori e una minore selettività, risultando pertanto più diffusi territorialmente. Per questa ragione, costituiscono una misura idonea a catturare la capacità inventiva diffusa del sistema produttivo locale, specialmente nei contesti caratterizzati da specializzazione manifatturiera tradizionale o da sistemi di impresa di piccola dimensione.

Diversamente, i brevetti depositati presso l'European Patent Office (EPO) intercettano invenzioni ad alto contenuto tecnologico, spesso associate a imprese strutturate e a strategie competitive orientate a mercati internazionali. Il deposito presso l'EPO comporta costi più elevati e una maggiore complessità procedurale, determinando un processo di selezione più stringente. Tali brevetti risultano quindi meno numerosi, più concentrati e generalmente associati a innovazione di frontiera. Proprio per queste caratteristiche, essi rappresentano una misura più selettiva e qualificata dell'attività inventiva più avanzata, segnalando la presenza di ecosistemi innovativi maturi e competitivi su scala sovranazionale. In termini spaziali, la loro distribuzione tende a evidenziare poli tecnologici di eccellenza piuttosto che diffusione capillare.

L'inclusione delle start-up innovative introduce una dimensione ulteriore dell'analisi. Una quota crescente dell'innovazione contemporanea (in particolare nei servizi, nel software, nelle piattaforme digitali) non assume necessariamente la forma di un brevetto. Le start-up rappresentano dunque un output alternativo dell'attività inventiva, in cui la conoscenza viene immediatamente tradotta in iniziativa imprenditoriale. Esse consentono di cogliere la dimensione dinamica e imprenditoriale dell'innovazione territoriale, intercettando la capacità di sfruttamento economico della conoscenza oltre la sua mera formalizzazione giuridica.

L'adozione congiunta dei tre indicatori consente pertanto di distinguere tra:

- (1) innovazione diffusa e incrementale (brevetti UIBM)
- (2) innovazione selettiva e tecnologicamente avanzata (brevetti EPO)
- (3) innovazione imprenditoriale (start-up innovative)

Questa articolazione risulta particolarmente rilevante in un'analisi spaziale, poiché permette di verificare se i territori presentino specializzazione in una sola dimensione dell'innovazione oppure

configurazioni coerenti lungo tutte le forme di output innovativo, restituendo così una rappresentazione più completa della geografia dell'innovazione.

3.1 Unità territoriale di riferimento

L'unità territoriale di riferimento è rappresentata dai Sistemi Locali del Lavoro (SLL) (Sforzi, 1997) nella configurazione definita da ISTAT con riferimento al Censimento 2021, per un totale di 515 aree (figura 1).

I SLL costituiscono aree funzionali delineate sulla base dei flussi di pendolarismo casa-lavoro e caratterizzate da un elevato grado di auto contenimento occupazionale. Dal punto di vista economico, essi approssimano mercati del lavoro locali integrati e rappresentano una scala territoriale coerente con l'ipotesi di localizzazione delle esternalità tecnologiche. La scelta di tale unità di analisi risponde all'esigenza di adottare una delimitazione spaziale fondata su interazioni effettive tra imprese e lavoratori, piuttosto che su confini amministrativi esogeni. In presenza di spillover di conoscenza mediati dalla prossimità geografica e dalla mobilità del lavoro qualificato, i SLL rappresentano un ambito empiricamente appropriato per l'analisi della distribuzione spaziale dell'innovazione.

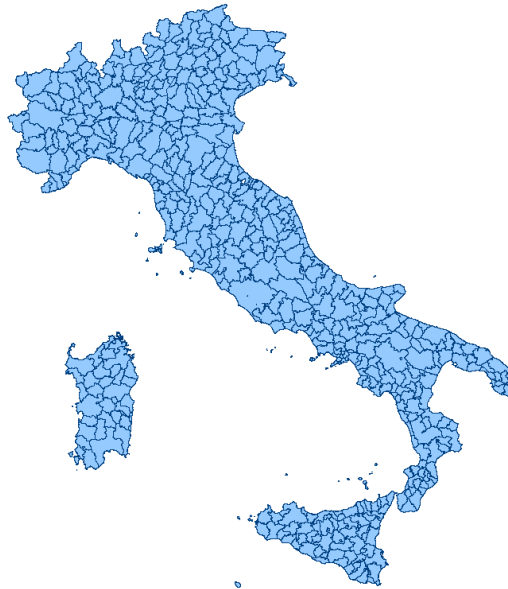


Figure 1: mappa dei SLL secondo l'aggiornamento del 2021

3.2 Orizzonte temporale

L'orizzonte temporale dell'analisi differisce in funzione della fonte informativa.

Per i brevetti EPO e per le start-up innovative la finestra di osservazione copre il periodo 2016 - 2020, mentre per i brevetti UIBM si estende al periodo 2016 - 2019. La scelta di adottare finestre pluriennali consente di attenuare la variabilità idiosincratICA annuale e di trattare l'output innovativo come processo cumulativo, riducendo il rumore statistico particolarmente rilevante nei territori di piccola dimensione.

3.3 Dati occupazionali e armonizzazione territoriale

I dati relativi al numero di addetti provengono dal Censimento ISTAT del 2011, sono disponibili a livello comunale e classificati per settore sulla base dei codici Ateco 2007. Poiché tra il 2011 e il 2021 si sono verificate modifiche nei confini amministrativi e nella delimitazione dei SLL, si è reso necessario un processo di armonizzazione territoriale.

In particolare, gli addetti originariamente osservati a livello comunale sono stati riallocati ai SLL 2021 sulla base della corrispondenza territoriale tra comuni e nuova delimitazione funzionale. Tale procedura garantisce coerenza spaziale tra la base occupazionale utilizzata per la normalizzazione degli indicatori e l'unità territoriale adottata nell'analisi.

Il numero di addetti costituisce la variabile di scala utilizzata per la costruzione degli indicatori di intensità innovativa, al fine di neutralizzare effetti puramente dimensionali e rendere comparabili territori caratterizzati da diversa base economica.

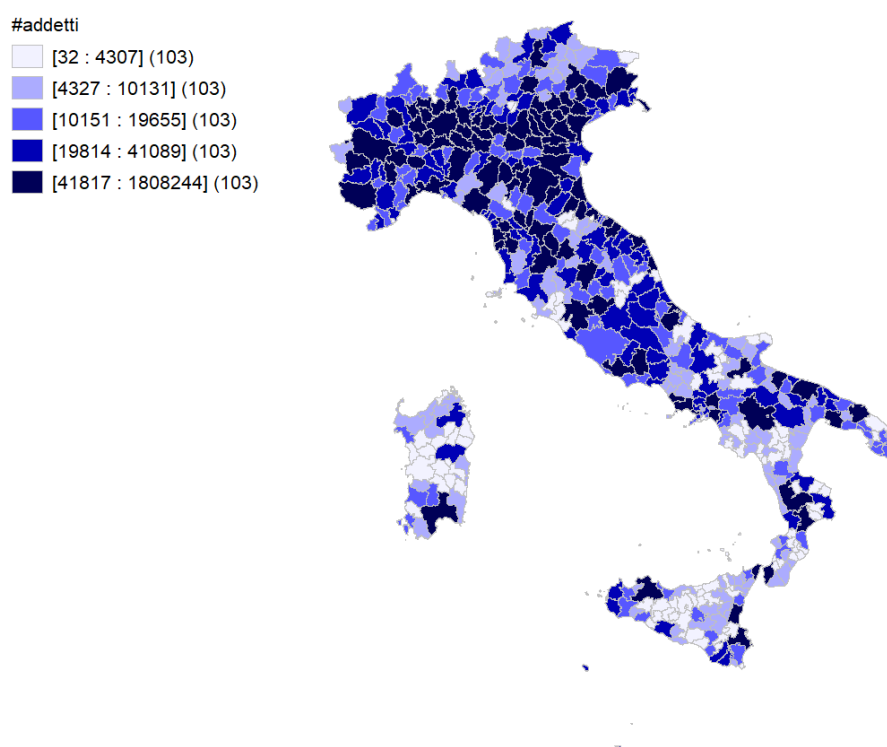


Figure 2: mappa coropletica (quantili) degli addetti per SLL secondo il censimento del 2011

La figura 2 mostra la distribuzione del numero di addetti nei 515 SLL italiani, classificata in cinque quantili. La rappresentazione evidenzia una forte concentrazione dell'occupazione nei principali sistemi locali del Nord, in particolare nelle aree metropolitane e nei poli produttivi maggiori. Nel Mezzogiorno prevalgono invece SLL collocati nei quantili inferiori della distribuzione.

La struttura occupazionale risulta dunque fortemente polarizzata, con pochi sistemi locali che concentrano una quota rilevante della base economica nazionale. Tale evidenza giustifica l'utilizzo di indicatori normalizzati rispetto al numero di addetti nelle successive analisi sull'innovazione.

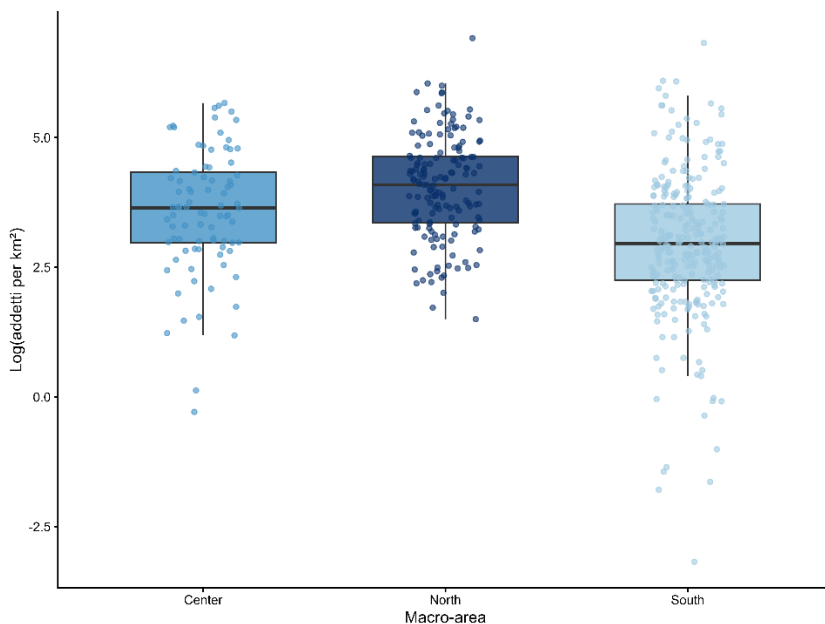


Figure 3: boxplot della densità di addetti per km², il confronto Nord - Centro - Sud

Il boxplot (figura 3) confronta la densità occupazionale nei Sistemi Locali del Lavoro (SLL) italiani, misurata come numero di addetti per km² e rappresentata in scala logaritmica, tra le tre macro-aree territoriali: Nord, Centro e Sud.

Il Nord presenta la mediana più elevata e una distribuzione complessivamente spostata verso valori superiori, evidenziando una maggiore concentrazione dell'attività economica e una struttura territoriale più densamente integrata. Il Centro si colloca in posizione intermedia, con livelli di densità inferiori rispetto al Nord ma superiori rispetto al Sud. Il Mezzogiorno mostra invece la mediana più bassa e una dispersione più ampia, con la presenza di numerosi sistemi locali a bassa densità e pochi poli relativamente concentrati.

L'evidenza suggerisce l'esistenza di un gradiente territoriale nella struttura occupazionale, coerente con il divario economico tra le macro-aree del Paese e rilevante ai fini dell'interpretazione delle successive analisi sulla distribuzione spaziale dell'innovazione.

I cluster sono stati costruiti secondo quanto indicato in tabella 1:

Table 1: costruzione dei cluster regionali per l'analisi boxplot della densità di addetti per km²

Cluster	Regioni
Nord	Piemonte, Valle d'Aosta, Liguria, Lombardia, Veneto, Friuli – Venezia Giulia, Trentino-Alto Adige, Emilia-Romagna
Centro	Toscana, Umbria, Marche, Lazio
Sud e isole	Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria, Sicilia, Sardegna

L'evidenza conferma l'esistenza di un divario territoriale nella struttura occupazionale, che costituisce un elemento rilevante nell'interpretazione delle successive analisi sulla distribuzione spaziale dell'innovazione.

3.4 Criteri di attribuzione territoriale delle attività innovative

Un elemento metodologicamente cruciale riguarda il criterio di georeferenziazione delle osservazioni relative all'output innovativo.

Per i brevetti depositati presso l'European Patent Office (EPO), l'attribuzione territoriale è stata effettuata utilizzando l'indirizzo dell'inventore (Breschi, 1997). Tale scelta riflette l'esigenza di identificare il luogo effettivo di generazione della conoscenza, piuttosto che la localizzazione formale della proprietà giuridica del brevetto. In presenza di imprese multi-stabilimento o multinazionali, la sede legale tende infatti a coincidere con l'headquarter societario, mentre l'attività di ricerca e sviluppo può essere svolta in unità produttive decentrate. L'indirizzo dell'inventore costituisce quindi una proxy più accurata del luogo di produzione dell'attività inventiva e consente una migliore approssimazione della dimensione territoriale degli spillover tecnologici.

Per le start-up innovative, l'assegnazione territoriale è stata effettuata sulla base dell'indirizzo della sede legale. Nel caso di imprese di nuova costituzione, la sede legale tende a coincidere con il luogo in cui si concentrano le funzioni strategiche e operative iniziali, essa rappresenta pertanto una proxy plausibile della localizzazione dell'attività imprenditoriale innovativa.

Per i brevetti UIBM, l'attribuzione territoriale è stata effettuata utilizzando l'indirizzo dell'applicant nel caso di soggetti fisici e la sede legale dell'impresa nel caso di soggetti giuridici. A differenza dei brevetti EPO, la base dati relativa ai brevetti UIBM non consente in modo sistematico di ricostruire l'indirizzo dell'inventore con un grado di completezza e standardizzazione analogo. In assenza di tale informazione, l'indirizzo dell'applicant rappresenta la migliore proxy disponibile per l'ancoraggio territoriale dell'invenzione.

Pur riconoscendo la potenziale divergenza tra sede legale e luogo effettivo di generazione dell'invenzione, tale scelta metodologica risulta giustificata dalla necessità di garantire omogeneità e completezza del dataset. Inoltre, nel caso di brevetti a valenza nazionale, la minore incidenza di strutture multinazionali complesse rispetto ai brevetti EPO riduce la probabilità di disallineamenti sistematici tra luogo di invenzione e sede giuridica.

3.5 Strumenti di analisi

3.5.1. Moran's I

Per verificare se l'innovazione presenti una struttura spaziale non casuale è necessario misurare formalmente la dipendenza spaziale. Lo strumento standard da utilizzare a questo scopo è il Moran's I globale, che quantifica il grado di autocorrelazione spaziale di una variabile (x) osservata su (n) unità territoriali (in questo caso, i Sistemi Locali del Lavoro). Intuitivamente, Moran's I misura se unità territoriali contigue tendono ad assumere valori simili (clustering) oppure dissimili (dispersione), rispetto a quanto ci si aspetterebbe in assenza di struttura spaziale.

Formalmente, dato un sistema di pesi spaziali $W = [w_{ij}]$ che descrive la struttura di vicinato tra SLL, il Moran's I è definito come:

$$I = \frac{n}{S_0} \frac{\sum_i \sum_j w_{ij} (x_i - \bar{x})(x_j - \bar{x})}{\sum_i (x_i - \bar{x})^2} \quad \text{con } S_0 = \sum_i \sum_j w_{ij}$$

dove:

i è il SLL di riferimento per l'analisi

j è un altro SLL, tipicamente un vicino di i

Il numeratore è una misura di covarianza spaziale: aumenta quando valori elevati (o bassi) di x_i si trovano vicino a valori elevati (o bassi), e diminuisce quando valori elevati sono circondati da valori bassi (e viceversa). Sotto l'ipotesi nulla di assenza di autocorrelazione spaziale, il valore atteso è:

$$E[I] = -\frac{1}{n-1}$$

Valori di I significativamente superiori al valore atteso indicano autocorrelazione positiva (clustering), mentre valori significativamente inferiori indicano autocorrelazione negativa (pattern "a scacchiera", cioè alternanza sistematica tra alto e basso).

Un aspetto cruciale è che Moran's I è associato ad un test d'ipotesi:

- H_0 : la variabile è spazialmente indipendente (assenza di autocorrelazione, assegnazione casuale rispetto alla struttura dei vicini);
- H_1 : presenza di autocorrelazione spaziale (positiva o negativa, a seconda dell'alternativa specificata).

L'inferenza può essere condotta sia con un approccio analitico (approssimazione asintotica della distribuzione di I sotto H_0) che in modo più robusto, con un test permutazionale (Monte Carlo) che costruisce la distribuzione empirica di I riassegnando casualmente i valori di x alle unità territoriali (randomizzazione) mantenendo fissa la matrice dei pesi W .

Nelle analisi si seguito riportate, l'uso congiunto di test analitico e permutazionale consente di verificare la solidità dell'evidenza statistica, soprattutto in presenza di distribuzioni asimmetriche o valori estremi. In tal senso Moran's I è rilevante per due motivi:

1. Consente di verificare se gli indicatori di output innovativo (brevetti e start-up per addetti) presentino una componente spaziale sistematica, coerente con l'idea che la produzione e la diffusione della conoscenza siano localmente mediate (spillover, prossimità, interazioni territoriali).
2. Fornisce una diagnosi preliminare essenziale per l'analisi successiva: in presenza di autocorrelazione, l'eterogeneità osservata nelle mappe coropletiche non va interpretata come semplice dispersione di valori, ma come possibile esito di interdipendenza tra territori.

Un ulteriore aspetto metodologico determinante riguarda la scelta della matrice dei pesi (W).

In questa analisi (W) è costruita su contiguità tra poligoni (queen contiguity) e standardizzata per riga (stile "W"), in modo che ogni SLL sia confrontato con la media ponderata dei suoi vicini e che le differenze nel numero di contatti non dominino meccanicamente il risultato. Infine, l'analisi è replicata su più ordini di contiguità (1°, 2°, 3°), ciò consente di valutare se un'eventuale dipendenza spaziale sia

limitata ai confinanti diretti o se si estenda a vicinati più ampi (vicini dei vicini), coerentemente con l'ipotesi che gli spillover possano operare a scale differenti.

3.5.2. LISA

Accanto alla misura globale, è utile impiegare indicatori locali di autocorrelazione che consentono di identificare dove la dipendenza spaziale è presente e con quale configurazione.

Il LISA (Local Indicators of Spatial Association) più utilizzato è il Local Moran's I, definito per ciascuna unità territoriale eventualmente con normalizzazioni standard. Operativamente, I_i misura l'associazione tra lo scostamento di x_i dalla media e lo scostamento medio ponderato dei valori nei territori vicini.

$$I_i = (x_i - \bar{x}) \sum_j w_{ij} (x_j - \bar{x})$$

Questa misura consente di classificare gli SLL in quattro configurazioni canoniche (in termini di quadranti del Moran scatterplot locale):

- High - High (HH): valori alti circondati da valori alti (cluster "caldo");
- Low - Low (LL): valori bassi circondati da valori bassi (cluster "freddo");
- High - Low (HL): valore alto circondato da valori bassi (outlier spaziale);
- Low - High (LH): valore basso circondato da valori alti (outlier spaziale);

A differenza della lettura puramente visuale delle coropletiche, i LISA introducono esplicitamente una componente inferenziale: ciascun I_i è associato a un test locale con relativo p-value, tipicamente ottenuto tramite permutazioni (randomizzazione locale) che valutano quanto l'associazione osservata sia compatibile con un processo casuale dato il vicinato.

Questo passaggio è centrale: la LISA map non segnala solo "aree alte" o "aree basse", ma identifica cluster/outlier statisticamente supportati, distinguendoli da configurazioni compatibili con rumore o con eterogeneità non strutturata.

Dopo la diagnosi globale tramite Moran's I, la LISA map consente di localizzare le eventuali aree in cui l'attività innovativa per addetto risulta concentrata in modo significativo (HH e LL) e di distinguere tali concentrazioni da situazioni di polarizzazione locale (HL e LH). In questo modo, l'analisi locale integra quella globale: anche quando il Moran globale non evidenzia un clustering "di sistema", i LISA permettono di rilevare micro-pattern territoriali, rendendo l'interpretazione spaziale più aderente alla natura spesso discontinua e policentrica dei fenomeni di innovazione.

3.6 Brevetti EPO: analisi di distribuzione spaziale

L'indicatore dei brevetti EPO è stato costruito a partire dal database estratto da PATSTAT selezionando le domande di brevetto con anno di deposito compreso tra il 2016 e il 2020.

L'attribuzione territoriale è effettuata sulla base dell'indirizzo dell'inventore e l'aggregazione avviene alla scala dei Sistemi Locali del Lavoro (SLL) 2021.

Poiché un brevetto può coinvolgere inventori localizzati in più territori, si è adottato un criterio di conteggio frazionario per territorio (SLL). La quota attribuita non dipende dal numero totale di inventori, bensì dal numero di SLL distinti coinvolti nel brevetto.

Per ciascun brevetto p , sia m_p il numero di SLL distinti associati agli inventori del brevetto, a ciascun SLL coinvolto viene attribuita una quota pari a $\frac{1}{m_p}$.

Formalmente, il numero di brevetti EPO attribuiti al SLL j è definito come:

$$EPO_j = \sum_{p \in P} \mathbf{1}(j \in SLL_p) \cdot \frac{1}{m_p}$$

dove P è l'insieme dei brevetti nel periodo considerato e m_p il numero di SLL distinti coinvolti nel brevetto p . Tale procedura garantisce che ogni brevetto contribuisca complessivamente per un'unità alla distribuzione territoriale, evitando sovrastime nei casi di collaborazione inter-territoriale.

Al fine di ottenere una misura di intensità inventiva comparabile tra territori di diversa dimensione economica, il conteggio frazionario è stato normalizzato per il numero di addetti del SLL j :

$$EPO_weight_j = \frac{EPO_j}{E_j}$$

dove E_j rappresenta il numero di addetti nel SLL j .

3.6.1. Mappe coropletiche

In linea con l'impostazione di Paci e Usai (2000), la rappresentazione cartografica è utilizzata come strumento esplorativo per valutare l'eterogeneità territoriale dell'output inventivo e l'eventuale presenza di pattern spaziali non casuali che saranno poi verificati formalmente con statistiche di autocorrelazione.

Le mappe di EPO_weight_j sono riportate secondo due schemi di classificazione.

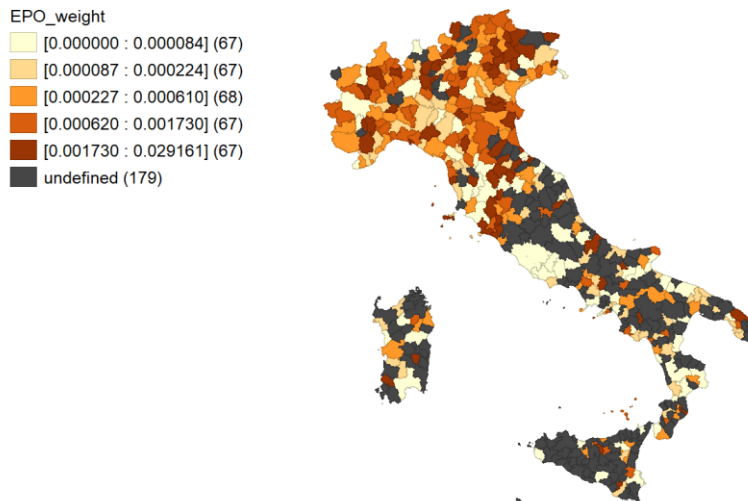


Figure 4: mappa per quantili di EPO_weightj su 5 classi di uguale numerosità

La mappa per quantili (*figura 4*) consente di evidenziare la posizione relativa dei sistemi locali nella distribuzione nazionale dell'intensità brevettuale, limitando l'effetto di pochi valori estremi.

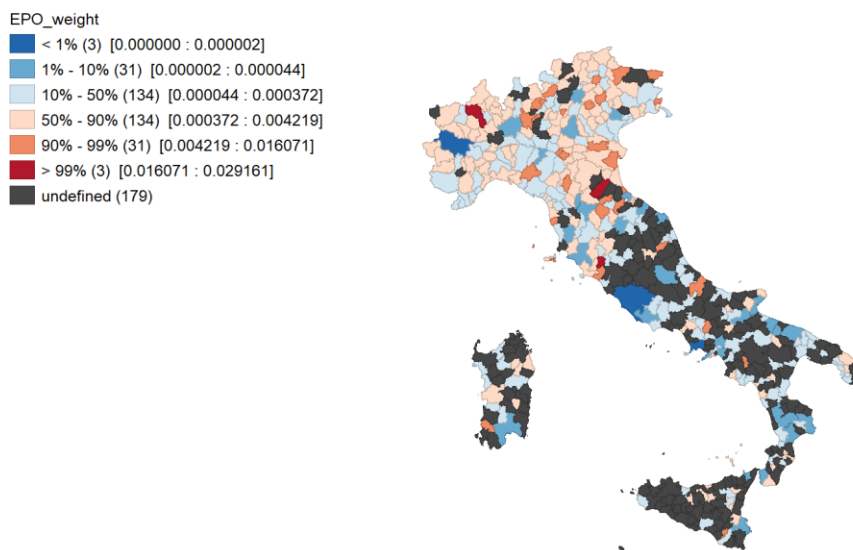


Figure 5: mappa per percentili di EPO_weightj

La mappa per percentili (*figura 5*) consente invece di isolare i territori che concentrano i livelli più elevati dell'indicatore, esplicitando la natura heavy-tailed della distribuzione. In entrambe le rappresentazioni emerge una marcata asimmetria: una quota rilevante di SLL presenta valori nulli o prossimi allo zero, mentre un numero ristretto di territori si colloca nelle classi superiori. Dal punto di vista geografico, i valori relativamente più alti risultano più frequenti nell'Italia Settentrionale e in alcuni sistemi del Centro, mentre nel Mezzogiorno prevalgono livelli molto bassi e un'elevata incidenza di valori nulli.

La lettura cartografica dell'intensità non restituisce grandi blocchi territoriali compatti: accanto a nuclei più innovativi si osserva una certa frammentazione, con sistemi ad alta intensità che non sempre risultano contigui e con ampie porzioni del Paese caratterizzate da valori bassi o nulli. La classificazione per percentili rende particolarmente evidente la concentrazione della coda superiore:

pochi sistemi locali collocati oltre il 90° e, soprattutto, oltre il 99° percentile concentrano i livelli più elevati di EPO_weight_j , indicando che l'attività inventiva di tipo EPO, quando misurata in termini relativi, resta fortemente polarizzata su un set limitato di territori.

Nelle mappe coropletiche della produttività brevettuale EPO emerge una distribuzione fortemente asimmetrica, con una coda superiore molto sottile che riguarda un numero limitato di Sistemi Locali del Lavoro.

La classificazione per percentili consente di identificare in modo chiaro gli SLL collocati oltre il 99° percentile, che rappresentano i territori con la più alta intensità relativa di brevetti EPO rispetto alla base occupazionale:

- Borgosesia (Piemonte);
- Pitigliano e Bibbiena (Toscana);
- Faenza (Emilia-Romagna);
- Comunanza (Marche);
- Telesse Terme (Campania);

È rilevante osservare che tali massimi non coincidono esclusivamente con grandi poli metropolitani al contrario, compaiono diversi sistemi locali di dimensione medio-piccola.

Questo risultato è coerente con la natura dell'indicatore, che normalizza l'output inventivo per il numero di addetti e quindi cattura una misura di intensità relativa (propensione brevettuale) più che la massa inventiva totale. L'analisi dei primi 20 SLL per valori di EPO_weight_j conferma inoltre una forte presenza del Centro-Nord: l'Emilia-Romagna risulta la regione più rappresentata, seguita da Toscana e Lombardia, mentre Veneto e Piemonte compaiono con minore frequenza. Sono presenti anche casi isolati in Lazio, Marche, Abruzzo, Campania e Sardegna.

Nel complesso, le mappe suggeriscono che la brevettazione EPO, una volta controllata per la scala occupazionale, tende a concentrarsi in un insieme ristretto di sistemi locali caratterizzati da elevata intensità relativa.

È opportuno sottolineare che 179 SLL presentano un valore nullo dell'indicatore. Tale evidenza, tuttavia, non può essere interpretata integralmente come assenza di attività inventiva. Su un totale di 515 SLL, in 468 sistemi locali è stata rintracciata attività brevettuale EPO in valore assoluto; pertanto, una quota dei valori nulli osservati nell'indicatore normalizzato non deriva dalla mancanza di brevetti, bensì da un problema informativo relativo al denominatore.

In particolare, per alcuni SLL non è disponibile il numero di addetti a seguito del processo di armonizzazione territoriale tra il Censimento 2011 e la delimitazione SLL 2021. In tali casi l'indicatore di intensità risulta nullo per costruzione (o non definibile), generando un'ulteriore massa di zeri che non riflette una reale assenza di output inventivo, ma una lacuna nei dati occupazionali. Questo fenomeno introduce una potenziale distorsione nella distribuzione dell'indicatore, amplificando artificialmente la quota di sistemi con intensità nulla e influenzando le statistiche di autocorrelazione spaziale.

Per evitare tale bias, nell'analisi di Moran si è adottato un criterio selettivo: il valore zero dell'indicatore è stato mantenuto esclusivamente per gli SLL nei quali il numero di brevetti attribuiti è effettivamente pari a zero, mentre sono stati esclusi dall'analisi spaziale quegli SLL in cui il valore nullo di EPO_weight_j è imputabile all'assenza del denominatore occupazionale. In questo modo si è evitato di alterare la distribuzione dell'indicatore con zeri non informativi, preservando la coerenza economica dell'analisi e garantendo una valutazione più robusta della dipendenza spaziale.

A completamento dell'analisi esplorativa, è utile confrontare queste evidenze con le mappe costruite sul numero assoluto di brevetti num_EPO_j

Tale confronto è rilevante perché num_EPO_j e EPO_weight_j misurano grandezze concettualmente diverse: la prima descrive la massa inventiva totale localizzata in ciascun SLL, mentre la seconda misura una propensione relativa alla brevettazione EPO condizionata alla scala economica locale.

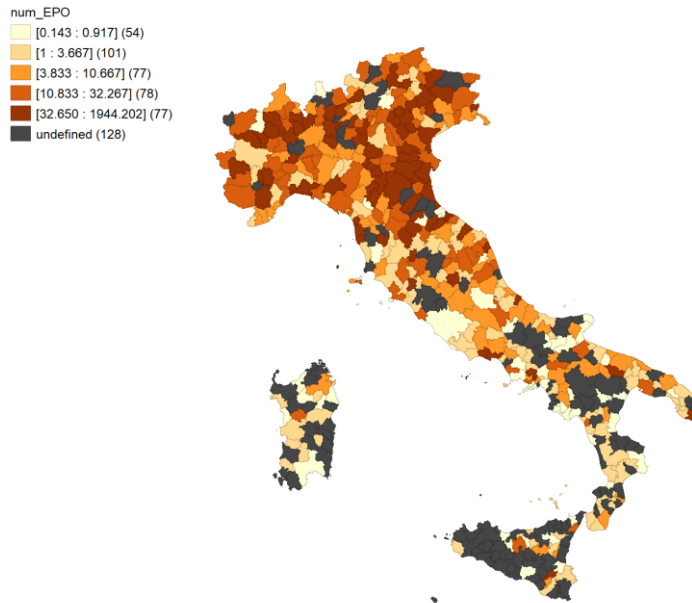


Figure 6: mappa per quantili di num_EPO_j , 5 classi di uguale dimensione

Nelle mappe basate sui conteggi assoluti (*figura 6 e figura 7*), la concentrazione nei principali poli urbani e produttivi risulta più netta e sistematica, poiché i territori con maggiore base occupazionale e produttiva tendono meccanicamente a generare più brevetti. Al contrario, la normalizzazione per addetti attenua l'effetto scala e può modificare il ranking territoriale: alcuni sistemi locali grandi riducono la propria posizione relativa, mentre alcuni sistemi intermedi possono emergere con intensità più elevata. In questa prospettiva, la maggiore presenza di valori nulli osservata nelle mappe di EPO_weight_j va interpretata con cautela: essa riflette in parte l'assenza effettiva di brevettazione EPO in numerosi sistemi locali, e in parte la struttura del denominatore L_j , che risente dell'armonizzazione della base occupazionale (Censimento 2011) alla delimitazione SLL 2021.

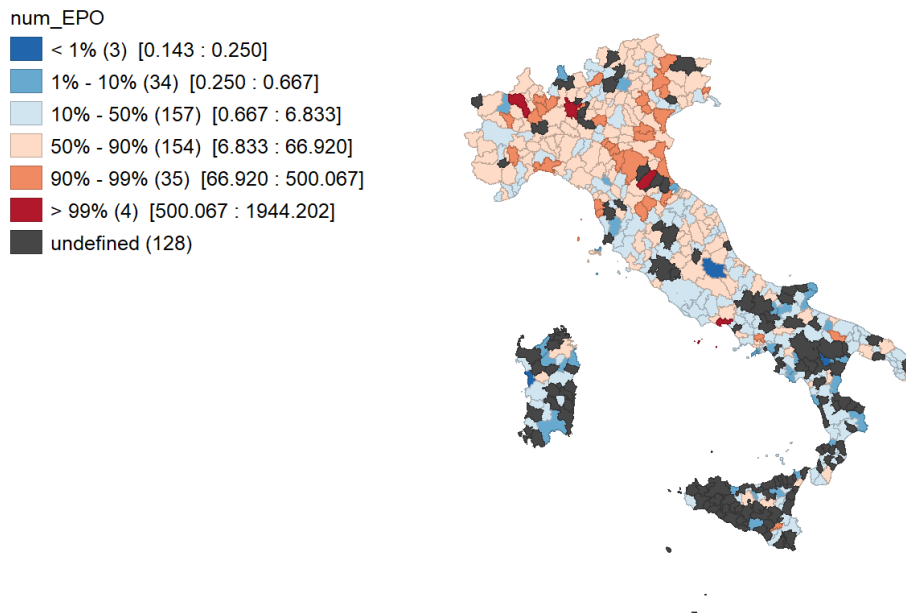


Figure 7: mappa per percentili di num_EPOj

Ne consegue che le mappe dei conteggi assoluti e quelle dell'intensità non sono alternative ma complementari: le prime descrivono dove si concentra l'output inventivo in termini di volume, le seconde dove la brevettazione EPO è relativamente più intensa rispetto alla base economica locale.

Le implicazioni di questa eterogeneità e la presenza (o meno) di dipendenza spaziale verranno poi valutate formalmente mediante Moran's I e LISA nella sezione successiva.

3.6.2. Distribuzione della funzione EPO_weight_j

Ai fini delle analisi di autocorrelazione spaziale su EPO_weight_j , dal dataset sono stati esclusi 81 SLL caratterizzati da un denominatore occupazionale non informativo ($E_j=0$), conseguenza dell'armonizzazione della base occupazionale dal Censimento 2011 alla delimitazione SLL 2021. Nelle rappresentazioni cartografiche tali unità sono evidenziate in nero (figura 8).

Questa scelta è metodologicamente rilevante, se si fossero considerate tali osservazioni come "zeri" dell'indicatore, si avrebbe ottenuto un'alterazione della distribuzione (media, varianza e asimmetria) con l'introduzione di una massa artificiale di valori nulli e con conseguenze dirette sui test di dipendenza spaziale.

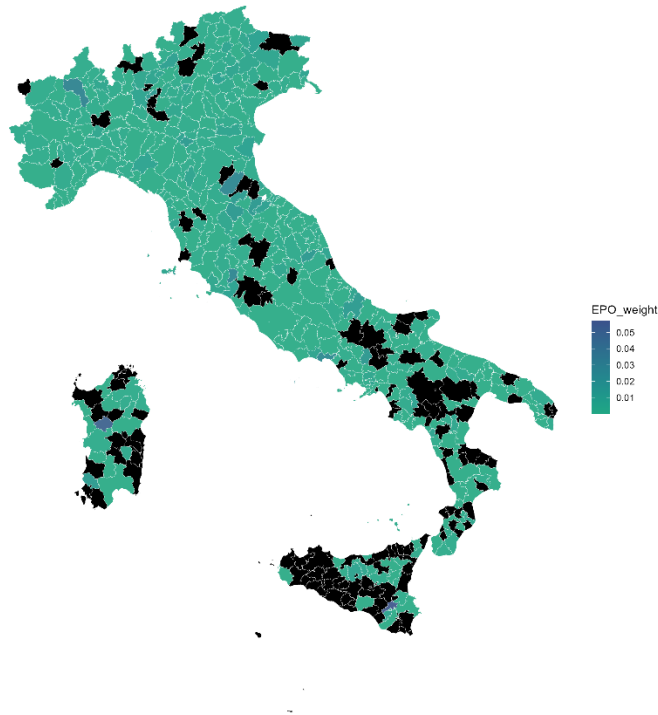


Figure 8: mappa coropletica di *EPO_weight*, in nero gli SLL esclusi dall'analisi

La distribuzione di *EPO_weight* risulta fortemente asimmetrica (right-skewed), la maggior parte degli SLL presenta valori molto bassi, mentre un numero limitato di osservazioni assume valori elevati, generando una coda destra pronunciata. Coerentemente con le statistiche descrittive, la mediana è inferiore alla media, indicando che pochi valori alti contribuiscono in modo rilevante al livello medio dell'indicatore. Questa configurazione suggerisce una forte eterogeneità territoriale nell'intensità brevettuale per addetto, i quantili confermano la presenza di pochi valori estremi (*Tabella 2*).

Table 2: riepilogo delle statistiche descrittive e quantili della distribuzione *EPO_weight*

STATISTICHE DESCRITTIVE		
MEDIANA 4.23E-04	MEDIA 1.91e-03	MASSIMO 5.72e-02
QUANTILI		
90° PERCENTILE 0,00455	95° PERCENTILE 0,00826	99° PERCENTILE 0,02455

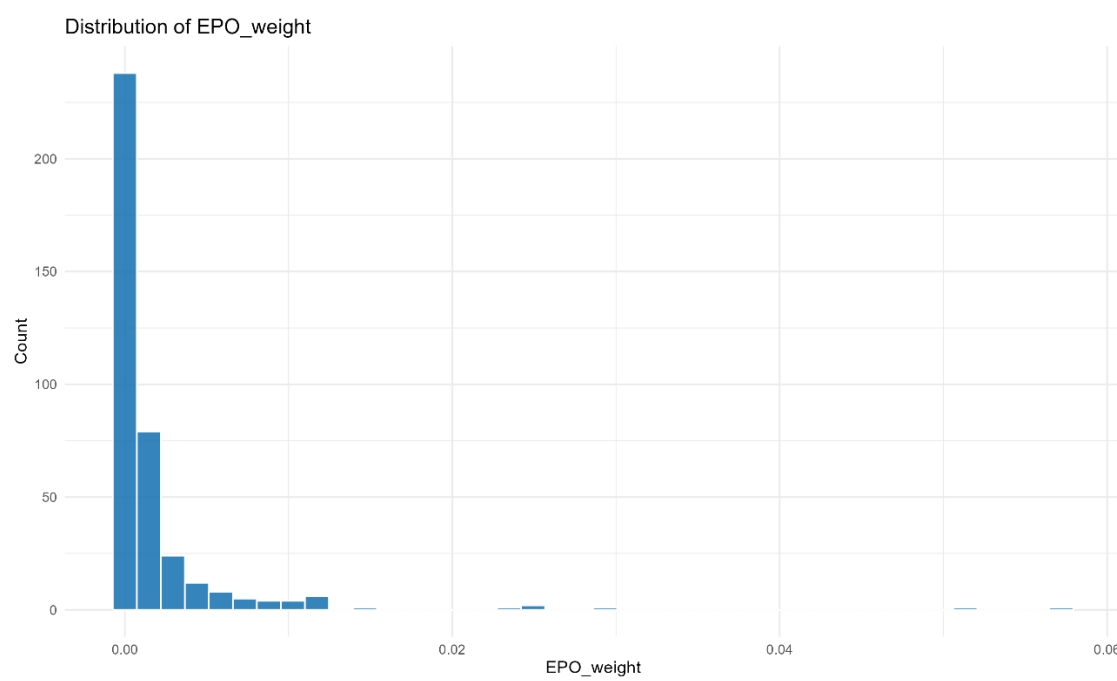


Figure 9: distribuzione dell'indicatore EPO_weight a livello di SLL, istogramma in scala naturale

La figura 9 mostra la distribuzione dell'indicatore in scala naturale, per rappresentarla in modo informativo si adotta un istogramma (figura 10) con asse orizzontale in scala logaritmica. La trasformazione logaritmica riduce l'influenza visiva della coda destra e consente di distinguere in modo più chiaro la struttura della distribuzione nella parte bassa (dove si concentra la maggioranza degli SLL). In altri termini, mentre la scala naturale evidenzia soprattutto la presenza di pochi valori estremi, la scala logaritmica permette di analizzare più finemente le differenze relative tra gli SLL a bassa intensità, rendendo evidente che la dispersione è presente lungo tutto l'intervallo ma si concentra prevalentemente su valori ridotti.

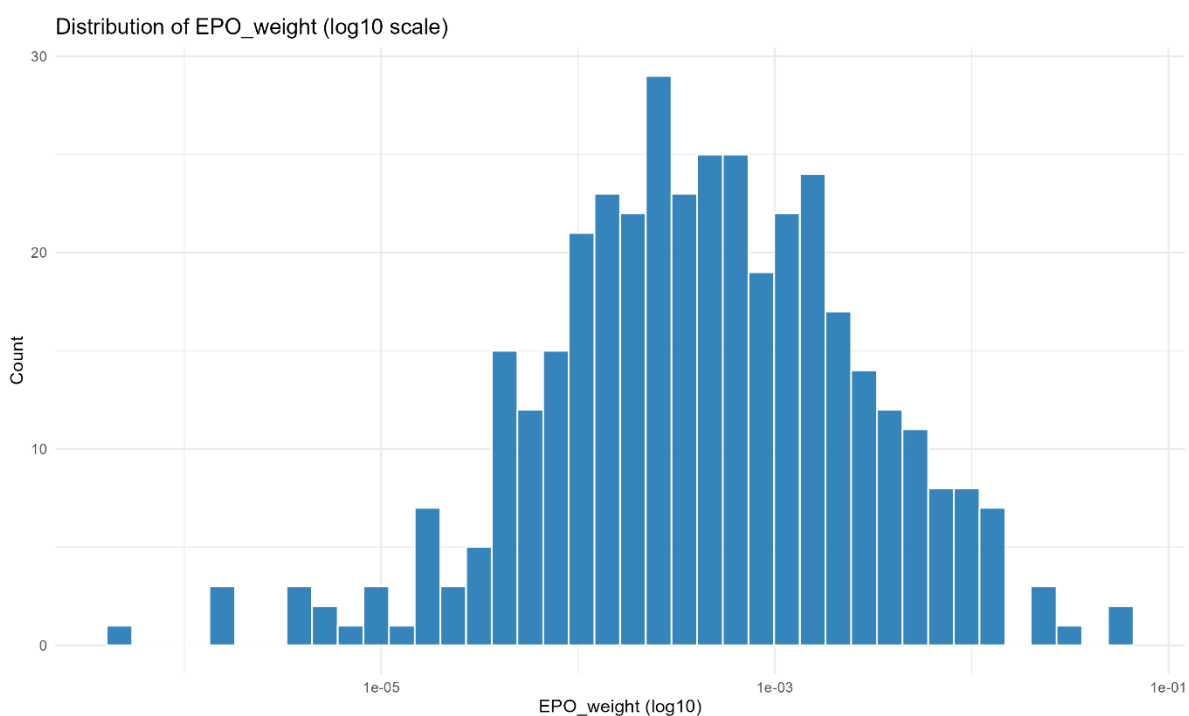


Figure 10: distribuzione dell'indicatore EPO_weight a livello di SLL, istogramma con asse orizzontale in scala logaritmica

Questa struttura distributiva è tipica di indicatori di intensità (pochi territori molto “performanti” e una massa di territori su valori bassi) e anticipa un possibile esito: valori elevati possono manifestarsi come episodi localizzati e non necessariamente come cluster spaziali contigui e statisticamente robusti.

3.6.3. Moran’s I globale e ordini di contiguità

L’indice di Moran dipende in modo cruciale dalla matrice dei pesi spaziali (W). In questa applicazione (W) è definita mediante contiguità queen tra poligoni e standardizzazione per riga (stile “ W ”), così che ogni SLL sia confrontato con una media ponderata dei valori osservati nei territori contigui, indipendentemente dal numero di vicini.

L’analisi è inoltre replicata su diversi ordini di contiguità, che introducono un “raggio spaziale” crescente: 1° ordine (confinanti diretti), 2° ordine (vicini dei vicini) e 3° ordine (vicinato più ampio). In presenza di esternalità locali robuste, ci si attenderebbe un valore di Moran positivo e statisticamente significativo al 1° ordine e una persistenza anche a ordini superiori, con intensità che può decrescere per effetto “distance decay”.

Nel caso di EPO_weight_j , i risultati non supportano l’esistenza di autocorrelazione spaziale globale positiva. Con la matrice di contiguità al 1° ordine si ottiene:

$$I = -0,001045$$

$$E[I] = -0,002646$$

Pertanto, non vi è evidenza di clustering globale dell’intensità brevettuale EPO tra unità contigue. Coerentemente anche la comparazione per ordini di contiguità non evidenzia un segnale che emerga a scale spaziali più ampie: per il 2° e 3° ordine Moran’s I resta prossimo a zero o leggermente negativo e non significativo (*Tabella 3*).

Table 3: Correlazione spaziale Moran’s I globale della produttività dell’innovazione per diversi ordini di contiguità

	Test analitico – Normale		Test permutazionale – Monte Carlo	
	I	Prob.	I	Prob.
1° ordine di contiguità	-0.0010445	0.4818524	-0.0010445	0.372
2° ordine di contiguità	-0.0067155	0.555800	-0.0067155	0.498
3° ordine di contiguità	-0.0262382	0.8113992	-0.0262382	0.871

La progressiva negatività al 3° ordine non è interpretabile come dispersione spaziale, poiché non accompagnata da significatività statistica: indica piuttosto che la correlazione spaziale media è assente o troppo debole rispetto all’eterogeneità interna dell’indicatore.

Questo risultato è coerente con il Moran scatterplot (*Figura 11*), in cui la massa dei punti si concentra vicino a valori bassi sull’asse delle ascisse (EPO_weight), mentre il lag spaziale non cresce sistematicamente con l’aumentare dell’indicatore. In termini interpretativi, l’assenza di un segnale

globale suggerisce che l'intensità brevettuale per addetto non si organizza in un pattern nazionale di cluster contigui, ma presenta piuttosto eterogeneità locale e discontinuità spaziale.

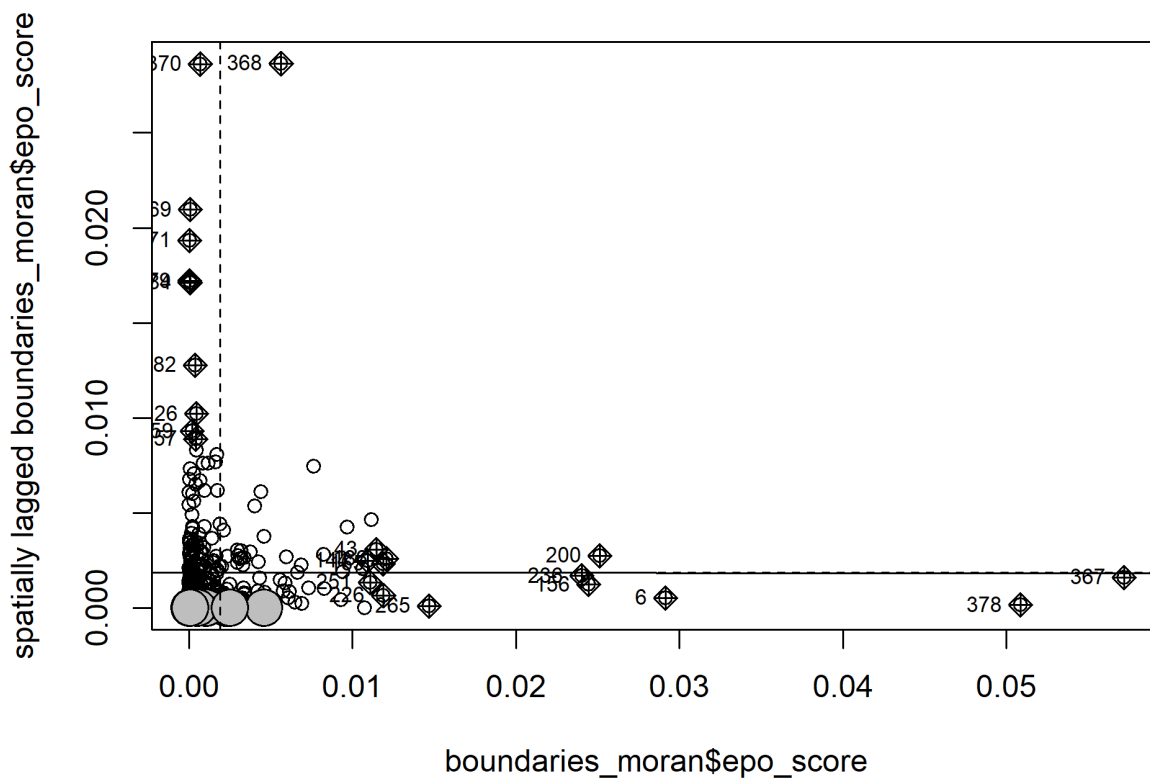


Figure 11: Moran scatterplot di EPO_weight a livello di SLL

Questa configurazione è tipica di indicatori altamente asimmetrici, con massa rilevante su valori bassi e una coda superiore ristretta: eventuali “poli” possono essere reali ma troppo localizzati per generare autocorrelazione globale.

3.6.4. LISA e significatività locale

L’analisi LISA (Local Indicators of Spatial Association) permette di passare dalla misura globale di autocorrelazione a una lettura locale della dipendenza spaziale. La classificazione dei cluster restituisce i risultati riepilogati in *tabella 4*.

Table 4: classificazione dei cluster secondo l’analisi LISA

Non significant	High – High	Low – Low	Low – High	High – Low	Undefined	Isolated
325 SLL	2 SLL	27 SLL	20 SLL	5 SLL	8 SLL	0 SLL

La mappa LISA (*figura 12*) evidenzia una netta prevalenza di aree Not significant, indicando che nella maggior parte degli SLL non emerge un pattern locale sufficientemente robusto di somiglianza/dissonanza territoriale. In questo quadro, i cluster significativi risultano pochi e frammentati: i rari High - High suggeriscono micro-aree in cui l'intensità elevata è accompagnata da un vicinato anch'esso elevato; i Low -Low indicano invece aree di persistenza locale di bassa intensità.

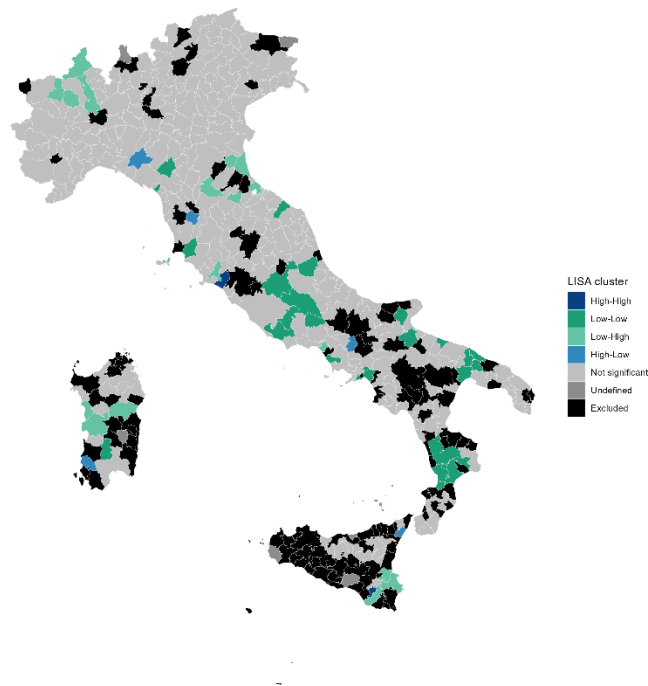


Figure 12: mappa LISA di EPO_weight a livello di SLL

La lettura dei cluster è naturalmente complementare alla mappa di significatività del Moran locale (p-value) (*figura 13*), quest'ultima fornisce la componente inferenziale, distinguendo le aree in cui l'associazione locale è statisticamente distinta dal caso dalle zone in cui le differenze territoriali osservate nella coropletica non si traducono in dipendenza spaziale significativa.

La maggior parte delle unità non presenta autocorrelazione locale statisticamente supportata:

- $p \leq 0,01$: 10 SLL
- $0,01 < p \leq 0,05$: 11 SLL
- $0,05 < p \leq 0,10$: 3 SLL
- $p > 0,10$: 355 SLL

L'assenza di Moran's I globale significativo è coerente con una mappa LISA dominata da unità non significative e da un numero limitato di cluster/outlier locali. Pertanto, per EPO_weight_j non emerge un processo di autocorrelazione spaziale generalizzato. La componente spaziale appare localizzata e concentrata in pochi episodi, insufficienti a produrre un segnale globale robusto.

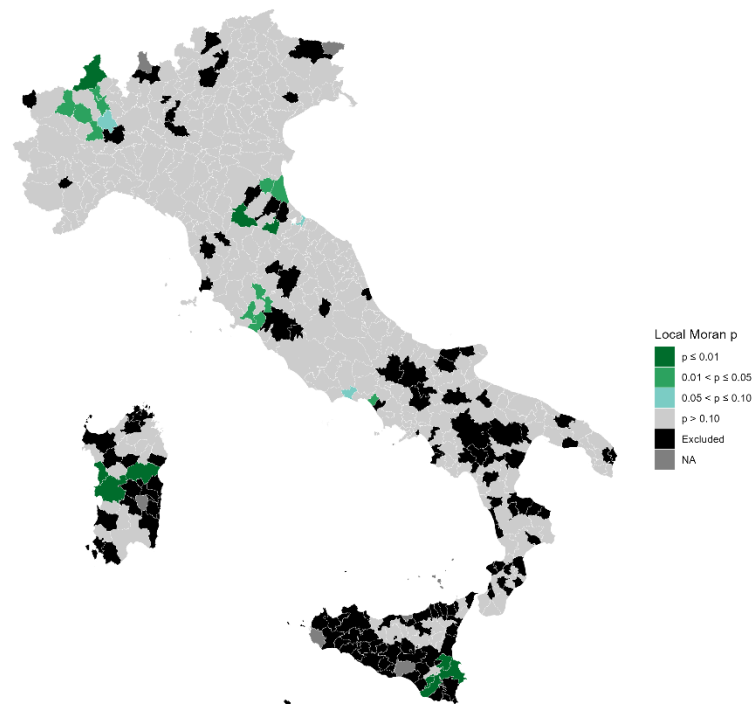


Figure 13: mappa di significatività del Moran locale di *EPO_weight* a livello di SLL

Dal punto di vista geografico, si osserva la comparsa di configurazioni “a corona”, in cui diversi SLL risultano Low - High attorno a uno SLL centrale che rimane Not significant.

Utilizzando la contiguità di primo ordine, si identificano cinque “centri” di questo tipo (tutti Not significant) con un numero elevato di vicini Low - High:

- Borgosesia (Piemonte) con 6 vicini Low - High: Biella, Novara, Borgomanero, Domodossola, Omegna, Ivrea
- Verbania (Piemonte) con 3 vicini Low - High: Borgomanero, Domodossola, Omegna
- Faenza (Emilia-Romagna) con 3 vicini Low - High: Ravenna, Lugo, Borgo San Lorenzo
- Mazzarone (Sicilia) con 3 vicini Low - High: Lentini, Siracusa, Ragusa
- Macomer (Sardegna) con 3 vicini Low - High: Nuoro, Oristano, Bosa

Questa struttura supporta l'interpretazione secondo cui, in alcune aree, l'intensità brevettuale non si organizza in un cluster “compatto” (High - High) sul territorio centrale, ma genera contrasti di prossimità: il centro non presenta un'associazione locale sufficientemente stabile da risultare significativo, mentre il vicinato evidenzia una configurazione coerente con Low - High, tipica di contesti polarizzati e disomogenei (micro-specializzazioni locali, differenze strutturali tra SLL adiacenti, e/o presenza di pochi picchi non accompagnati da un “blocco” contiguo di valori elevati).

L'integrazione tra mappa LISA e mappa di significatività del Moran locale suggerisce che la struttura spaziale di *EPO_weight* sia prevalentemente fatta di segnali locali e discontinui (outlier e micro-cluster), più che di grandi aree contigue e statisticamente robuste.

3.7. Start-up innovative: analisi della distribuzione spaziale

L'indicatore relativo alle start-up innovative è costruito a partire da un database amministrativo ottenuto tramite estrazione dal Registro delle Imprese. All'interno di tale archivio sono state selezionate le imprese osservate nella finestra 2016 - 2020.

La scelta di una finestra quinquennale è coerente con il quadro normativo che disciplina questa tipologia di impresa: lo status di start-up innovativa è infatti temporaneo e limitato ad un orizzonte di 5 anni. Questa scelta consente quindi di misurare il fenomeno in modo più realistico, cogliendone la dinamica di entrata e uscita e riducendo la volatilità tipica dei conteggi annuali, soprattutto nei territori di dimensioni minori.

L'attribuzione territoriale è effettuata sulla base dell'indirizzo della sede legale e l'aggregazione avviene alla scala dei Sistemi Locali del Lavoro (SLL) 2021. In questo contesto, la sede legale rappresenta il criterio di localizzazione più appropriato non solo per ragioni di disponibilità informativa, ma anche perché riflette in modo credibile dove l'attività innovativa è effettivamente ancorata e organizzata.

Le start-up innovative, soprattutto nelle fasi iniziali, presentano infatti assetti operativi spesso fluidi: team distribuiti su più città, lavoro da remoto, utilizzo di coworking e incubatori, e talvolta una distinzione non netta tra "sede operativa" e luoghi in cui si svolgono attività quotidiane.

La sede legale, al contrario, costituisce l'ancoraggio istituzionale dell'impresa: è il luogo che concentra gli elementi di governance e responsabilità giuridico-amministrativa (registrazione, adempimenti, rappresentanza legale) e, in molti casi, coincide con l'ecosistema che sostiene la start-up (relazioni con incubatori, professionisti, università, investitori e servizi specializzati). Ne consegue che la sede legale è spesso la miglior proxy disponibile per identificare il territorio in cui l'innovazione imprenditoriale si insedia, si struttura e si connette al contesto locale.

A differenza dei brevetti, che possono coinvolgere inventori localizzati in più territori e richiedono quindi un criterio di conteggio frazionario, ciascuna start-up è associata a una sola sede legale e contribuisce integralmente al territorio in cui essa ricade.

Formalmente, il numero di start-up attribuite al SLL j è definito come:

$$Startup_j = \sum_{s \in S} \mathbf{1}(HQ_s \in SLL_j)$$

dove S è l'insieme delle start-up nella finestra considerata, HQ_s indica la sede legale della start-up s , e $\mathbf{1}(\cdot)$ è una funzione indicatrice che vale 1 se la condizione è soddisfatta e 0 altrimenti.

Per ottenere una misura di intensità imprenditoriale innovativa comparabile tra territori di diversa dimensione economica, il conteggio è stato normalizzato per il numero di addetti del SLL j :

$$Startup_weight_j = \frac{Startup_j}{E_j}$$

dove E_j rappresenta il numero di addetti nel SLL j .

3.7.1. Mappe coropletiche

Le mappe coropletiche rappresentano la distribuzione territoriale delle start-up innovative (finestra temporale 2016 - 2020) su base SLL 2021, confrontando due misure:

1. **#startup** (valore assoluto): quante start-up ricadono nel singolo SLL.
2. **start_core** ($\#start-up_i / addetti_i$): quante start-up si osservano a parità di dimensione economica, normalizzando per il numero di addetti.

Le mappe sono presentate in due versioni di classificazione: una a percentili, che enfatizza la “coda” dei valori più alti) ed una a quantili che rende più leggibile il confronto tra gruppi di SLL. In tutte le mappe, gli SLL in grigio “undefined” sono quelli privi di valore perché non registrano attività innovativa nel periodo.

(i) Indicatore **#startup_i** in valore assoluto per SLL

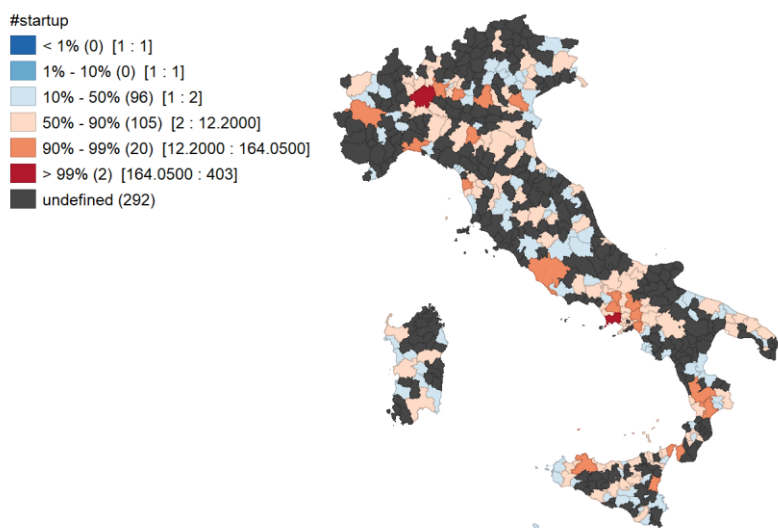


Figure 14: mappa percentile di #startup a livello di SLL

La lettura della mappa a percentili (figura 14) rende evidente quanto la distribuzione sia “sbilanciata” verso pochi territori:

- La classe > 99% comprende solo 2 SLL e copre l’intervallo [164; 403] start-up: due veri “poli” che emergono nettamente rispetto al resto del Paese. Visivamente, uno è collocato nel Nord (area padana) e l’altro nel Mezzogiorno tirrenico.
- La classe 90 - 99% mostra un secondo livello di territori ad alta concentrazione, distribuiti soprattutto lungo la dorsale del Nord e in alcune aree metropolitane/di nodo anche nel Centro-Sud.

Questa struttura indica che il fenomeno start-up, in termini di massa, è fortemente polarizzato.

La mappa a quantili (figura 15) per #startup conferma due aspetti:

- Una quota ampia di territori osservati rientra nelle classi basse: 96 SLL hanno esattamente 1 start-up, e 80 SLL ne hanno tra le 2 e le 4. Questo suggerisce che, anche tra i territori attivi, spesso la presenza di start-up è episodica o comunque numericamente contenuta.
- La classe più alta (5 - 403 start-up) include 47 SLL: quindi la presenza significativa (almeno 5 start-up) riguarda una minoranza di casi e tende a organizzarsi in aree contigue soprattutto nel Nord e lungo alcuni corridoi del Centro-Sud. Dal momento che questa classe agglomera SLL con valori molto diversi di #startup, la *tabella 5* mostra i 9 SLL che hanno valori più alti di numero di startup.

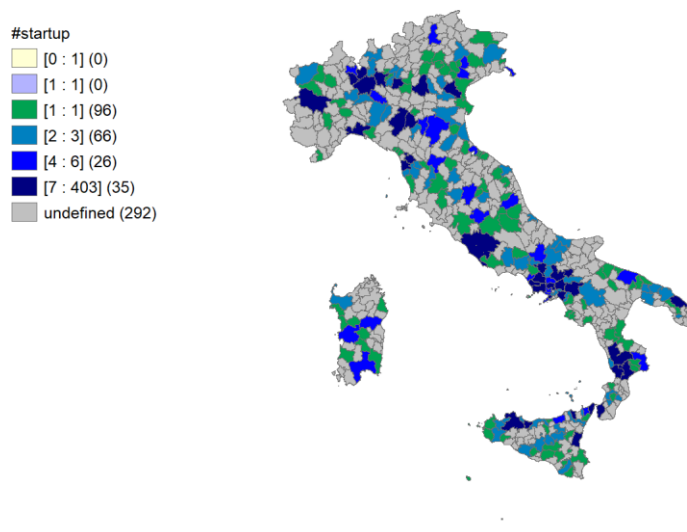


Figure 15: mappa a quantili di #startup a livello di SLL

Il dato più strutturale è però la quantità di “undefined”: 292 SLL risultano senza start-up. Questo non è un dettaglio cartografica. Indica che, nel quinquennio, una parte molto ampia del territorio non esprime output imprenditoriale innovativo misurabile.

Table 5: i 9 UIBM SLL con il #startup più alto

SLL 2021	#startup per SLL
<i>Milano</i>	403
<i>Napoli</i>	248
<i>Roma</i>	133
<i>Palermo</i>	102
<i>Caserta</i>	56
<i>Torino</i>	42
<i>Genova</i>	38
<i>Messina</i>	32
<i>Avellino</i>	30

Ne consegue una geografia a macchie: poli e fasce attive intervallate da aree con assenza totale di fenomeno.

(ii) Indicatore $startup_weight_j$ ($\#start-up_j / addetti_j$)

Normalizzare per addetti cambia la domanda a cui risponde la mappa, in questo caso si osserva dove le start-up sono più dense rispetto alla dimensione del mercato del lavoro.

Nella mappa a percentili (*figura 16*) di $startup_weight_j$ (richiamato come $startup_core$ nella legenda delle mappe) i valori massimi (> 99%) sono ancora 2 SLL, ma questa volta la loro collocazione è chiaramente nelle Isole (si distinguono due hotspot in Sicilia).

Questo è un risultato importante: indica che alcuni territori non sono necessariamente campioni per volume assoluto, ma diventano estremi quando si guarda la propensione relativa (start-up per addetto).

Parallelamente, molte aree del Nord che erano ben visibili nella parte alta con $\#startup$ rientrano ora in classi intermedie: non perché perdano importanza in termini assoluti, ma perché l'elevato numero di addetti diluisce l'intensità.

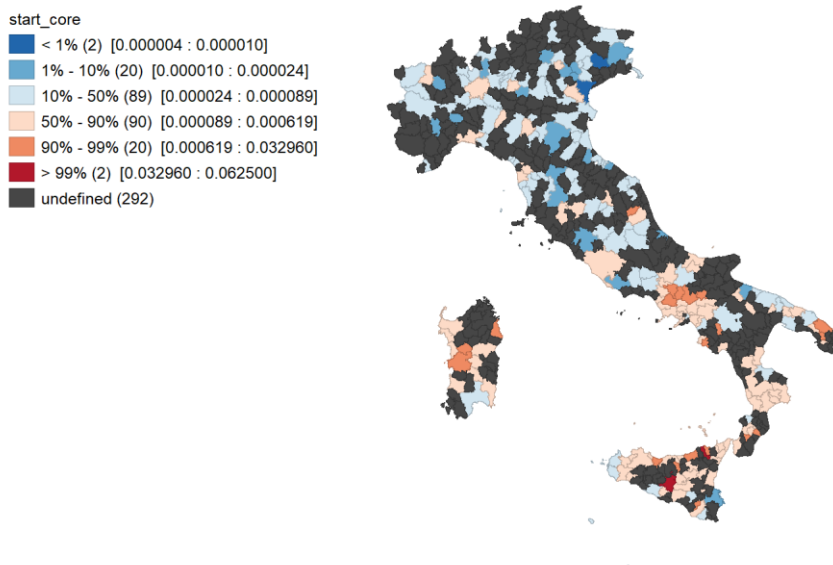


Figure 16: mappa a percentili di $startup_weight_j$ a livello di SLL

Nella mappa a quantili (*figura 17*) di $startup_weight_j$ (richiamato come $startup_core$ nella legenda delle mappe), la classe più alta (intervallo ampio fino a 0.0625) si concentra in modo più evidente nel Mezzogiorno e nelle Isole, con una presenza particolarmente marcata in Sicilia e segnali rilevanti anche in Sardegna.

Il Nord resta certamente presente, ma l'immagine complessiva suggerisce che la normalizzazione fa emergere una componente diversa: territori dove l'innovazione imprenditoriale è più intensa rispetto alla base occupazionale locale.

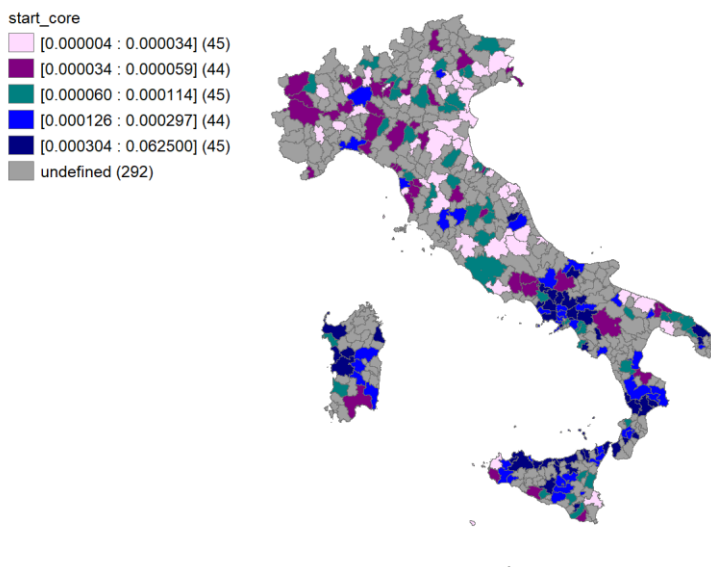


Figure 17: mappa a quantili di $startup_weight_j$ a livello di SLL

Il confronto tra $\#startup$ e $startup_weight_j$, mette in luce due configurazioni che coesistono:

- Poli di massa (valori assoluti): pochi SLL dominano per numerosità e strutturano gerarchie forti, intorno si osservano altri SLL mediamente attivi, ma la coda superiore resta molto concentrata.
- Poli di intensità (normalizzati): alcuni territori, soprattutto nelle Isole e in parte nel Centro-Sud, emergono come “outlier” quando l’output innovativo è letto in rapporto alla dimensione economica, segnalando una propensione relativamente elevata alla creazione di start-up innovative.

Il numero elevato di SLL con zero start-up e la presenza di pochi poli estremi suggeriscono un fenomeno che non è diffuso uniformemente, ma organizzato in hub e corridoi. Gli hotspot di intensità non coincidono necessariamente con quelli di volume.

3.7.2. Distribuzione della funzione $startup_weight_j$

La distribuzione di $startup_weight_j$ risulta fortemente zero-inflated: su 515 SLL complessivi, 292 presentano $startup_weight_j = 0$, pari al 56,7% del totale. Questo significa che in oltre metà dei sistemi locali, nel quinquennio 2016 - 2020, non è presente alcuna start-up innovativa attribuibile al SLL e dunque l’indicatore assume valore nullo.

È importante sottolineare che, in questo caso, lo zero è pienamente informativo: non deriva da elementi di armonizzazione territoriale ma segnala direttamente l’assenza effettiva di start-up innovative nel periodo considerato per quello specifico SLL. Di conseguenza, la distribuzione non è solo asimmetrica, ma presenta una marcata massa di zeri che riflette una reale eterogeneità territoriale nell’emersione dell’imprenditorialità innovativa.

Tra gli SLL con valore positivo, l’indicatore mostra inoltre una forte concentrazione e una coda superiore lunga. Le statistiche confermano che la maggior parte dei valori è estremamente bassa: il 25° percentile (Q1) e la mediana sono entrambi pari a 0, mentre il 75° percentile (Q3) è appena $6,13 \times 10^{-5}$. Anche guardando alla parte alta della distribuzione, i valori restano contenuti fino ai percentili elevati:

il 90° percentile è $2,68 \times 10^{-4}$, il 95° percentile è $5,56 \times 10^{-4}$, mentre il 99° percentile sale a $7,355 \times 10^{-3}$. Il confronto con il massimo osservato (0,0625) evidenzia la presenza di pochissimi outlier estremi: oltre il 99° percentile, l'intensità può crescere di quasi un ordine di grandezza, indicando che una quota molto limitata di SLL concentra livelli eccezionalmente elevati di start-up per addetto.

Nel complesso, $startup_weight_j$ descrive quindi un fenomeno territorialmente discontinuo: una parte ampia di SLL non registra alcuna attività innovativa, mentre tra i territori attivi l'intensità è generalmente molto bassa e soltanto pochi sistemi locali si collocano nella coda superiore, con valori nettamente superiori al resto della distribuzione. Questa configurazione suggerisce già in fase esplorativa un'elevata polarizzazione e rende particolarmente rilevante la verifica formale di eventuali pattern di dipendenza spaziale tramite Moran's I e LISA.

3.7.3. Moran's I globale e ordini di contiguità

Per verificare se l'intensità di start-up innovative per addetto tende ad agglomerarsi nello spazio, l'autocorrelazione è stimata con Moran's I usando una matrice dei pesi costruita su contiguità queen e normalizzazione per riga.

Table 6: Correlazione spaziale Moran's I globale di $startup_weight_j$ per diversi ordini di contiguità

	Test analitico – Normale		Test permutazionale – Monte Carlo	
	I	Prob.	I	Prob.
1° ordine di contiguità	0.0050665	0.3801463	0.0050665	0.149
2° ordine di contiguità	0.0133417	0.1780729	0.0133417	0.114
3° ordine di contiguità	-0.0080053	0.6627208	-0.0080053	0.641

L'analisi è ripetuta su più ordini di contiguità per capire se un eventuale segnale emerge solo a scala più ampia. I risultati mostrati in *tabella 6* indicano che non c'è evidenza di un pattern globale robusto: ai primi due ordini l'indice è leggermente positivo ma non significativo, mentre al terzo ordine diventa leggermente negativo e resta non significativo. In pratica, l'intensità di start-up per addetto non sembra seguire una struttura nazionale regolare in cui territori vicini presentano sistematicamente valori simili.

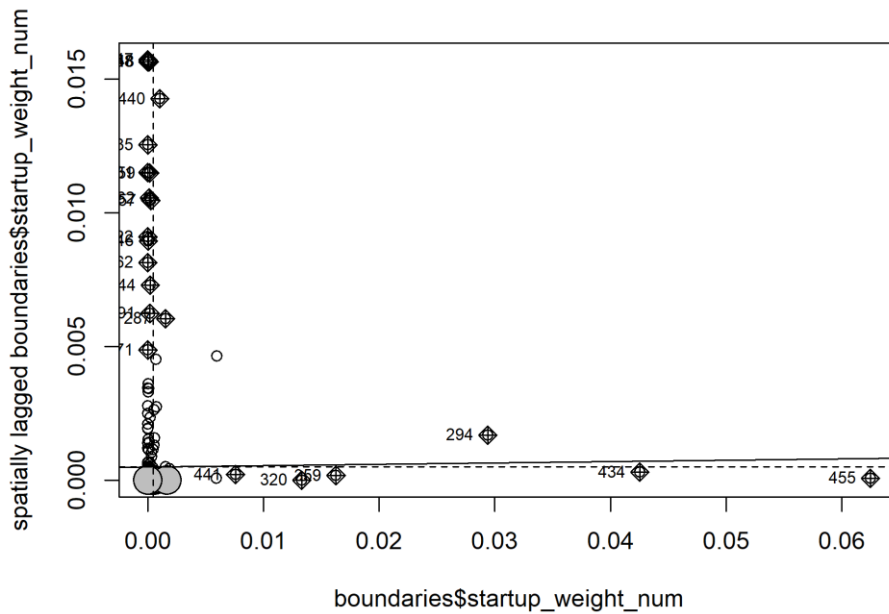


Figure 18: Moran scatterplot *startup_weight_{tj}*

Questa lettura è coerente con il Moran scatterplot (*figura 18*), quasi tutta la nuvola dei punti è compressa vicino allo zero sull'asse orizzontale (molti SLL con valore nullo o quasi nullo), e il lag spaziale non cresce in modo ordinato all'aumentare di *startup_weight_{tj}*. Anche i pochi casi con valori molto elevati appaiono come osservazioni isolate rispetto al contesto circostante: questo tipo di distribuzione (fortemente zero-inflated e con coda lunga) rende difficile che pochi “picchi” bastino a generare autocorrelazione globale.

3.7.4. LISA e significatività locale

Se il test globale non restituisce un segnale netto, diventa molto interessante capire se esistono episodi locali di associazione spaziale. La classificazione LISA (*Tabella 7*) mostra una prevalenza schiacciante di SLL senza pattern locale robusto (469 su 515 sono *Not significant*). Accanto a questa maggioranza, però, emergono alcuni fenomeni circoscritti (*figura 19*).

Table 7: classificazione dei cluster secondo l'analisi LISA

Non significant	High – High	Low – Low	Low – High	High – Low	Undefined	Isolated
469 SLL	8 SLL	13 SLL	22 SLL	1 SLL	2 SLL	0 SLL

Il punto chiave è che la dipendenza spaziale, quando c'è, non si presenta come una grande area compatta, ma come una combinazione di pochi cluster ristretti e diversi episodi di “discontinuità” (soprattutto Low -High), cioè gradienti bruschi tra SLL adiacenti.

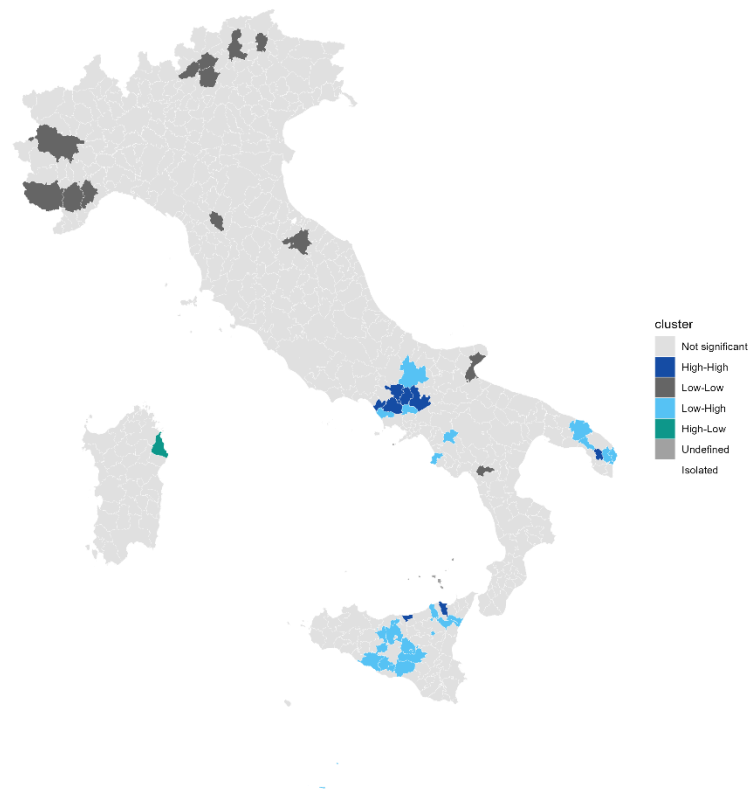


Figure 19: mappa LISA di startup_weight a livello di SLL

Il cluster High - High più evidente è localizzato in Campania, e comprende gli SLL di Caserta, Mondragone, Piedimonte Matese, Benevento e Telesse Terme. In termini LISA questo significa che in quest'area non c'è solo un singolo punto dinamico, ma un piccolo blocco contiguo in cui l'intensità di start-up per addetto è alta e lo è anche nel vicinato immediato: una configurazione coerente con l'idea di un ecosistema territoriale locale (non necessariamente metropolitano) in cui l'attività imprenditoriale innovativa risulta relativamente più intensa.

Diversi SLL vengono classificati come Low - High, cioè territori con intensità bassa ma collocati accanto a vicini più elevati. Questa "corona" Low - High suggerisce che l'innovazione imprenditoriale non si distribuisce in modo omogeneo nello spazio: il nucleo HH indica un'area dove l'intensità è localmente "autorinforzante", mentre l'intorno mette in evidenza contrasti di prossimità (territori contigui ma meno dinamici), un pattern tipico quando un fenomeno è polarizzato e la diffusione spaziale non è uniforme.

In Sicilia emergono due SLL High - High lungo l'asse tirrenico: Patti e Santo Stefano di Camastra. Anche qui non è un singolo outlier isolato, ma un segnale locale coerente, valori relativamente elevati che si accompagnano a un vicinato anch'esso elevato.

Intorno (e più in generale nell'isola) compaiono numerosi Low - High, che includono Capo d'Orlando, Taormina, Randazzo, Enna, Piazza Armerina, Petralia Soprana, Mussomeli, Mazzarino, oltre a Agrigento, Canicattì e Campobello di Licata. L'impressione complessiva è che l'isola combini un nucleo HH (tirrenico) con un insieme di SLL che, pur avendo intensità bassa, risultano collocati in prossimità di aree relativamente più dinamiche: ancora una volta, non un blocco compatto, ma una geografia fatta di poli e discontinuità.

Infine, tra gli High - High compare Galatina (Puglia), e l'elenco Low - High include diversi SLL del Salento e dell'area pugliese come Maglie, Otranto e Copertino (oltre a Brindisi). Questo rafforza l'idea che l'intensità di start-up per addetto possa emergere come specializzazione locale anche fuori dai grandi poli. Un risultato coerente con la natura dell'indicatore normalizzato, che può far emergere hotspot relativi in sistemi locali non metropolitani.

La mappa di significatività del Moran locale (*figura 20*) va letta come la controparte inferenziale: evidenzia che la significatività statistica è concentrata in poche zone, mentre il resto del territorio rimane non significativo. Questo si lega bene ai risultati precedenti: un Moran globale non significativo non esclude che esistano cluster locali, ma suggerisce che quei cluster sono troppo pochi e/o troppo circoscritti per spostare la correlazione media nazionale.

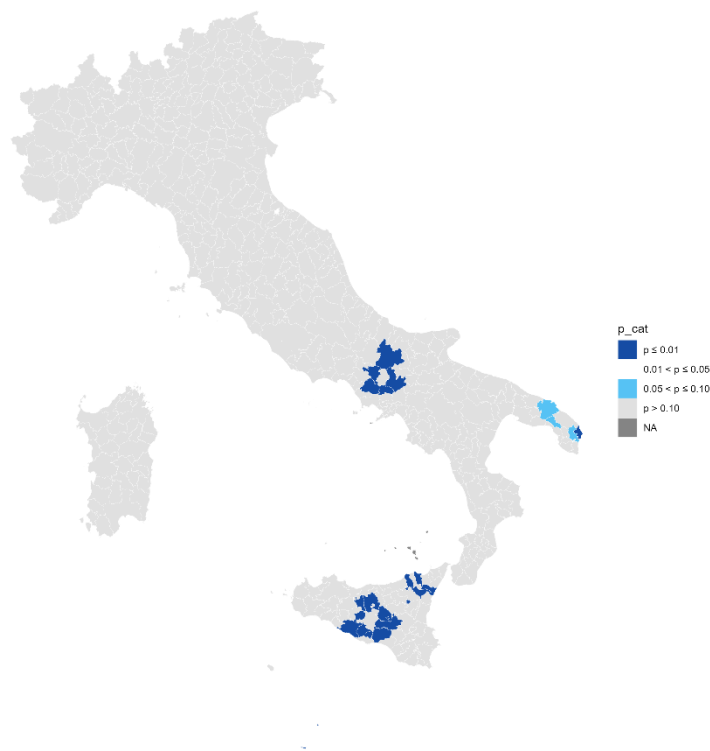


Figure 20: mappa di significatività del Moran locale di *startup_weight* a livello di SLL

In sintesi, *startup_weight_j* non sembra organizzarsi secondo un meccanismo di clustering territoriale diffuso. Ciò che emerge è piuttosto una geografia fatta di pochi nuclei locali (High - High e Low - Low) e di numerosi contrasti di prossimità (Low - High), coerente con la struttura dell'indicatore: molti zeri (assenza di start-up) e pochi valori elevati che non necessariamente trascinano anche i territori contigui.

3.8. Brevetti UIBM: analisi della distribuzione spaziale

L'indicatore UIBM è costruito selezionando le domande di brevetto con anno di deposito 2016 - 2019, dalla banca dati DGPI-UIBM. L'annualità 2020 è esclusa per incompletezza e disomogeneità informativa nelle variabili necessarie alla georeferenziazione e all'armonizzazione territoriale, che renderebbero meno robusta la comparabilità del periodo.

L'attribuzione territoriale avviene alla scala dei SLL 2021 mediante un processo di armonizzazione comunale che ricomponete le variazioni amministrative intervenute nel periodo (fusioni/riorganizzazioni e modifiche anagrafiche), assicurando la riconducibilità dei record comunali 2016 - 2019 alla geografia di riferimento utilizzata per l'aggregazione agli SLL.

A differenza dell'indicatore EPO, per i brevetti UIBM l'allocazione spaziale è basata sul depositante (applicant) e non sull'inventore. La scelta è motivata da due aspetti metodologici:

- In primo luogo, nei dati UIBM la disponibilità e la qualità delle informazioni sugli inventori (identificativi e indirizzi) risultano meno complete e meno uniformi rispetto alle anagrafiche dei depositanti, rendendo più affidabile la geolocalizzazione del brevetto attraverso l'applicant.
- In secondo luogo, l'applicant rappresenta il soggetto che sostiene i costi di deposito, detiene i diritti e tipicamente decide le strategie di protezione dell'innovazione. Ne consegue che l'indirizzo dell'applicant (o la sede legale nel caso di persona giuridica) costituisce una proxy coerente della base organizzativa in cui l'output brevettuale viene formalmente internalizzato.

Operativamente, per persone fisiche si utilizza l'indirizzo dichiarato in domanda, per le imprese si assume la sede legale.

Poiché una domanda può includere più depositanti potenzialmente localizzati in SLL diversi, si adotta un criterio di conteggio frazionario a livello di SLL. Per ciascun brevetto p , sia m_p il numero di SLL distinti associati ai depositanti del brevetto. A ogni SLL coinvolto viene attribuita una quota pari a $\frac{1}{m_p}$.

Il conteggio UIBM attribuito al SLL j è quindi:

$$UIBM_j = \sum_{p \in P} 1(j \in SLL_p) \frac{1}{m_p}$$

dove P è l'insieme delle domande nel periodo considerato, SLL_p è l'insieme degli SLL distinti associati ai depositanti del brevetto p , e $1(\cdot)$ è una variabile indicatrice. Tale procedura garantisce che ogni brevetto contribuisca complessivamente per un'unità alla distribuzione territoriale, evitando sovrastime nei casi di co-deposito inter-territoriale.

Infine, per ottenere una misura comparabile tra territori di diversa dimensione economica, il conteggio frazionario è normalizzato per il numero di addetti del SLL j :

$$UIBM_weight_j = \frac{UIBM_j}{E_j}$$

dove E_j indica gli addetti nel SLL.

3.8.1. Mappe coropletiche

Le mappe coropletiche sono utilizzate per fornire una lettura esplorativa della distribuzione territoriale dei brevetti UIBM, distinguendo tra il numero assoluto di brevetti (num_UIBM_j) e l'intensità relativa rispetto alla dimensione economica locale ($UIBM_weight_j$).

Come per i due casi analizzati in precedenza, la rappresentazione è proposta sia tramite classi per percentili, più adatte a evidenziare la presenza di code estreme e fenomeni di forte concentrazione, sia tramite classi per quantili, che permettono un confronto più omogeneo tra territori. Considerate congiuntamente, queste due scelte di discretizzazione consentono di separare ciò che dipende prevalentemente da effetti di scala da ciò che riflette differenze più strutturali nella propensione alla brevettazione.

(i) Indicatore ***num_UIBM_j*** in valore assoluto per SLL

Nel caso dei conteggi assoluti (num_UIBM), la mappa per percentili (*figura 21*) mette chiaramente in evidenza una distribuzione fortemente asimmetrica: un numero limitato di SLL ricade nelle classi più alte e concentra valori molto superiori rispetto al resto, segnalando una dinamica dominata da pochi poli. Questa evidenza è coerente con un meccanismo di accumulazione legato sia alla maggiore densità di imprese e addetti, sia alla localizzazione di funzioni direzionali e amministrative che, nel caso UIBM, risultano particolarmente rilevanti perché l'allocazione territoriale è basata sul depositante.

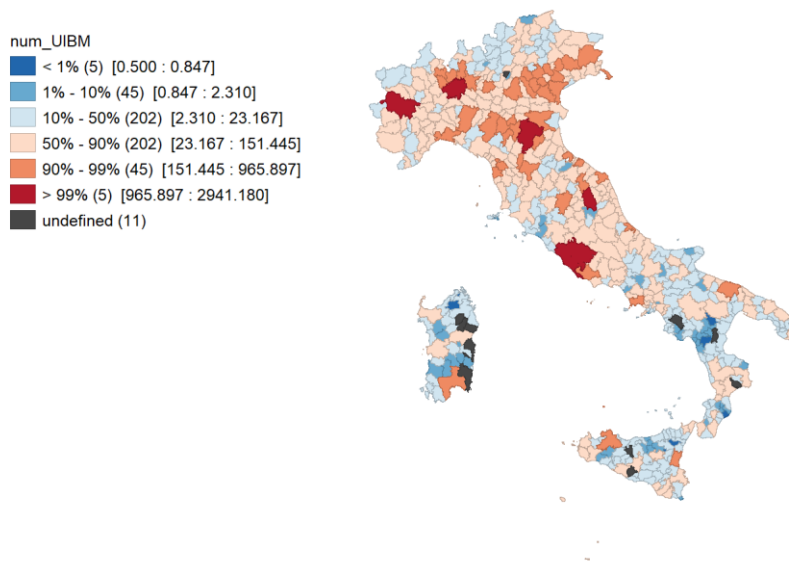


Figure 21: mappa per percentili num_UIBM per SLL

Nella mappa per quantili (*figura 22*) emerge con chiarezza un gradiente territoriale Nord - Sud: gli SLL collocati nelle classi superiori risultano nettamente più frequenti e territorialmente continui nelle regioni settentrionali, mentre nel Mezzogiorno prevalgono classi medio-basse con presenze nelle classi alte più sporadiche e puntuali. Poiché i quantili assegnano a ciascuna classe lo stesso numero di SLL, questa evidenza segnala una differenza strutturale nella distribuzione. A parità di numerosità di unità territoriali, una quota relativamente maggiore di SLL del Nord presenta livelli di deposito UIBM più elevati.

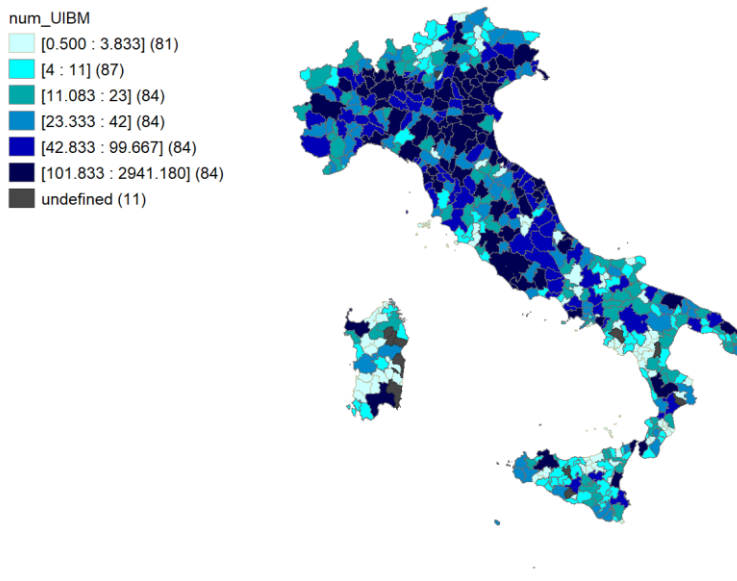


Figure 22: mappa per quantili num_UIBM per SLL

In termini interpretativi, ciò è coerente con la maggiore densità di attività manifatturiere e di imprese innovatrici, con una più elevata dotazione di capitale umano e infrastrutture di supporto e con la localizzazione più concentrata di sedi aziendali e funzioni amministrative (rilevanti nel caso UIBM perché l’allocazione avviene sulla base dell’applicant). Ne deriva che il Nord non ospita soltanto pochi poli eccezionali, ma mostra una diffusione più ampia di sistemi locali con valori relativamente alti anche lungo la parte centrale della distribuzione, mentre nel Sud la brevettazione UIBM appare più polarizzata e meno capillarmente distribuita tra gli SLL.

(ii) Indicatore **UIBM_weight_j**; (#UIBM/addetti per SLL)

La normalizzazione per addetti modifica in modo sostanziale l’interpretazione economica della mappa, poiché riduce l’effetto meccanico della dimensione locale e fornisce una misura della produttività inventiva relativa. Anche qui, la mappa per percentili (*figura 23*) è utile per identificare chiaramente i picchi di intensità (top tail), che non coincidono con quelli osservati nei conteggi: se un territorio è alto in num_UIBM ma non in UIBM_weight_j (indicato come UIBM_core nelle figure 23 e 24) la sua centralità è principalmente trainata dalla scala; viceversa, valori elevati di UIBM_weight_j in territori non dominanti nei conteggi suggeriscono una capacità brevettuale sproporzionata rispetto alla dimensione, dunque un profilo più “intensivo” che “estensivo”.

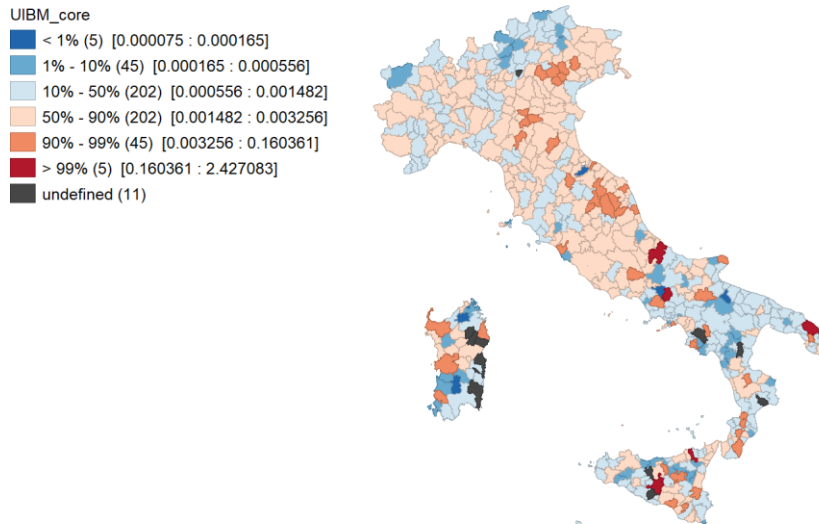


Figure 23: mappa per percentili UIBM_weightj per SLL

La versione in quantili rafforza questa lettura: l'intensità tende a presentare una geografia più frammentata e meno "gerarchica" rispetto ai valori assoluti, perché dipende maggiormente dalla composizione produttiva e dalla concentrazione di pochi soggetti particolarmente brevettanti.

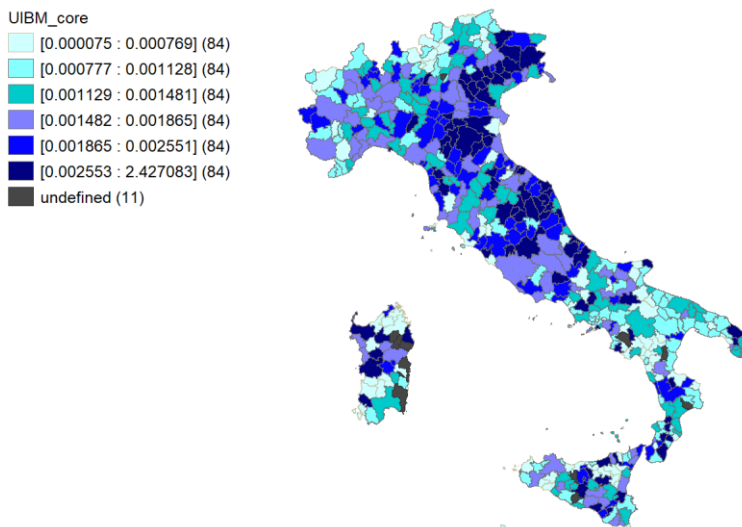


Figure 24: mappa per quantili di UIBM_weightj per SLL

La mappa per quantili (figura 24) suggerisce che l'intensità UIBM non è omogenea, ma presenta una struttura spaziale con macro-aree a maggiore densità di SLL nelle classi alte, in particolare nel Nord, dove si osserva una presenza estesa di unità territoriali nei quantili superiori. Questo risultato è coerente con la compresenza di:

- un tessuto produttivo più orientato a settori a maggiore propensione brevettuale;
- ecosistemi locali in cui la formalizzazione della proprietà industriale è più sistematica anche a parità di base occupazionale.

Rispetto alla mappa dei conteggi, qui l'effetto di scala è attenuato: la permanenza del vantaggio del Nord nella metrica per addetto indica che non si tratta soltanto di più imprese e più addetti, ma di una maggiore intensità media di brevettazione.

Un aspetto rilevante è la maggiore granularità territoriale: rispetto a una concentrazione esclusivamente metropolitana, l'intensità appare distribuita in modo più eterogeneo, con SLL ad alto $UIBM_weight_j$ che possono emergere anche fuori dai grandi poli.

Questa configurazione è compatibile con la presenza di specializzazioni produttive locali (distretti o sistemi locali con elevata propensione alla brevettazione) che, pur non generando necessariamente i massimi livelli assoluti, mostrano performance elevate una volta normalizzate per addetti. In termini interpretativi, l'indicatore intercetta territori nei quali una parte relativamente maggiore dell'attività economica è associata a depositi UIBM, segnalando una combinazione di capacità organizzativa (propensione al deposito) e struttura settoriale favorevole.

Va inoltre considerato un elemento metodologico specifico degli UIBM: l'allocazione è basata sull'applicant (sede legale per persone giuridiche). Nella metrica per addetto questo può produrre due effetti contrapposti. Da un lato, può rafforzare alcuni SLL dove la localizzazione amministrativa delle imprese coincide con la base produttiva e dove l'attività di deposito è effettivamente parte delle pratiche innovative locali; dall'altro, può redistribuire parte del segnale verso SLL che ospitano sedi legali di imprese con stabilimenti altrove. Tuttavia, il fatto che la mappa di $UIBM_weight_j$ continui a mostrare un profilo geografico coerente (e non solo pochi punti anomali) suggerisce che l'intensità non è guidata unicamente da un effetto di headquarter, ma riflette anche differenze territoriali persistenti nella propensione alla protezione brevettuale.

Infine, la presenza di unità undefined va interpretata come assenza di brevettazione.

3.8.2. Distribuzione della funzione $UIBM_weight_j$

$UIBM_weight_j$ mostra una distribuzione estremamente sbilanciata a destra: quasi tutti gli SLL presentano valori molto prossimi allo zero, mentre pochi casi concentrano livelli di intensità nettamente superiori.

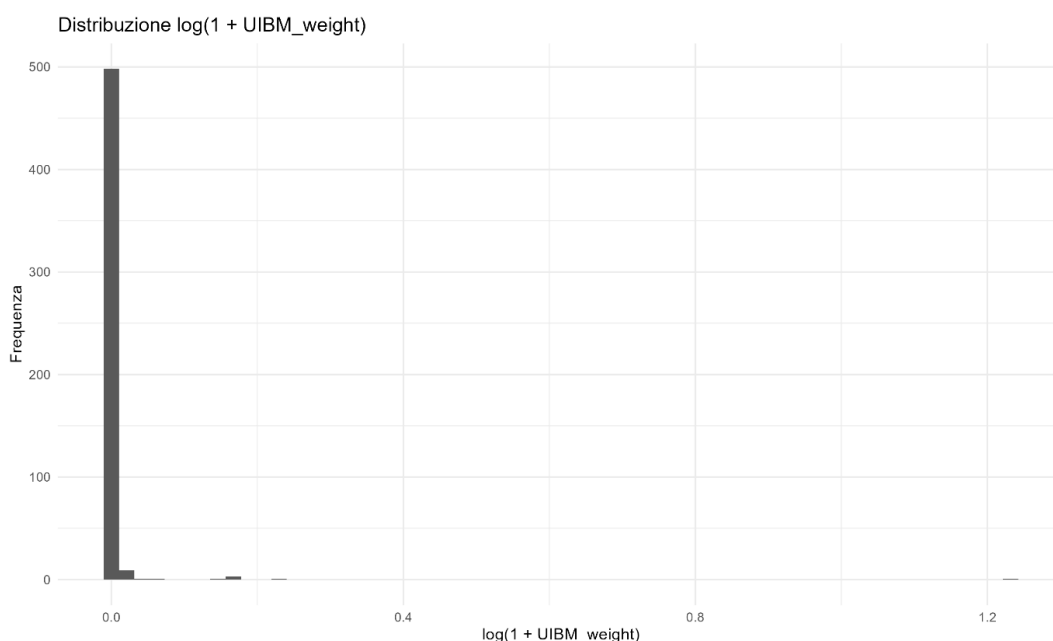


Figure 25: istogramma di $\log(1 + UIBM_weightj)$

Pur con una quota di zeri contenuta (11 SLL su 515, pari a 2,14%), l'istogramma (figura 25) e la tabella 8 mostrano che la frequenza si addensa immediatamente sopra lo zero: la mediana è 0,001457 e il 75° percentile è 0,002088, indicando che per la maggioranza dei territori l'intensità brevettuale rimane entro un intervallo ristretto (IQR 0,001168).

Al contrario, la coda destra è molto pronunciata e trascina la media verso 0,008634: il 99° percentile raggiunge 0,137261 e il massimo è 2,427083, evidenziando outlier estremi. La dispersione risulta quindi elevatissima (SD = 0,1085; CV = 12,566), coerentemente con un indicatore normalizzato per addetti: i valori più alti possono riflettere contesti con propensione al deposito particolarmente intensa e/o una base occupazionale relativamente contenuta.

In ottica Moran - LISA, questa struttura (massa vicino a zero più coda lunga) implica che i risultati possono essere influenzati da pochi casi estremi, l'interpretazione dei cluster deve quindi distinguere tra pattern territoriali sistematici e outlier spaziali.

Table 8: riepilogo delle statistiche descrittive della funzione $UIBM_weightj$

Statistiche descrittive								
N.ro osservazioni	Valore minimo	Valore massimo	Q1	Mediana	Media	Q3	Dev. Standard	Coefficiente variazione
515	0	2,427	0,0009	0,00145	0,0086	0,0021	0.1085	12,566

3.8.3. Moran's I globale e ordini di contiguità

Dai risultati del test di Moran's I su 3 ordini di contiguità (tabella 9) emerge un quadro molto netto: assenza di autocorrelazione spaziale globale.

Table 9: Correlazione spaziale Moran's I globale di $UIBM_weightj$ per diversi ordini di contiguità

	Test analitico – Normale		Test permutazionale – Monte Carlo	
	I	Prob.	I	Prob.
1° ordine di contiguità	-0.0064128	0.7362981	-0.0064128	0.991
2° ordine di contiguità	-0.0041525	0.678590	-0.0041525	0.713
3° ordine di contiguità	-0.0049452	0.7658903	-0.0049452	0.897

I valori di I sono molto piccoli in valore assoluto e non significativi a tutti gli ordini, non si può rigettare l'ipotesi nulla di random spatial pattern (assenza di autocorrelazione globale). La leggera negatività è trascurabile e non stabile (oscilla tra -0.004 e -0.006), quindi non indica una vera "dispersione" sistematica. E' più plausibilmente che sia il risultato dell'effetto combinato di una distribuzione estremamente concentrata vicino a zero e di pochi outlier che assumono configurazioni HL/LH (valori estremi non accompagnati da vicini simili).

Il fatto che anche aumentando l'ordine di contiguità (2° e 3°) Moran resti vicino a zero suggerisce che non c'è un segnale di clustering che emerge a scala più ampia: l'intensità brevettuale sembra guidata da discontinuità e da pochi casi estremi.

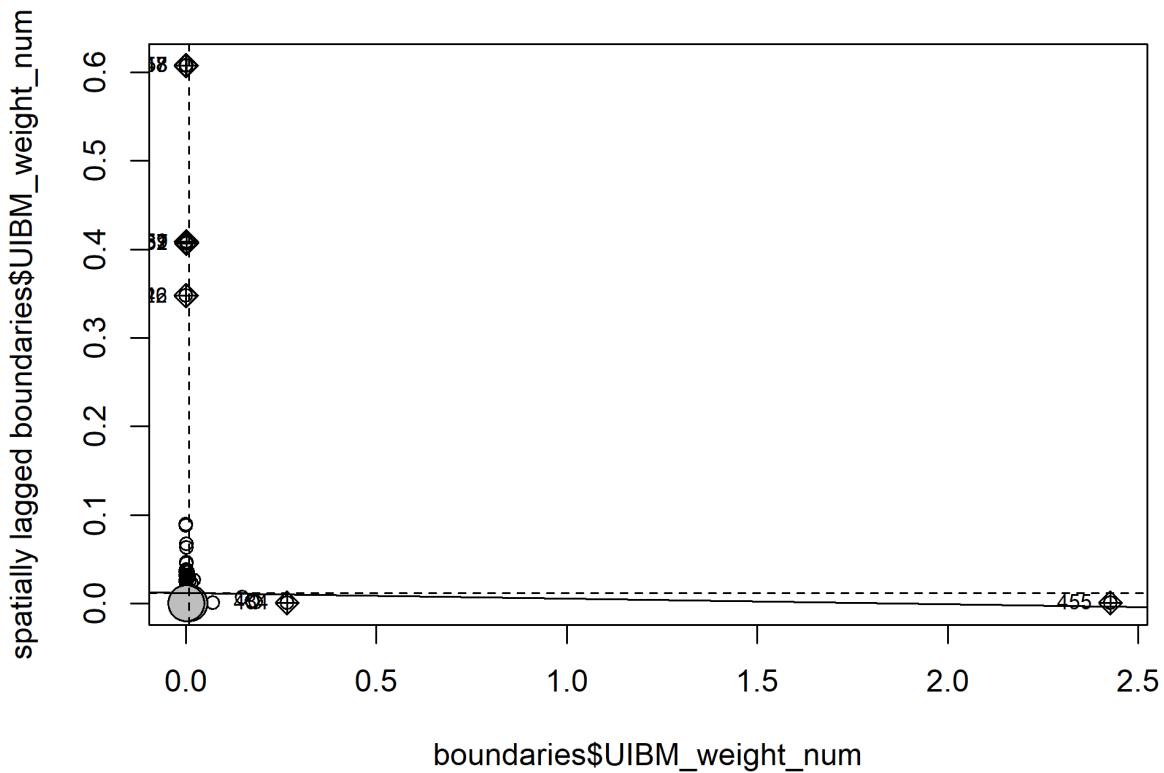


Figure 26: Moran Scatterplot per la funzione $UIBM_weight_j$

Il Moran Scatterplot (figura 26) evidenzia tre aspetti tipici di un indicatore molto asimmetrico con outlier:

1. Massa di osservazioni addossata all'origine: la maggior parte degli SLL ha valori molto bassi e anche i vicini hanno valori bassi, molta nube concentrata vicino a (0,0).
2. Pochi punti "estremi" che rompono la linearità:
 - un outlier con X molto alto (circa 2.4) ma W_y vicino a 0: è il classico profilo High - Low (alto valore isolato in un intorno di bassi), che *spinge* la relazione verso lo zero o verso valori negativi.
 - alcuni punti con X vicino a 0 ma W_y molto alto (asse y molto elevato): sono casi Low - High (basso valore circondato da vicini relativamente alti), anch'essi incompatibili con un clustering "omogeneo" e più coerenti con una struttura fatta di poli e località isolate.
3. Pendenza quasi nulla: la retta di regressione nel Moran Scatterplot è (in forma standardizzata) proporzionale a Moran's I. Visivamente la pendenza è praticamente piatta, ci si aspetta un Moran globale prossimo a zero.

3.8.4. LISA e significatività locale

A livello locale, i LISA cluster (*tabella 10*) mostrano che la struttura spaziale, quando presente, è selettiva e territorialmente circoscritta.

Table 10: classificazione dei cluster secondo l'analisi LISA

Non significant	High - High	Low - Low	Low - High	High - Low	Undefined	Isolated
422 SLL	58 SLL	28 SLL	5 SLL	2 SLL	0 SLL	0 SLL

L'analisi LISA (figura 27) individua 58 SLL classificati come High-High, pari a circa il 11,2% del totale. Tuttavia, tali casi non si organizzano in un'ampia configurazione territoriale continua, risultando piuttosto frammentati e concentrati in un numero limitato di aree. Ne deriva una struttura locale presente ma circoscritta, che non si traduce in un pattern di autocorrelazione spaziale globale robusto

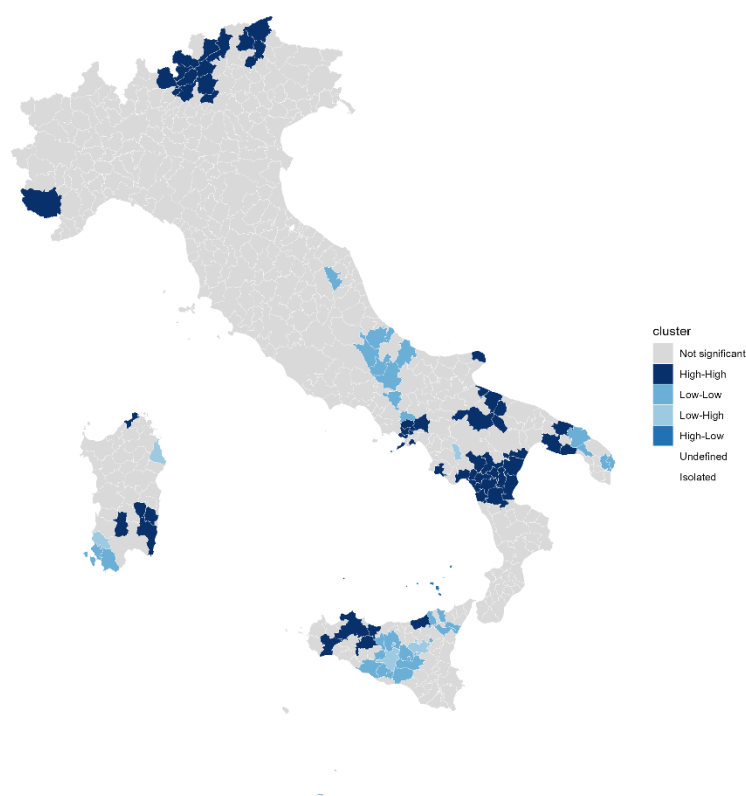


Figure 27: mappa LISA di UIBM_weightj a livello di SLL

Tale evidenza è ulteriormente rafforzata dalla mappa di significatività del Local Moran (*figura 28*), in cui quasi l'intero territorio nazionale ricade nella classe $p > 0.10$, segnalando assenza di pattern locale robusto.

L'unica area chiaramente significativa ($p \leq 0.01$) è localizzata in Sicilia, suggerendo che qui l'intensità brevettuale per addetto non è solo alta in singoli SLL, ma tende a presentare dipendenza locale (valori elevati associati a vicinati relativamente elevati oppure una struttura coerente di outlier locali).

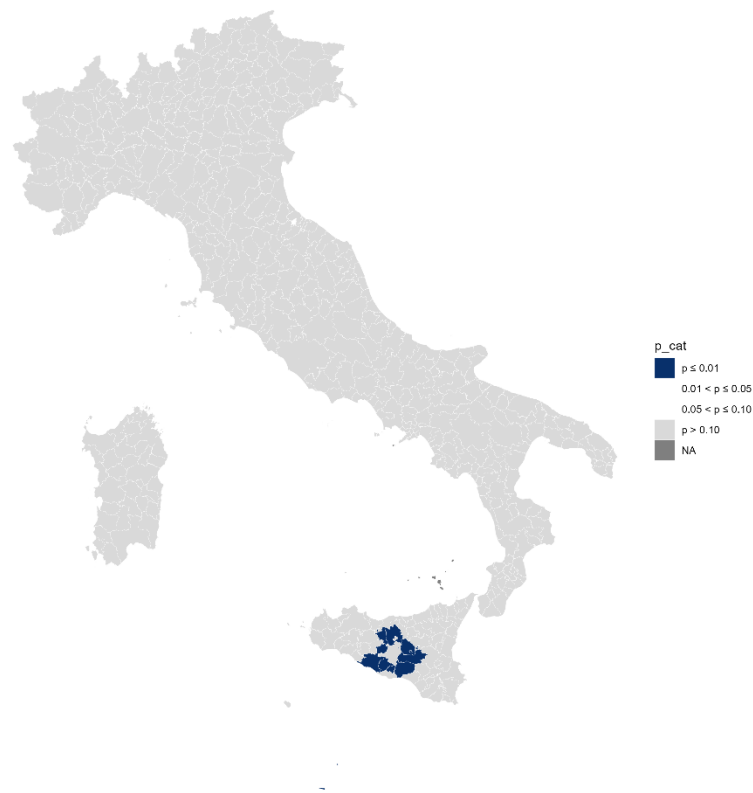


Figure 28: mappa di significatività del Moran locale di $UIBM_weight_j$ a livello di SLL

La geografia dei 20 SLL per valore più elevato di $UIBM_weight_j$ è perfettamente compatibile con questa lettura.

Il massimo assoluto è Caltanissetta ($UIBM_weight \approx 2,427$), seguito da altri SLL meridionali e insulari (tra cui Brolo, Gela, Leonforte, Comiso in Sicilia; Iglesias, Sassari, Siniscola, Oristano in Sardegna; Pizzo e Soriano Calabro in Calabria; Lecce in Puglia; Telesse Terme, Polla e Caserta in Campania).

Molte di queste “eccellenze” risultano spazialmente discontinue (isole ad alta intensità circondate da aree a bassa intensità), cioè più vicine a configurazioni High - Low che a veri cluster estesi.

In sintesi, per $UIBM_weight_j$ l'evidenza empirica suggerisce assenza di autocorrelazione globale e presenza, al più, di micro-strutture locali e di outlier territoriali, con un segnale statisticamente robusto concentrato in una porzione limitata del territorio (in particolare in Sicilia) piuttosto che in un pattern nazionale diffuso.

4. Il modello di regressione OLS

CAPITOLO 4

Questo capitolo definisce l'impianto empirico adottato nell'analisi e ne illustra i principali elementi costitutivi. L'obiettivo è costruire un quadro metodologico coerente che consenta di interpretare i risultati econometrici alla luce delle diverse dimensioni settoriali e territoriali dell'innovazione.

4.1 Il modello empirico

L'adozione di un modello di regressione multivariata risponde alla necessità di sottoporre a verifica empirica le principali ipotesi formulate dalla letteratura sull'economia geografica dell'innovazione, secondo cui la capacità innovativa di un territorio non dipende esclusivamente dalle caratteristiche interne delle imprese, ma è il risultato dell'interazione tra struttura produttiva locale, economie di agglomerazione, varietà settoriale, opportunità tecnologiche e processi di diffusione della conoscenza.

In questa prospettiva, il riferimento metodologico fondamentale è rappresentato dal contributo di Paci e Usai (2000), i quali propongono un modello encompassing volto a stimare congiuntamente l'effetto delle esternalità di specializzazione e di diversità sulla distribuzione spaziale delle attività innovative, nonché a verificare la presenza di spillover di conoscenza tra sistemi locali contigui.

Il presente lavoro si colloca in tale tradizione di ricerca, ma ne estende l'impostazione in funzione di uno specifico obiettivo analitico: utilizza gli SLL 2021 come unità territoriale di analisi, normalizza l'output innovativo rispetto agli addetti anziché alla popolazione, replica la stima su tre diverse proxy dell'innovazione e introduce due variabili aggiuntive specificamente rilevanti per il caso italiano, ossia la presenza di infrastrutture universitarie e una dummy territoriale che isola il Nord rispetto al resto del Paese. In questo senso, il modello impiegato conserva l'impianto teorico di riferimento, ma viene ricalibrato in funzione dell'obiettivo specifico della ricerca, vale a dire individuare le determinanti dell'innovazione alla luce della persistente frattura territoriale italiana.

Coerentemente con tale impostazione, la variabile dipendente è costruita come misura di produttività innovativa della coppia settore-territorio.

Essa esprime il rapporto tra l'output innovativo osservato nel settore i e nel sistema locale del lavoro j e il numero di addetti appartenenti alla medesima cella (i,j) :

$$y_{ij} = \frac{P_{ij}}{E_{ij}}$$

dove:

P_{ij} indica il numero di innovazioni attribuite alla cella settore i e SLL j ;

E_{ij} rappresenta il numero di addetti del settore i nello SLL j ;

L'adozione di questa formulazione consente di depurare l'analisi dagli effetti puramente dimensionali e di confrontare contesti locali e settoriali eterogenei in termini di scala economica. L'interesse dell'analisi non è infatti rivolto alla semplice distribuzione assoluta dell'innovazione, bensì alla sua intensità relativa, ossia alla capacità di un determinato settore, in un determinato territorio, di generare output innovativo in rapporto alla propria base occupazionale.

I settori considerati sono definiti sulla base della classificazione ATECO 2007, adottando un livello di aggregazione a due digit. Tale scelta risponde a un'esigenza di coerenza metodologica, poiché consente di armonizzare fonti informative differenti mantenendo al tempo stesso un grado di dettaglio sufficiente a cogliere le principali differenze intersettoriali. Su questa base, l'analisi individua complessivamente 85 settori.

L'analisi empirica viene quindi articolata in tre specificazioni distinte, corrispondenti a tre diverse proxy dell'output innovativo. Nel primo caso, l'innovazione è misurata attraverso i brevetti depositati presso l'European Patent Office (EPO), nel secondo caso la proxy è costituita dai brevetti depositati presso l'Ufficio Italiano Brevetti e Marchi (UIBM) ed infine nel terzo caso, l'output innovativo è rappresentato dal numero di start-up innovative, che consentono di osservare la traduzione della conoscenza in nuova iniziativa imprenditoriale.

Ne derivano tre variabili dipendenti formalmente omogenee:

$$y_{ij}^{EPO} = \frac{P_{ij}^{EPO}}{E_{ij}}$$

$$y_{ij}^{UIBM} = \frac{P_{ij}^{UIBM}}{E_{ij}}$$

$$y_{ij}^{Start Up} = \frac{P_{ij}^{Start Up}}{E_{ij}}$$

Pur differendo per fonte statistica e contenuto economico, le tre misure condividono la medesima struttura concettuale: in tutti i casi, la variabile dipendente esprime l'output innovativo per addetto nella cella (i,j) . Ciò consente di mantenere costante l'architettura interpretativa dell'analisi e, al tempo stesso, di verificare se i fattori esplicativi considerati incidano in modo uniforme sulle diverse manifestazioni dell'innovazione oppure producano effetti differenziati a seconda della proxy adottata.

Sulla base di tale definizione, il modello di regressione assume, nelle tre applicazioni, una forma generale comune:

$$y_{ij}^m = \alpha + \beta PS_{ij} + \gamma PD_{ij} + \delta SBS_{ij} + \theta SBD_{ij} + \vartheta TO_i^m + \mu_1 PD_j * DM + \mu_2 PD_j * DHT + \rho NORD_j + \sigma UNIV_j + \varepsilon_{ij}$$

dove m identifica di volta in volta la proxy utilizzata dell'output innovativo: *EPO*, *UIBM*, *Start Up*.

In questa formulazione, la produttività innovativa della cella settore-territorio è spiegata come funzione congiunta di caratteristiche del settore, del contesto locale e delle loro interazioni. Più precisamente, il modello include variabili riconducibili alla specializzazione produttiva, alla diversità settoriale, alla prossimità cognitiva tra attività science-based, alle opportunità tecnologiche del comparto, alla metropolitaneità, alla propensione innovativa del settore, alla presenza di infrastrutture universitarie e al dualismo territoriale Nord/resto del Paese.

Nel complesso, il capitolo presenta dunque un impianto empirico che, pur richiamando la tradizione inaugurata da Paci e Usai, viene riformulato in modo coerente con i dati utilizzati e con l'obiettivo specifico della ricerca.

4.2 Definizione e costruzione delle variabili di regressione

Le sezioni che seguono illustrano, in modo analitico, la definizione, il significato teorico e le modalità di costruzione di ciascuna variabile inclusa nel modello.

4.2.1. La specializzazione produttiva PS_{ij}

La prima variabile esplicativa considerata nel modello è l'indice di specializzazione produttiva PS_{ij} , introdotto come misura delle esternalità di specializzazione di tipo Marshall.

L'idea teorica sottostante è che l'attività innovativa di un determinato settore possa risultare più intensa laddove il sistema locale del lavoro presenti una specializzazione produttiva proprio in quel medesimo settore.

In un contesto territorialmente specializzato, la concentrazione di imprese appartenenti alla stessa industria può favorire la circolazione di conoscenze tacite, la formazione di competenze specifiche, la mobilità del lavoro qualificato e l'emergere di meccanismi di apprendimento localizzato.

La formula della variabile è la seguente:

$$PS_{ij} = \frac{\left(\frac{E_{ij}}{\sum_j E_{ij}} \right)}{\left(\frac{\sum_j E_{ij}}{\sum_i \sum_j E_{ij}} \right)}$$

dove E_{ij} rappresenta il numero di addetti del settore i nel sistema locale del lavoro j .

Il numeratore misura il peso del settore i sull'occupazione totale del sistema locale j , mentre il denominatore esprime il peso dello stesso settore sull'occupazione complessiva nazionale.

Ne consegue che valori elevati di PS_{ij} identificano situazioni in cui il settore i presenta, nel territorio j , un'incidenza occupazionale superiore a quella osservata nell'economia nazionale, segnalando quindi una specializzazione relativa del sistema locale in quella particolare attività produttiva.

Ai fini della stima econometrica, l'indice è standardizzato nella forma $(PS - 1)/(PS + 1)$, così da essere vincolato all'intervallo $(-1,1)$.

4.2.2. La diversità produttiva PD_j

L'indice di diversità produttiva del sistema locale PD_j è costruito per cogliere le esternalità di diversità di matrice jacobsoniana, fondate sull'idea che la capacità innovativa di un territorio possa essere favorita

non tanto dalla concentrazione in una sola attività produttiva, quanto dalla compresenza di settori differenti all'interno dello stesso contesto locale.

In questa prospettiva, la varietà industriale accresce le possibilità di interazione tra competenze, conoscenze e tecnologie eterogenee, creando condizioni favorevoli a processi di fertilizzazione incrociata, apprendimento intersettoriale e generazione di nuove combinazioni produttive.

L'indice è definito come il complementare ad 1 del coefficiente di Gini calcolato sulla distribuzione degli addetti tra i settori del sistema locale:

$$PD_j = \left[\frac{2}{(n-1)Q_n} \sum_{i=1}^{n-1} Q_i \right]$$

dove:

- Q_i rappresenta la cumulata degli addetti E fino al settore i , una volta ordinati i settori in senso crescente rispetto al numero di addetti;
- Q_n indica il totale complessivo degli addetti del sistema locale;
- n è il numero dei settori considerati sulla base della classificazione Ateco 2007 che sono 85;

L'indice assume valori compresi nell'intervallo (0,1) e cresce all'aumentare della varietà produttiva del territorio. Poiché il coefficiente di Gini misura la concentrazione, il suo complementare ad 1 fornisce dunque una misura inversa della concentrazione stessa e quindi una proxy della diversificazione del sistema locale.

In termini interpretativi, PD_j consente di verificare se la produttività innovativa osservata nella coppia settore-territorio sia più elevata nei contesti locali caratterizzati da una struttura produttiva maggiormente articolata e differenziata. Tale indicatore non è utilizzato per contrapporre in modo meccanico specializzazione e diversità, ma per cogliere una dimensione specifica della struttura economica locale, distinta da quella misurata dall'indice di specializzazione PS_{ij} . Più precisamente, mentre PS_{ij} rileva il grado di specializzazione del settore i nel sistema locale j , PD_j misura il livello complessivo di varietà produttiva del sistema locale. Ne consegue che le due variabili non vengono trattate come alternative, bensì come caratteristiche differenti e potenzialmente coesistenti del contesto territoriale, entrambe suscettibili di incidere sulla produttività innovativa.

4.2.3. La specializzazione science-based SBS_{ij}

L'indice di specializzazione science-based SBS_{ij} , è introdotto per cogliere una forma più specifica di esternalità di specializzazione. Come illustrato da Feldman e Audretsch (1996), i benefici della prossimità produttiva non si esauriscono necessariamente all'interno del solo settore i , ma possano estendersi anche ai settori complementari che condividono con esso una medesima base scientifica e tecnologica. In questa prospettiva, l'attività innovativa di un settore può essere favorita non soltanto dalla concentrazione locale di imprese appartenenti allo stesso comparto, ma anche dalla presenza, nello stesso sistema locale, di attività produttive affini sotto il profilo cognitivo, poiché tali attività rendono più probabili processi di apprendimento incrociato, circolazione di conoscenze tacite e scambio di competenze specialistiche.

Per identificare tali relazioni di prossimità cognitiva, Feldman e Audretsch (1999) costruiscono una tassonomia delle industrie fondata sulla vicinanza nello spazio tecnologico e cognitivo, individuando sei cluster k .

La classificazione originaria è definita sui codici SIC a quattro digit e deriva dall'utilizzo della Yale Survey of R&D Managers di Levin et al. (1987), che misura la rilevanza relativa di diverse discipline scientifiche di base per l'attività innovativa dei vari comparti produttivi.

Sulla base di tali informazioni, gli autori identificano sei cluster science-based: agri-business, ingegneria Chimica, macchine per ufficio, macchinari industriali, high-tech computing e biomedicale ciascuno dei quali raggruppa industrie accomunate da profili simili di dipendenza dalla conoscenza scientifica, e non semplicemente da affinità merceologiche o produttive.

La *tabella 11* riporta la classificazione proposta dagli autori, evidenziando per ciascun cluster i principali dipartimenti accademici di riferimento e alcune industrie particolarmente rappresentative.

Table 11: classificazione dei settori industriali in cluster k

Agri-business	Chimica (6,06); Scienze Agrarie (4,65); Informatica (4,18); Biologia (4,09)	SIC 2013: Salsicce SIC 2038: Prodotti surgelati SIC 2087: Estratti aromatici SIC 2092: Alimenti confezionati
Ingegneria chimica	Scienza dei Materiali (5,32); Chimica (4,80); Informatica (4,50); Fisica (4,12)	SIC 3861: Apparecchi fotografici SIC 3443: Lavorazioni in lamiera SIC 2821: Materiali plastici SIC 3559: Macchinari industriali speciali
Macchine per ufficio	Informatica (6,75); Scienze Mediche (5,75); Matematica (5,49); Matematica Applicata (4,64)	SIC 3576: Bilance e strumenti di misura SIC 3579: Macchine per ufficio SIC 3535: Nastri trasportatori SIC 2751: Stampa commerciale
Macchinari industriali	Scienza dei Materiali (5,03); Informatica (4,76); Fisica (3,94); Chimica (3,88)	SIC 3551: Attrezzature per la lavorazione alimentare SIC 3523: Macchinari SIC 3546: Utensili manuali SIC 3629: Apparecchiature industriali
High-tech computing	Scienza dei Materiali (5,92); Informatica (5,63); Fisica (5,45); Matematica (4,76)	SIC 3573: Macchine per il calcolo SIC 3662: Apparecchi radio/TV SIC 3823: Strumenti di controllo di processo SIC 3674: Semiconduttori
Biomedicale	Chimica (5,53); Scienze Mediche (5,47); Informatica (5,32); Scienza dei Materiali (5,02)	SIC 3842: Apparecchi chirurgici SIC 3841: Strumenti medicali SIC 2834: Prodotti farmaceutici SIC 3811: Strumenti scientifici

In altri termini, il criterio di aggregazione non è settoriale in senso descrittivo, bensì cognitivo-tecnologico: settori differenti possono essere inclusi nel medesimo cluster quando condividono una base scientifica sottostante analoga. Nel presente lavoro, tale classificazione è stata ricondotta ai codici Ateco a 2 digit, al fine di renderla compatibile con la struttura dei dati utilizzati nell'analisi empirica. Questa riclassificazione non modifica l'impostazione concettuale originaria, ma ne costituisce un adattamento al diverso sistema di classificazione settoriale adottato, preservando l'idea metodologica di fondo secondo cui gli spillover rilevanti possono emergere tra attività produttive distinte ma appartenenti a una medesima *science base*. La *tabella 12* mostra la dimensione dei cluster individuati

Table 12: distribuzione dei settori industriali (Ateco a 2 digit) nei cluster k

Cluster k	Nome del cluster k	N.ro codici Ateco
1	Agri – business	3
2	Ingegneria chimica	7
3	Macchine per l'ufficio	2
4	Macchinari industriali	2
5	High-tech computing	6
6	Biomedicale	1

Coerentemente con tale impostazione, SBS_{ij} misura il grado di specializzazione del sistema locale del lavoro j nelle industrie complementari al settore i , vale a dire nell'insieme dei settori che condividono con esso la medesima *science base* (nello stesso cluster k). La variabile è definita come:

$$SBS_{ij} = \frac{\left(\frac{E_{ij}^k}{\sum_i E_{ij}} \right)}{\left(\frac{\sum_j E_{ij}^k}{\sum_i \sum_j E_{ij}} \right)}$$

dove E_{ij}^k indica gli addetti nelle industrie appartenenti al medesimo cluster science-based k del settore i , al netto degli addetti del settore i . In forma esplicita:

$$E_{ij}^k = \sum_{h \in k(i)} E_{hj} - E_{ij}$$

con h indice dei settori appartenenti al cluster $k(i)$. Inoltre, $\sum_i E_{ij}$ rappresenta il totale degli addetti nel sistema locale del lavoro j , mentre $\sum_i \sum_j E_{ij}$ indica il totale nazionale degli addetti.

Valori elevati di SBS_{ij} segnalano quindi una concentrazione relativamente alta, nel sistema locale j , di attività appartenenti alla medesima base cognitiva e tecnologica del settore i , configurando un contesto potenzialmente favorevole alla generazione di spillover di conoscenza.

4.2.4. La diversità science-based SBD_{ij}

Il modello include anche l'indice di diversità science-based SBD_{ij} , volto a cogliere il grado di varietà interna al cluster k cui appartiene il settore i .

L'idea teorica sottostante è che gli spillover di conoscenza non dipendano soltanto dalla presenza locale di attività produttive affini, ma anche dal livello di differenziazione che caratterizza tali attività all'interno di una comune base scientifica. In questa prospettiva, non rileva solo che in un sistema locale siano presenti settori riconducibili alla medesima *science base*, ma anche che tali settori risultino sufficientemente articolati e diversificati, così da favorire processi di fertilizzazione incrociata, combinazione di competenze e circolazione di conoscenze complementari.

Seguendo l'impostazione di Paci e Usai (2000), SBD_{ij} è quindi costruito come misura della diversità occupazionale all'interno del cluster science-based associato al settore i , ed è basato sul reciproco del coefficiente di Gini calcolato sugli addetti dei settori che compongono il cluster k . In tal modo, la

variabile consente di valutare se l'attività innovativa risulti più elevata nei contesti locali in cui la comune base scientifica non si traduce in una concentrazione ristretta in pochi comparti, ma in una configurazione produttiva più ampia e differenziata.

SBD_{ij} misura dunque il grado di diversità interna al cluster science-based $k(i)$ nel sistema locale del lavoro j , ossia il livello di articolazione dei settori che condividono con il settore i la medesima *science base*. L'indice è definito come:

$$SBD_{ij} = \left[\frac{2}{(n_k - 1)Q_{kn}} \sum_{i=1}^{n_k-1} Q_{ki} \right]$$

dove:

- Q_{ki} rappresenta la cumulata degli addetti nei settori appartenenti al cluster k , ordinati in senso crescente rispetto al numero di addetti;
- Q_{kn} indica il totale degli addetti del cluster k nel sistema locale j ;
- n_k è il numero dei settori appartenenti a quel cluster presenti nel sistema locale considerato;

Ne consegue che valori più elevati di SBD_{ij} identificano contesti in cui l'occupazione del cluster science-based risulta distribuita in modo più equilibrato tra più settori, segnalando quindi una maggiore varietà interna della base cognitiva locale.

Anche questa variabile è costruita mantenendo la medesima logica metodologica proposta in letteratura, ma riportando la classificazione dei cluster al livello dei codici Ateco a 2 digit, così da renderla coerente con la struttura settoriale adottata nell'analisi empirica.

In sintesi, mentre SBS_{ij} coglie la concentrazione del sistema locale nelle attività appartenenti alla stessa base scientifica del settore i , SBD_{ij} ne misura invece il grado di varietà interna.

4.2.5. Le opportunità tecnologiche settoriali TO_i

Tra le variabili di controllo del modello rientra l'indicatore di opportunità tecnologiche settoriali TO_i che misura il volume complessivo di attività innovativa osservata nel settore i a livello nazionale e può quindi essere interpretato come una proxy della quantità di conoscenza specifica già disponibile in quel settore, suscettibile di ulteriori sviluppi cumulativi.

In altri termini, consente di tenere conto del fatto che i settori non si collocano tutti sul medesimo piano in termini di fertilità tecnologica: alcuni comparti operano infatti in traiettorie scientifiche e produttive intrinsecamente più dinamiche e presentano quindi maggiori possibilità di generare nuova innovazione indipendentemente dalle caratteristiche del contesto locale. Coerentemente con questa interpretazione, nel modello ci si attende un coefficiente positivo, poiché una più elevata dotazione di conoscenza specifica settoriale dovrebbe favorire l'ulteriore produzione di output innovativo.

Formalmente, la variabile è definita come:

$$TO_i = \sum_j P_{ij}$$

dove P_{ij} rappresenta l'output innovativo attribuito al settore i nel SLL j .

Dal momento che l'analisi empirica è replicata utilizzando tre differenti proxy dell'innovazione: brevetti EPO, brevetti UIBM e start-up innovative anche TO_i è costruito in tre versioni corrispondenti, coerenti con la diversa natura delle fonti utilizzate.

4.2.5.a Brevetti EPO

I brevetti EPO sono osservati, da un lato, con una localizzazione territoriale ricostruibile attraverso l'indirizzo dell'inventore, e dall'altro con una classificazione tecnologica in classi IPC4.

Poiché l'IPC è una classificazione tecnologica e non industriale, i brevetti non sono immediatamente riconducibili ai settori produttivi utilizzati nel modello. Per questa ragione, ciascun brevetto è stato ricondotto ai settori industriali mediante una tabella di concordanza IPC4 - NACE Rev. 2, coerente con ATECO 2007, così da rendere omogenea la classificazione settoriale dei brevetti rispetto a quella impiegata per i dati occupazionali.

L'allocazione è stata effettuata con criterio frazionario lungo più dimensioni. Sul piano territoriale, come visto nel capitolo 3, ogni brevetto è stato ripartito tra gli SLL distinti associati agli inventori, attribuendo a ciascun territorio una quota proporzionale all'inverso del numero di SLL coinvolti.

In secondo luogo, sul piano tecnologico, nei casi in cui una stessa domanda riporti più classi IPC4, il peso del brevetto è stato ulteriormente suddiviso tra le diverse classi.

In terzo luogo, poiché una stessa classe IPC4 può rinviare a più settori NACE Rev. 2, anche la quota già attribuita è stata ulteriormente ripartita tra i settori compatibili secondo la concordanza adottata.

L'adozione del criterio fractional permette di evitare che le corrispondenze multiple tra classi tecnologiche e settori determinino un sovradimensionamento artificiale del contributo dei brevetti multi-classe o multi-mappati. Tale procedura consente di preservare l'unitarietà del peso complessivo di ciascun brevetto lungo l'intera allocazione

Il conteggio P_{ij}^{EPO} così ottenuto rappresenta quindi il numero frazionario di brevetti EPO attribuibili al settore i nel SLL j , e la corrispondente variabile di opportunità tecnologiche è definita come:

$$TO_i^{EPO} = \sum_j P_{ij}^{EPO}$$

In questo modo, TO_i^{EPO} misura il volume complessivo di conoscenza brevettuale EPO riferibile al settore i a livello nazionale.

L'ancoraggio territoriale all'indirizzo dell'inventore è coerente con la scelta metodologica adottata nel Capitolo 3, volta a cogliere il luogo effettivo di produzione dell'attività inventiva. Inoltre, il passaggio dalla classificazione IPC alla classificazione industriale riprende la logica, già presente in Paci e Usai, di ricondurre i brevetti originariamente classificati per tecnologia ai corrispondenti settori di manifattura mediante una procedura di attribuzione proporzionale.

E' essenziale osservare che la procedura di raccordo tra la classificazione tecnologica IPC4 e la classificazione settoriale NACE Rev. 2 a due cifre non produce una copertura completa (*Tabella 13*). In particolare, 9 codici IPC4 risultano privi di corrispondenza nella concordanza adottata e interessano complessivamente 344 brevetti su 15.764 brevetti EPO.

Table 13: IPC4 non mappati per assente corrispondenza con NACE Rev.2

IPC4	N.ro brevetti
B29L	100
B29K	68
F24S	24
G16H	20
C12R	11
F24T	2
F24V	2
G16B	1
G16C	1

Le classi non mappate appartengono a ambiti tecnologici tra loro eterogenei e non riconducibili a un unico insieme settoriale coerente. Per tale ragione, si è esclusa l'introduzione di una categoria residuale ZZ, in quanto l'aggregazione dei codici unmatched in un'unica classe avrebbe determinato una rappresentazione settoriale artificiale e metodologicamente poco fondata.

La strategia adottata è stata pertanto selettiva. Sono stati esclusi esclusivamente i 22 brevetti associati unicamente a codici IPC4 non mappati, poiché per tali osservazioni non era possibile effettuare alcuna attribuzione settoriale. Al contrario, i brevetti caratterizzati dalla presenza congiunta di codici IPC4 mappati e non mappati sono stati mantenuti nel campione, così da preservare l'informazione disponibile. Ne consegue una riduzione del numero complessivo di brevetti da 15.764 a 15.742 unità.

Nel calcolo dei pesi fractional, la quota del singolo brevetto è stata inizialmente ripartita sull'insieme completo dei codici IPC4 ad esso associati, inclusi quelli non mappati. In una fase successiva, tuttavia, ai fini della costruzione della matrice settore-territorio, sono state considerate esclusivamente le componenti corrispondenti a codici IPC4 per i quali fosse disponibile una mappatura verso la NACE a due cifre.

4.2.5.b Brevetti UIBM

Per i brevetti UIBM, la costruzione della variabile segue una logica analoga sul piano settoriale, ma differisce sul piano territoriale. Anche in questo caso, infatti, i brevetti sono inizialmente osservati con classificazione IPC4 e vengono successivamente ricondotti ai settori industriali mediante la concordanza IPC4 - NACE Rev. 2. Tuttavia, diversamente dai brevetti EPO, l'allocazione territoriale non è basata sull'inventore, bensì sul depositante (applicant) o, nel caso di persone giuridiche, sulla sede legale.

Anche in questo caso si adotta un criterio frazionario: se una domanda coinvolge più depositanti localizzati in SLL diversi, il brevetto viene ripartito tra tali SLL; se una domanda riporta più classi IPC4, il peso viene suddiviso tra le classi, infine se una stessa classe IPC4 corrisponde a più settori NACE, la quota viene ulteriormente ripartita tra i settori compatibili.

Si ottiene così un conteggio frazionario P_{ij}^{UIBM} per ciascuna cella settore - territorio, dal quale deriva:

$$TO_i^{UIBM} = \sum_j P_{ij}^{UIBM}$$

Anche in questo caso, la variabile misura il numero complessivo di brevetti UIBM attribuibili al settore i sull'intero territorio nazionale, fornendo una proxy delle opportunità tecnologiche specifiche del comparto.

Per i brevetti UIBM, la concordanza incompleta tra IPC4 e Nace Rev.2 comporta la perdita di 57 osservazioni su 19.082 brevetti distinti, dunque una quota del tutto trascurabile.

4.2.5.c Start-up innovative

Nel caso delle start-up innovative, la costruzione di $TO_i^{Start Up}$ è invece più diretta. Le imprese sono infatti già osservate con una classificazione settoriale coerente con ATECO e con una localizzazione territoriale univoca, ricostruita sulla base della sede legale e successivamente attribuita allo SLL di appartenenza. A differenza dei brevetti, non è quindi necessaria alcuna procedura di concordanza tecnologico-settoriale né alcun conteggio frazionario multi-territoriale o multi-classe. L'output innovativo nella cella settore - territorio coincide semplicemente con il numero di start-up innovative appartenenti al settore i e localizzate nello SLL j, che possiamo indicare con S_{ij} . La corrispondente misura di opportunità tecnologiche settoriali è pertanto:

$$TO_i^{Start Up} = \sum_j S_{ij}$$

ossia il numero complessivo di start-up innovative osservate nel settore i a livello nazionale. In questo caso, TO_i può essere interpretato come una misura della vivacità innovativa-imprenditoriale del comparto, vale a dire della sua capacità di generare nuove iniziative imprenditoriali innovative su scala nazionale.

In sintesi, pur essendo costruita a partire da basi informative differenti, la logica della variabile resta la stessa in tutte e tre le applicazioni: TO_i è una variabile di settore, comune a tutti i territori, che serve a controllare per l'eterogeneità tecnologica intrinseca dei comparti. In tal modo, il modello evita di attribuire alle sole esternalità territoriali differenze nell'output innovativo che dipendono invece, almeno in parte, dal fatto che alcuni settori presentano opportunità tecnologiche più elevate di altri.

4.2.6. Le aree metropolitane DM_j

La variabile DM_j è introdotta per cogliere l'effetto delle aree metropolitane sulla capacità innovativa dei sistemi locali del lavoro. L'ipotesi sottostante è che le principali realtà urbane possano beneficiare di economie di agglomerazione più intense, di una maggiore densità di interazioni economiche e cognitive e di una più elevata disponibilità di servizi avanzati, fattori potenzialmente favorevoli alla produzione di innovazione (Ron A. BOSCHMA 2005). DM_j è stata costruita a partire dalla classificazione dei sistemi locali del lavoro proposta dall'Istat nel volume *Forme, livelli e dinamiche dell'urbanizzazione in Italia*, riferita agli SLL 2011 e successivamente aggiornata nella revisione 2018. Tale classificazione distingue tre categorie: principali realtà urbane, città medie e altri sistemi locali del lavoro.

Le principali realtà urbane comprendono gli SLL dei comuni capoluogo delle città metropolitane, nonché i sistemi locali di grandi dimensioni, con oltre 500 mila abitanti e quelli con un comune capoluogo di almeno 200 mila abitanti.

La classificazione proposta da Istat è stata armonizzata con la definizione dei SLL del 2021 e ricondotta a una dummy dicotomica, definita come:

$$DM_j = \begin{cases} 1 & \text{se il SLL } j \text{ è stato classificato come 'Principale realtà urbana'} \\ 0 & \text{altrimenti} \end{cases}$$

In tal modo, la variabile non misura genericamente il grado di urbanizzazione del territorio, ma isola specificamente i sistemi locali che presentano, secondo la classificazione Istat, il più elevato livello di centralità urbana e metropolitana. Operativamente, la classificazione è stata dapprima attribuita agli SLL della base originaria e successivamente armonizzata alla geografia territoriale utilizzata nell'analisi empirica, così da mantenere coerenza tra la costruzione della variabile territoriale e la partizione finale degli SLL.

DM_j non distingue l'urbano dal non urbano in senso lato, ma separa le principali realtà urbane da tutti gli altri sistemi locali (*figura 29*).

Nel modello, DM_j è introdotta in interazione con PD_j , per verificare se l'effetto della diversità produttiva sia più marcato nelle principali realtà urbane. L'ipotesi è infatti che i contesti metropolitani, per effetto della maggiore densità di relazioni e della più elevata prossimità tra attività differenti, possano amplificare i benefici innovativi associati alla varietà settoriale.

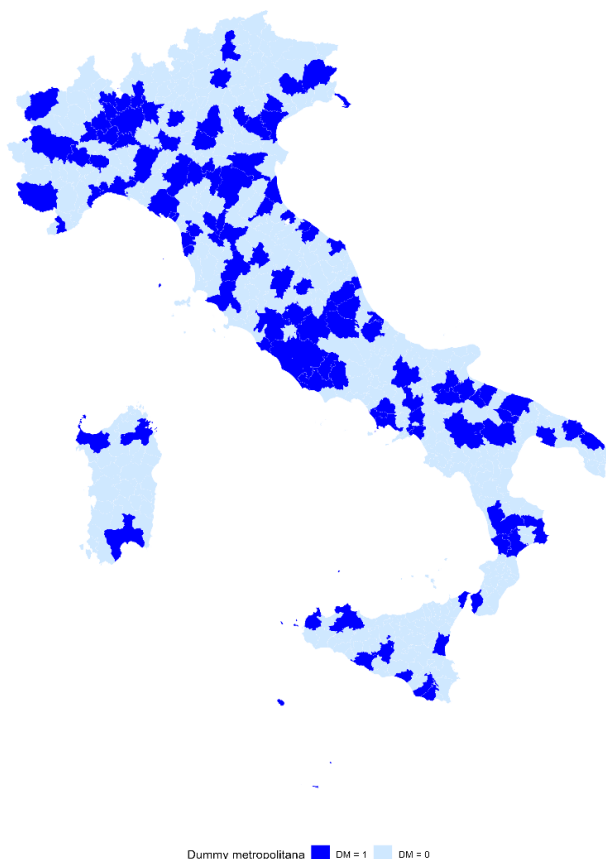


Figure 29: mappatura degli SLL per i quali la $DM_j = 1$ dunque classificabili come aree metropolitane

4.2.7. I settori a più elevata propensione innovativa DHT_i

La variabile DHT_i è stata introdotta per individuare i settori caratterizzati da una più elevata propensione all'innovazione. La sua costruzione si basa sulla 'Rilevazione sull'innovazione nelle imprese. Anni 2018-2020' dell'Istat, indagine che raccoglie informazioni sui processi innovativi delle imprese dell'industria e dei servizi nel triennio di riferimento. In tale fonte, l'innovazione è osservata con riferimento alle principali tipologie di innovazione introdotte dalle imprese, tra cui l'innovazione di prodotto e l'innovazione di processo, nell'ambito della più ampia rilevazione CIS coordinata a livello europeo.

Nel presente lavoro, la dummy è costruita a partire dalla quota di imprese innovatrici calcolata per ciascun settore, considerando le imprese che hanno portato a termine processi di innovazione di prodotto e di processo. La classificazione è stata ricondotta al livello dei codici Ateco a 2 digit, così da renderla coerente con l'unità settoriale adottata nell'intera analisi empirica. In particolare, è stato attribuito valore pari a 1 ai soli settori per i quali tale quota risulta superiore al 40 per cento, e valore pari a 0 in tutti gli altri casi. La variabile è quindi definita come:

$$DHT_i = \begin{cases} 1 & \text{se la quota di imprese innovatrici nel settore } i \text{ è superiore al } 40\% \\ 0 & \text{altrimenti} \end{cases}$$

In termini interpretativi, DHT_i consente di distinguere i comparti con una struttura innovativa relativamente più intensa da quelli caratterizzati da una minore propensione all'innovazione, introducendo nel modello una misura sintetica delle differenze intersettoriali nella capacità di generare e assorbire nuove conoscenze. La soglia del 40% per cento permette infatti di isolare i settori che, nel periodo considerato, mostrano una diffusione dell'innovazione sistematicamente più elevata rispetto agli altri comparti dell'industria e dei servizi.

Diversamente da TO_i , che misura il volume complessivo delle opportunità tecnologiche del settore, DHT_i ne coglie invece il profilo qualitativo, distinguendo i comparti a più elevata intensità innovativa.

Nel modello, la variabile è considerata in interazione con l'indice di diversità produttiva PD_j , al fine di verificare se l'effetto della varietà settoriale locale risulti diverso nei comparti caratterizzati da una più elevata propensione all'innovazione. Tale interazione risponde all'ipotesi che i benefici della diversificazione produttiva non siano uniformi tra i settori, ma possano manifestarsi con maggiore intensità proprio in quelli a più alto contenuto innovativo, nei quali la presenza di conoscenze eterogenee e complementarità intersettoriali può tradursi più facilmente in nuovi processi di combinazione tecnologica.

4.2.8. La presenza di infrastrutture universitarie $UNIV_j$

La variabile $UNIV_j$ è introdotta per cogliere il possibile ruolo della presenza universitaria nella capacità innovativa dei sistemi locali del lavoro. L'ipotesi sottostante è che le università, in quanto sedi di produzione e trasmissione di conoscenza, possano favorire i processi innovativi locali attraverso la formazione di capitale umano qualificato, la diffusione di conoscenze codificate e tacite e la creazione di interazioni con il tessuto produttivo circostante (Anselin, L, Varga, A. and Acs, Z.J. (1997)).

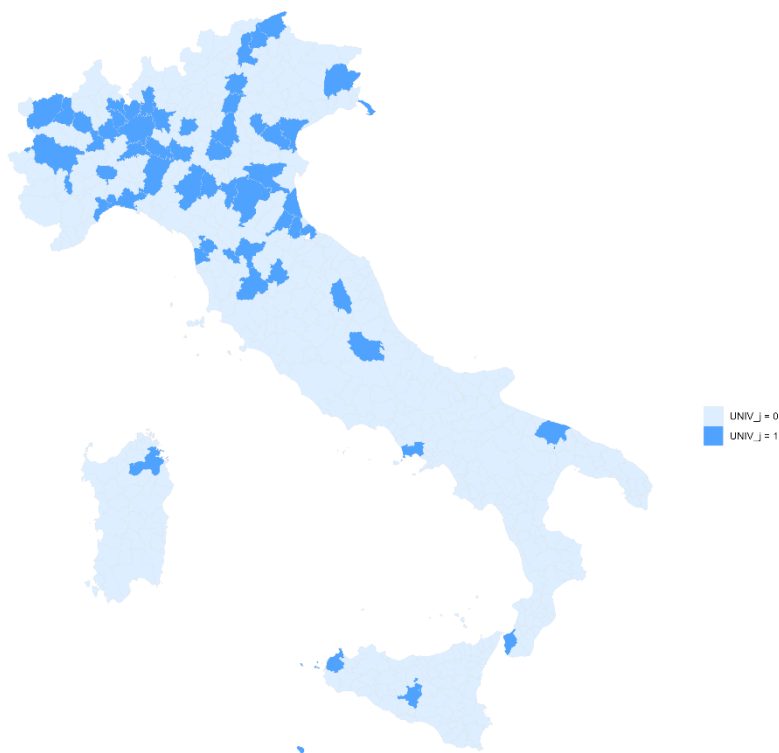


Figure 30: mappatura della presenza universitaria nei SLL

Nel presente lavoro, $UNIV_j$ è costruita come variabile dummy, assumendo valore pari a 1 se nel sistema locale del lavoro j è presente almeno una sede universitaria, e valore pari a 0 in caso contrario (Figura 30).

Ai fini della costruzione della variabile, sono state considerate sia le sedi principali sia le sedi decentrate o secondarie degli atenei, includendo le università statali, le università non statali legalmente riconosciute e gli istituti ad ordinamento speciale.

Le informazioni utilizzate per identificare la presenza delle sedi universitarie sono state reperite direttamente dalle fonti ufficiali del Ministero dell'Università e della Ricerca. La variabile assume pertanto la seguente forma:

$$UNIV_j = \begin{cases} 1 & \text{se nel SLL } j \text{ è presente almeno un'università riconosciuta} \\ 0 & \text{altrimenti} \end{cases}$$

L'inclusione della variabile $UNIV_j$ costituisce un'estensione del modello originario di Paci e Usai, introdotta nel presente lavoro al fine di cogliere in modo più esplicito la dimensione *knowledge-based* dello sviluppo innovativo territoriale. In questa prospettiva, la dummy $UNIV_j$ consente di verificare se, a parità delle altre condizioni, i territori caratterizzati dalla presenza di almeno un'istituzione universitaria presentino livelli più elevati di produttività innovativa.

4.2.9. La presenza di infrastrutture universitarie $NORD_j$

L'ultima variabile introdotta nel modello è la dummy territoriale $NORD_j$, finalizzata a cogliere in modo esplicito la presenza di un differenziale geografico nella capacità innovativa dei sistemi locali del lavoro. La sua inclusione risponde a uno degli obiettivi centrali del presente lavoro, vale a dire verificare se, a parità delle altre condizioni considerate nel modello, permanga un divario territoriale riconducibile alla storica frattura tra Nord e Sud del Paese. In questa prospettiva, la variabile consente di valutare se l'innovazione presenti una distribuzione sistematicamente differenziata lungo la direttrice territoriale italiana, oltre quanto già spiegato dalla struttura produttiva, dalle economie di agglomerazione, dalle opportunità tecnologiche e dalla presenza di infrastrutture knowledge-based.

Operativamente, la dummy è definita come:

$$NORD_j = \begin{cases} 1 & \text{se lo SLL } j \text{ appartiene al cluster Nord} \\ 0 & \text{se lo SLL } j \text{ appartiene al cluster Sud} \end{cases}$$

La tabella 14 riporta nel dettaglio la classificazione degli SLL tra i due gruppi territoriali.

Table 14: classificazione degli SLL nei due cluster Nord - Sud

Cluster	Regioni	N.ro regioni
Nord	Piemonte, Valle d'Aosta, Lombardia, Trentino-Alto Adige, Veneto, Friuli-Venezia Giulia, Liguria, Emilia-Romagna	8
Sud	Toscana, Lazio, Abruzzo, Molise, Umbria, Campania, Basilicata, Marche, Puglia, Calabria, Sardegna, Sicilia	12

In particolare, al cluster Nord appartengono gli SLL localizzati nelle regioni dell'Italia settentrionale, mentre nel cluster Sud sono stati ricompresi non soltanto i sistemi locali del Mezzogiorno e delle Isole, ma anche quelli appartenenti alle regioni del Centro.

Tale scelta non deriva da un intento meramente geografico - descrittivo, bensì da un'esigenza analitica: l'obiettivo è isolare la specificità del Nord rispetto al resto del Paese, verificando se esso presenti un profilo innovativo differenziale in termini strutturali. In questo senso, la contrapposizione adottata non coincide perfettamente con la tradizionale ripartizione geografica in Nord, Centro e Mezzogiorno, ma assume la forma di una dicotomia tra Nord e Centro-Sud, funzionale a mettere alla prova l'esistenza di un vantaggio territoriale sistematico dell'Italia settentrionale nei processi di innovazione.

In termini interpretativi, $NORD_j$ non misura quindi una semplice appartenenza geografica, ma introduce nel modello una proxy sintetica del dualismo territoriale italiano. Un coefficiente statisticamente significativo associato a tale variabile indicherebbe che, al netto degli altri fattori considerati, la collocazione nel Nord del Paese continua a rappresentare un elemento rilevante nella spiegazione della produttività innovativa dei sistemi locali del lavoro.

La dummy $NORD_j$ è dunque introdotta per verificare se il divario innovativo territoriale italiano rifletta una componente strutturale riconducibile alla contrapposizione tra Nord e resto del Paese.

4.3 Conclusioni

L'impianto empirico così definito costituisce quindi la base analitica attraverso cui indagare le determinanti dell'innovazione nel contesto italiano, mettendo in relazione la dimensione settoriale con quella territoriale e verificando il ruolo delle diverse forme di esternalità richiamate dalla letteratura. La costruzione delle variabili e la specificazione dei tre modelli consentono infatti di osservare il fenomeno innovativo secondo prospettive complementari, ma metodologicamente omogenee.

Muovendo da questo quadro, il capitolo seguente sarà dedicato all'analisi dei risultati delle regressioni, così da valutare la capacità esplicativa del modello e il grado di coerenza tra le evidenze empiriche ottenute e le ipotesi teoriche discusse nei capitoli precedenti.

5. Analisi dei risultati econometrici

CAPITOLO 5

Nel presente capitolo sono illustrati e discussi i risultati delle stime econometriche relative ai tre modelli considerati, costruiti rispettivamente sulla base dei brevetti EPO, dei brevetti UIBM e delle start-up innovative. Prima di procedere all'interpretazione dei coefficienti stimati, è tuttavia necessario chiarire la struttura del campione effettivamente utilizzato nelle regressioni, successivamente si analizzano i risultati delle tre specificazioni, valutandone sia il contenuto economico sia i principali aspetti diagnostici.

5.1 Dataset di regressione

Prima di discutere i risultati della regressione, è necessario chiarire la struttura del campione effettivamente utilizzato nelle stime. Poiché l'analisi è impostata su una doppia dimensione, settoriale i e territoriale j , l'unità di osservazione non coincide né con il solo settore né con il solo territorio, ma con la coppia settore-territorio (i,j) . Ne consegue che l'ampiezza del campione dipende sia dalla copertura delle due dimensioni considerate separatamente, sia dalla loro effettiva intersezione. Per questo motivo, prima di interpretare i coefficienti stimati, è opportuno verificare la consistenza del dataset di partenza e chiarire in quale misura la numerosità finale derivi dalla struttura stessa dei dati.

PS_{ij} e PD_{ij}

Con riferimento alle variabili costruite sugli addetti, e in particolare a PS_{ij} e PD_{ij} , la base informativa territoriale risulta completa. Gli SLL distinti presenti nel dataset sono 515, il totale degli addetti osservati è pari a 19.946.950 e rappresenta la base occupazionale su cui vengono costruite le variabili esplicative legate alla struttura produttiva locale.

I settori distinti osservati sono 85, corrispondenti all'intero insieme dei settori NACE a 2 digit considerati. Tuttavia, la piena copertura delle due dimensioni non implica che ogni settore sia presente in ogni territorio.

Le combinazioni settore-territorio effettivamente osservate sono infatti 35.713, un numero inferiore a quello teorico ottenibile dal prodotto tra 85 settori e 515 SLL. Questo scarto non segnala un problema del dataset, ma riflette la struttura reale del sistema produttivo: non tutti i territori presentano addetti in tutti i settori. La matrice (i,j) da cui parte la regressione è quindi incompleta per ragioni economiche e non per errori di costruzione del campione.

Di conseguenza, una parte della riduzione del numero di osservazioni va interpretata come un tratto fisiologico dei dati, legato all'assenza empirica di alcune combinazioni settore-territorio, e non come una perdita informativa generata dalle successive operazioni di merge o dalla costruzione delle variabili.

SBS_{ij} e SBD_{ij}

SBS_{ij} e SBD_{ij} sono definite solo per i settori appartenenti ai cluster $k = 1, \dots, 6$. Per questa ragione, una parte consistente delle 35.713 coppie (i,j) iniziali resta esclusa per costruzione: 27.707 osservazioni

appartengono infatti a settori non riconducibili a nessuno dei $cluster_k$ individuati e non rientrano nel dominio teorico di questi indicatori. Restano quindi 8.006 coppie potenzialmente eleggibili.

Nel caso di SBD_{ij} , la variabile risulta effettivamente osservata in 7.596 casi. I 410 valori mancanti residui non dipendono da problemi del dataset, ma dalla definizione stessa dell'indicatore: quando, all'interno di uno stesso SLL e cluster, è presente un solo settore osservato ($n_k = 1$), la diversificazione interna non è calcolabile. Di conseguenza, SBD_{ij} è disponibile nel 94,9% delle osservazioni teoricamente eleggibili, ma soltanto nel 21,3% del totale delle coppie (i, j). La riduzione del campione non va quindi interpretata come una perdita anomala di dati, bensì come una conseguenza diretta della natura selettiva della variabile.

Per SBS_{ij} , la riduzione del numero di osservazioni dipende interamente dal dominio teorico della variabile. Su un totale di 35.713 coppie settore-territorio (i, j), SBS_{ij} risulta infatti definita in 8.006 casi e non definita nelle restanti 27.707 osservazioni. La variabile è osservata nel 22,4% del totale delle coppie (i, j), ma nel 100% delle osservazioni teoricamente eleggibili. A differenza di quanto avviene per SBD_{ij} , quindi, nel caso di SBS_{ij} , una volta ristretto il campione ai soli settori appartenenti ai cluster considerati, la copertura della variabile è completa. Questo implica che la selezione del campione associata a SBS_{ij} è interamente determinata dalla scelta teorica di limitare l'analisi ai settori clusterizzati, e non da problemi informativi residui nella costruzione dell'indicatore.

$DM_j, DHT_i, UNIV_j, NORD_j$

Le variabili di controllo $DM_j, DHT_i, UNIV_j$ e $NORD_j$ non esercitano un effetto selettivo sul campione. $UNIV_j$ e $NORD_j$ sono variabili definite a livello territoriale e risultano attribuite all'intero insieme degli SLL considerati. Esse entrano quindi nel modello per tutte le coppie (i, j), assumendo valore pari a 0 o 1 a seconda della condizione identificata.

Anche DM_j è definita a livello territoriale, mentre DHT_i è definita a livello settoriale. Nel modello queste due variabili rilevano attraverso il loro termine di interazione. Ciò non modifica la logica del campione: il valore 0 non comporta l'esclusione dell'osservazione, ma segnala soltanto l'assenza della caratteristica considerata. Di conseguenza, anche nel caso di DM_j e DHT_i , così come per la loro interazione, le coppie settore - territorio restano nel dataset indipendentemente dal fatto che la variabile assuma valore 0 o 1. L'unico caso in cui tali variabili potrebbero ridurre il numero di osservazioni sarebbe la presenza di valori mancanti, non il loro contenuto informativo e non è questo il caso.

TO_i

Nella costruzione di TO_i quando riferita ai brevetti (EPO ed UIBM) si riscontra una perdita di informazione che non dipende dall'aggregazione finale sulla dimensione settoriale in sé, ma dalla fase precedente di costruzione della matrice P_{ij} , da cui TO_i deriva per somma sulla dimensione territoriale.

In particolare, il problema emerge quando l'unità di analisi originaria è classificata secondo una tassonomia tecnologica e deve quindi essere ricondotta ai settori produttivi mediante una procedura di armonizzazione con la NACE Rev. 2. In questi casi, la conversione dalla classificazione tecnologica a quella settoriale non è perfetta e una parte dell'informazione si perde poiché 9 codici IPC4 non trovano corrispondenza nella concordanza adottata.

Table 15: brevetti persi nella costruzione di $P_{i,j}$ a causa della mappatura IPC4 - NACE Rev. 2

	#Brevetti di partenza	#Brevetti esclusi	#Brevetti su cui è definita TO_i
Brevetti EPO	15.764	22	15.742
Brevetti UIBM	19.082	57	19.025

Questo fenomeno riguarda specificamente le variabili costruite a partire dai brevetti, e quindi i brevetti EPO e UIBM, dal momento che in entrambi i casi la classificazione originaria è di tipo tecnologico e deve essere successivamente tradotta in termini settoriali. La *tabella 15* mostra il numero di osservazioni escluse dal campione per effetto dell'armonizzazione.

Diverso è invece il caso delle startup, qui non si osserva alcuna perdita informativa riconducibile alla costruzione di TO_i , poiché le unità sono già classificate direttamente secondo codici Ateco e non richiedono alcun passaggio di armonizzazione tra classificazioni differenti.

$$y_{ij}^{EPO}, y_{ij}^{UIBM}, y_{ij}^{Start Up}$$

La costruzione della variabile dipendente richiede, in tutti e tre i casi considerati, l'allineamento tra la matrice dell'attività innovativa e la base occupazionale settore-territorio. L'unità finale di osservazione resta infatti la coppia (i, j) , per cui la variabile dipendente può essere definita solo laddove l'informazione sull'evento innovativo risulti associabile a un settore i e ad un territorio j per i quali sia disponibile anche la corrispondente cella degli addetti. Ne consegue che il campione effettivamente utilizzato nella regressione non coincide con il numero iniziale di brevetti o startup osservati, ma con il sottoinsieme di osservazioni che, una volta ricondotte alla dimensione settoriale e territoriale, trovano effettiva corrispondenza nella matrice degli addetti. In questo senso, la restrizione del campione dipende dall'intersezione tra due basi informative distinte: da un lato la struttura produttiva locale, dall'altro la distribuzione settore-territorio del fenomeno innovativo considerato.

La *tabella 16* quantifica la restrizione del campione su cui sono definite le variabili dipendenti dei tre modelli di regressione

Table 16: numero di osservazioni (i,j) su cui sono definite le variabili dipendenti

Proxy dell'innovazione	Coppie (i,j) con output innovativo osservato	Coppie (i,j) senza match con la base addetti	Coppie (i,j) con output innovativo e addetti disponibili
Brevetti EPO	3371	973	2398
Brevetti UIBM	3998	262	3736
Start Up Innovative	719	0	719

Nel caso delle start-up innovative, il valore 719 non coincide con il numero finale di osservazioni utilizzate nella regressione, ma con il numero di coppie settore-territorio in cui si osserva almeno una start-up. Il campione del modello è invece più ampio, poiché include anche le coppie (i, j) nelle quali la variabile dipendente assume valore nullo. Di conseguenza, le 6.580 osservazioni della regressione comprendono sia celle positive sia celle con assenza di start-up innovative.

Come già visto nel paragrafo precedente, nel caso dei brevetti, una prima riduzione può già emergere nella fase di costruzione della matrice innovativa, poiché l'attribuzione ai settori produttivi richiede il raccordo tra classificazioni tecnologiche e classificazioni settoriali, con conseguente possibile perdita di copertura. Nel caso delle startup, invece, non vi è alcuna perdita dovuta alla classificazione settoriale né alla corrispondenza con la base degli addetti. Pertanto, le eventuali riduzioni del numero di osservazioni rilevanti per la regressione vanno lette come il risultato del progressivo allineamento tra le diverse fonti informative necessarie alla definizione della variabile dipendente, più che come una semplice perdita di dati in senso stretto.

5.2 Modello di regressione OLS per i brevetti EPO

Nel presente paragrafo si analizzano i risultati delle regressioni relative ai brevetti EPO, preceduti da un esame preliminare della distribuzione delle variabili impiegate nel modello e della struttura delle correlazioni tra i regressori. Questo passaggio è utile non solo per inquadrare le principali caratteristiche del campione, ma anche per interpretare in modo più consapevole le stime econometriche che verranno discusse di seguito.

5.2.1 Le variabili della regressione: statistiche descrittive e correlazioni

La *figura 31* riporta le statistiche descrittive delle variabili utilizzate nel modello EPO. Il primo elemento che emerge riguarda l'elevata eterogeneità della variabile dipendente.

Variabile	Tipo di variabile	Descrizione	Osservazioni	Media	Dev. Std	Min	Max
y_{ij}^{EPO}	Variabile dipendente	Produttività dell'innovazione (EPO) del settore i nel SLL j	2398	0,1151	0,7073	3,00E-06	20,75
PS_{ij}	Variabile indipendente	Specializzazione produttiva del settore i nel SLL j	2398	-0,1773	0,4729	-0,997	0,9643
PD_j	Variabile indipendente	Diversità produttiva nel SLL j	2793	0,2619	0,0527	0,1074	0,3852
TO_{ij}^{EPO}	Variabile indipendente	Numero di brevetti del settore i nel SLL j	3371	12279	12337	11783	42860
SBS_{ij}	Variabile di controllo	Specializzazione science - based del settore i nel SLL j	1598	0,6146	0,5255	-1	0,9911
SBD_{ij}	Variabile di controllo	Diversità science - based del settore i nel SLL j	1512	0,3968	0,1843	0,0023	1
$UNIV_j$	Variabile di controllo	Presenza di università nel SLL j	2793	0,1754	0,3804	0	1
$NORD_j$	Variabile di controllo	Appartenenza del SLL j al Nord Italia	2793	0,5825	0,4932	0	1
DM_j	Variabile di controllo	Classificazione del SLL j in area metropolitana	2793	0,2574	0,4373	0	1
DHT_i	Variabile di controllo	Classificazione di i come settore ad alta propensione innovativa	3197	0,9621	0,1908	0	1

Figure 31: statistiche descrittive delle variabili utilizzate nella regressione OLS dei brevetti EPO

La produttività dell'innovazione misurata tramite i brevetti EPO presenta una media pari a 0.1151, una deviazione standard di 0.7073 e un valore massimo molto elevato a fronte di un minimo prossimo allo zero. Si tratta di una distribuzione fortemente asimmetrica, coerente con la natura selettiva e fortemente concentrata dell'attività brevettuale EPO: la maggior parte delle combinazioni settore-territorio registra valori molto bassi di intensità innovativa, mentre un numero ristretto di casi concentra livelli assai più elevati di output brevettuale. In altri termini, l'innovazione misurata attraverso i brevetti europei appare distribuita in modo fortemente diseguale nello spazio economico considerato.

Un secondo aspetto rilevante concerne il diverso numero di osservazioni disponibili per le singole variabili. Mentre per la maggior parte dei regressori territoriali il numero di osservazioni si colloca intorno a 2.793 unità, le variabili SBS_{ij} e SBD_{ij} risultano disponibili solo per un sottoinsieme più limitato di coppie settore-territorio, pari rispettivamente a 1.598 e 1.512 osservazioni.

Ciò conferma quanto già discusso in precedenza: l'inclusione di tali variabili nelle specificazioni più ricche comporta una riduzione del campione stimato e questo elemento va tenuto presente nel confronto tra modelli, poiché differenze nei coefficienti possono dipendere non solo dalla struttura del modello ma anche dal diverso insieme di osservazioni utilizzate.

Anche sul lato delle variabili esplicative emergono alcune informazioni significative. PS_{ij} che misura la specializzazione produttiva del settore i nel sistema locale j , presenta valori sia negativi sia positivi ed evidenzia una dispersione relativamente ampia. Ciò segnala una marcata eterogeneità nei livelli di specializzazione locale. PD_j che cattura la diversità produttiva del sistema locale, mostra invece una variabilità più contenuta, suggerendo che la distribuzione di questa caratteristica territoriale sia più concentrata.

La variabile TO_i^{EPO} che misura le opportunità tecnologiche settoriali, si colloca su una scala numerica molto diversa rispetto agli altri regressori: la media è pari a 12.279 e la deviazione standard a 12.337. Questo implica che il coefficiente associato a tale variabile non sia direttamente confrontabile, in termini di ampiezza numerica, con quelli delle altre covariate, ma debba essere interpretato tenendo conto della specifica unità di misura e della natura settoriale dell'indicatore.

Le variabili dummy forniscono inoltre alcune indicazioni utili sulla composizione del campione. Il 58,3% delle osservazioni appartiene a sistemi locali del Nord Italia, il 25,7% ricade in aree metropolitane e il 17,5% in sistemi locali con presenza universitaria. Particolarmente elevata è la quota di osservazioni per cui DHT_i assume valore unitario: con una media pari a 0,9621, la variabile segnala che il campione EPO è fortemente concentrato nei settori ad alta propensione innovativa. Questo dato è coerente con la natura stessa del brevetto europeo, che tende a rappresentare le attività inventive più strutturate, codificabili e tecnologicamente avanzate.

Accanto alle statistiche descrittive, la matrice di correlazione (figura 32) consente una prima valutazione della struttura delle relazioni lineari tra i regressori e fornisce indicazioni preliminari sull'eventuale presenza di problemi di multicollinearità.

variabile	PS_ij	PD_j	TO_i_EPO	SBS_ij	SBD_ij	DHT_i	DM_j	UNIV_j	NORD_j
PS_ij	1	0.2706	-0.1131	0.0681	-0.1667	-0.0122	0.0083	0.0488	0.1828
PD_j	0.2706	1	-0.1052	0.0931	0.0434	-0.0599	0.4906	0.4838	0.2779
TO_i_EPO	-0.1131	-0.1052	1	-0.0356	0.2818	0.1935	-0.0261	-0.0391	-0.0916
SBS_ij	0.0681	0.0931	-0.0356	1	0.2365	-0.057	-0.0329	-0.0242	0.0396
SBD_ij	-0.1667	0.0434	0.2818	0.2365	1	0.0058	0.0718	0.0722	-0.0965
DHT_i	-0.0122	-0.0599	0.1935	-0.057	0.0058	1	-0.017	-0.0249	-0.0291
DM_j	0.0083	0.4906	-0.0261	-0.0329	0.0718	-0.017	1	0.6305	0.0268
UNIV_j	0.0488	0.4838	-0.0391	-0.0242	0.0722	-0.0249	0.6305	1	0.2931
NORD_j	0.1828	0.2779	-0.0916	0.0396	-0.0965	-0.0291	0.0268	0.2931	1

Figure 32: matrice di correlazione tra le variabili del modello di regressione OLS per i brevetti EPO

Nel complesso, i coefficienti di correlazione risultano per lo più contenuti e non emergono associazioni bivariate così elevate da far presumere seri problemi di collinearità tra le variabili esplicative. La correlazione tra specializzazione produttiva e diversità produttiva è positiva ma moderata (0,2706), segnalando che le due dimensioni non sono sovrapponibili empiricamente: i territori più diversificati non coincidono necessariamente con quelli più specializzati nei singoli settori. Anche le correlazioni che coinvolgono TO_i^{EPO} restano relativamente basse, a conferma del fatto che le opportunità tecnologiche settoriali colgono una dimensione distinta rispetto alle caratteristiche territoriali locali.

Le associazioni più elevate riguardano soprattutto alcune variabili di controllo territoriali. In particolare, la correlazione tra aree metropolitane e presenza universitaria è la più alta dell'intera matrice, mentre PD_j risulta correlata positivamente sia con DM_j sia con $UNIV_j$. Questi valori suggeriscono che i sistemi locali più diversificati tendono più frequentemente a coincidere con contesti metropolitani e universitari, il che appare economicamente plausibile: le aree urbane più grandi e dotate di infrastrutture conoscitive presentano in genere una struttura produttiva più articolata. Poiché tali correlazioni restano inferiori ai livelli che solitamente destano maggiore preoccupazione, esse sembrano indicare più una parziale sovrapposizione tra caratteristiche dei territori che non un vero problema di multicollinearità.

Nel complesso, l'analisi preliminare suggerisce dunque due considerazioni utili per la lettura delle regressioni.

- (i) Da un lato, la forte asimmetria della variabile dipendente conferma che l'innovazione brevettuale EPO è un fenomeno estremamente concentrato e selettivo;
- (ii) Dall'altro, la struttura delle correlazioni tra regressori mostra una certa prossimità tra alcune dimensioni territoriali, ma non tale da compromettere in modo evidente l'identificazione dei singoli effetti.

Su queste basi è quindi possibile procedere all'interpretazione dei risultati econometrici, prestando particolare attenzione al ruolo della specializzazione, della diversità produttiva e delle principali caratteristiche di contesto che connotano i sistemi locali osservati.

5.2.2. Risultati della regressione per i brevetti EPO

Si esaminano a questo punto i risultati dell'analisi econometrica dei brevetti EPO ottenuti a partire da 6 diversi modelli di regressione. L'evidenza delle analisi restituisce un risultato molto chiaro: la specializzazione produttiva presenta in tutte le specificazioni un coefficiente negativo e statisticamente significativo, mentre la diversità produttiva non mostra un effetto positivo stabile e generalizzabile.

	(1)	(2)	(3)	(4)	(5)	(6)
Specializzazione produttiva	-0,317*** (0,058)	-0,2337*** (0,0298)	-0,3055*** (0,0586)	-0,2127*** (0,0263)	-0,2245*** (0,0261)	-0,2044*** (0,0241)
Diversità produttiva	0,1315 (0,5123)	-0,2660 (0,2341)	-0,1682 (0,3599)	-1,8460** (0,7546)	-0,0649* (0,2801)	-0,4945 (1,5876)
Opportunità tecnologiche settoriali	-2,93e-06 (8,2134e-06)	1,64e-05*** (3,56e-06)	1,27e-05*** (3,8e-06)		1,3870e-05 (4,1391e-06)	
Appartenenza al Nord Italia	-0,0248 (0,0416)	-0,0037 (0,0244)	0,0226 (0,1033)	-0,0013 (0,0241)		
Presenza di università	0,0035 (0,0242)	0,0225 (0,0189)	0,0188 (0,0171)	0,0231 (0,0188)	0,0180*** (0,0205)	0,0239 (0,02003)
Diversità produttiva * Aree metropolitane	-0,3188** (0,1639)	-0,2036** (0,0728)	-0,1995** (0,0741)	-0,1988** (0,0718)	-0,1664** (0,0708)	-0,1867** (0,0722)
Diversità produttiva * Settori high tech	0,1204 (0,1282)	-0,1237 (0,1271)	-0,1952 (0,1331)	1,4805** (0,7440)	0,0957 (0,2074)	-0,0182 (1,5452)
Specializzazione science-based		0,0307 (0,0275)	0,0256 (0,0276)	-0,0316 (0,0595)	0,0161 (0,0292)	-0,0585 (0,0625)
Diversità science-based		-0,1430 (0,0702)	-0,1381* (0,070)	-0,1203* (0,0714)	-0,1149 (0,0723)	-0,0916 (0,0741)
Appartenza al Nord Italia * Diversità produttiva			0,0032 (0,3723)			
Appartenza al Nord Italia * Specializzazione produttiva			0,1136** (0,0636)			
Dummy Settore	NO	NO	NO	SI	NO	SI
Dummy Regione	NO	NO	NO	NO	SI	SI
Intercetta	0,041 (0,119)	0,1772*** (0,0493)	0,1553* (0,0838)	0,1438** (0,0578)	0,2185*** (0,0674)	0,1759*** (0,0648)
Osservazioni	1487	2252	1509	1509	1509	1509
R-quadro corretto	0,1013	0,0439	0,1028	0,1106	0,1227	0,1332

Nota: errori standard robusti tra parentesi (*p < 0,10 **p < 0,05 ***p < 0,001). La variabile dipendente è il numero di brevetti EPO per settore i e SLL j normalizzato rispetto al numero di addetti del settore i nel SLL j

Figure 33: regressione OLS per i brevetti EPO

Tra le diverse specificazioni, il *modello (2)* può essere assunto come riferimento principale dell'analisi. Esso rappresenta infatti il miglior compromesso tra completezza informativa e interpretabilità economica dei coefficienti. Da un lato si nutre di un numero di osservazioni più elevato rispetto ai modelli che introducono interazioni territoriali o effetti fissi più stringenti, dall'altro consente di includere esplicitamente le variabili SBS_{ij} e SBD_{ij} , che sono particolarmente rilevanti rispetto all'obiettivo della tesi perché permettono di verificare se la dimensione science-based del tessuto produttivo locale eserciti un ruolo specifico sull'attività brevettuale. Questa informazione andrebbe invece persa, o quantomeno resa meno leggibile, nelle specificazioni con dummy settoriali, dove parte dell'eterogeneità tecnologica viene assorbita dagli effetti fissi.

Il modello è globalmente significativo, ma spiega una quota relativamente contenuta della variabilità della dipendente. In tal senso, è plausibile che una parte importante della variabilità dipenda da fattori idiosincratici, tecnologici e organizzativi che non sono interamente catturati dagli indici di struttura produttiva locale.

Il grafico in *figura 34* mostra che il *modello 2* intercetta solo in parte la variabilità della dipendente: la maggior parte delle osservazioni è compressa vicino a zero, mentre pochi casi presentano valori molto elevati e restano lontani dalla retta di adattamento. Questo è coerente con la natura selettiva e fortemente concentrata dell'output EPO, che come illustrato nel capitolo 3 è descritto come un fenomeno polarizzato e con coda destra pronunciata. In questo senso, il grafico riflette una proprietà sostanziale del fenomeno, la distribuzione è fortemente right-skewed e non organizzata in grandi cluster spaziali continui, ma in episodi localizzati e discontinui

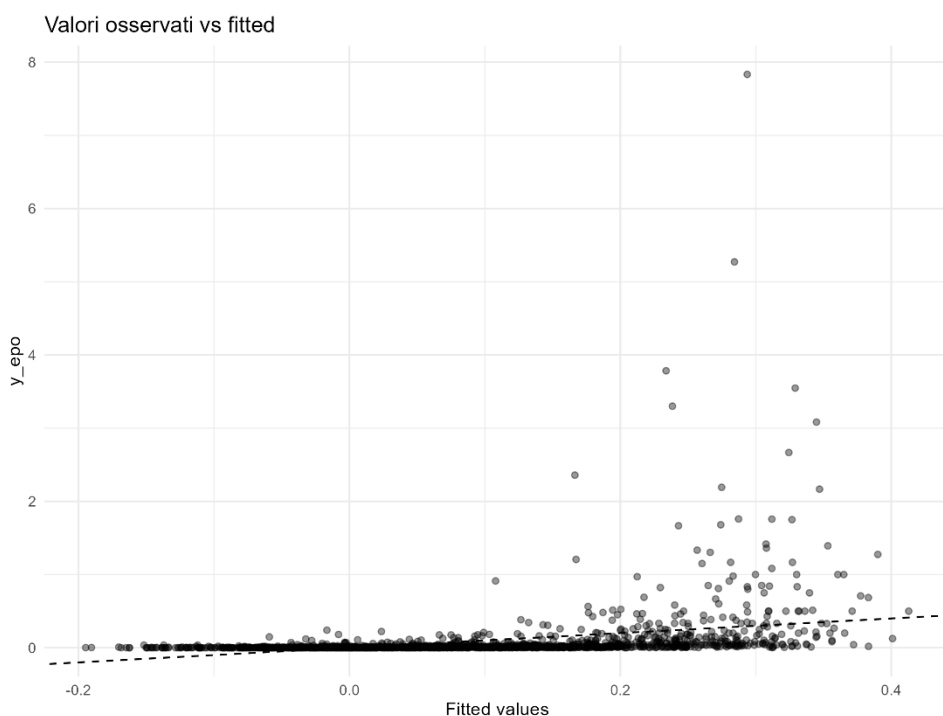


Figure 34: grafico di confronto tra valori osservati e valori stimati, modello 2

I test diagnostici (tabella 17) segnalano che l'OLS lineare va interpretato con cautela.

Table 17: risultati dei test diagnostici condotti sul modello di regressione OLS (2) sui brevetti EPO

Test diagnostico	Statistica	p-value
Breusch – Pagan	19,08931	0,0244
RESET	42,26701	1,2682e-20

Il Breusch - Pagan è significativo, quindi l'ipotesi di omoschedasticità è respinta, per questo motivo la lettura delle relazioni tra le variabili è basata sugli errori standard robusti HC1.

Inoltre, il RESET è fortemente significativo, il che suggerisce una probabile miss - specificazione della forma funzionale: il modello lineare non esaurisce bene la relazione tra variabili esplicative e intensità brevettuale.

Il grafico in figura 35 mostra infatti una struttura non casuale dei residui, la dispersione cresce al crescere dei valori stimati e la linea smooth non resta piatta intorno a zero, ma assume una curvatura evidente.

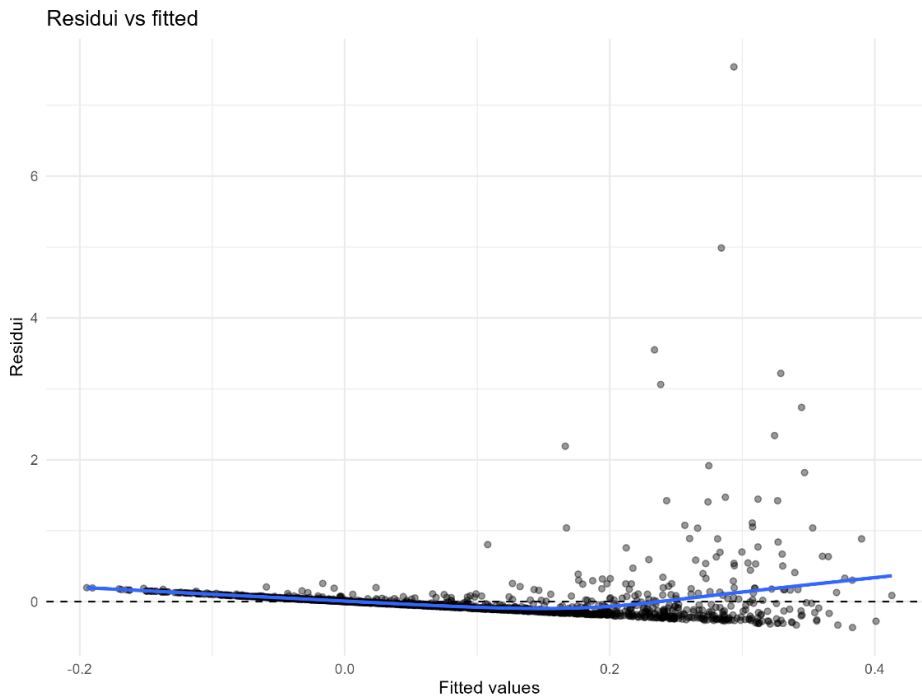


Figure 35: grafico dei residui rispetto ai valori stimati per EPO, modello 2

L'evidenza empirica ottenuta per i brevetti EPO (figura 33) restituisce un quadro solo parzialmente coerente con le attese teoriche derivate da Paci e Usai e proprio per questo, particolarmente interessante sotto il profilo interpretativo.

Il risultato più forte riguarda PS_{ij} che presenta coefficiente negativo e altamente significativo anche in presenza di errori robusti. Nel campione considerato, dunque, una maggiore specializzazione produttiva locale nello stesso settore non si associa a una più elevata intensità brevettuale, ma, al contrario, a una sua riduzione.

Questo esito non va letto come un'indicazione del fatto che, nel caso dei brevetti EPO, la geografia della produzione e la geografia dell'innovazione formalizzata non coincidono necessariamente. I contesti territoriali fortemente specializzati dal punto di vista occupazionale possono infatti identificare aree produttive mature ed efficienti sotto il profilo manifatturiero, ma non per questo centrali nei processi di generazione della conoscenza maggiormente codificata, selettiva e internazionalmente orientata che l'indicatore EPO tende a catturare.

Inoltre, poiché la variabile dipendente è costruita come intensità rapportata agli addetti, è plausibile che nei territori ad alta specializzazione il denominatore occupazionale cresca più rapidamente del numeratore brevettuale, contribuendo così a spiegare il segno negativo stimato.

Sul versante delle esternalità di Jacobs i risultati non forniscono conferme robuste: PD_j presenta segno negativo ma non significativo, mentre l'interazione $PD_j * DM_j$ risulta negativa e significativa, segnalando che nei sistemi locali classificati come metropolitani la diversificazione produttiva non si traduce in un vantaggio brevettuale, ma appare anzi associata a una minore intensità innovativa.

Questo risultato suggerisce che la varietà urbana, da sola, non sia sufficiente a produrre maggiore brevettazione EPO. Nella stessa direzione va, pur senza significatività statistica, l'interazione $PD_j * DHT_i$ che non offre evidenza a favore di un effetto positivo della diversità nei settori ad alta tecnologia. I risultati relativi alle variabili science-based confermano ulteriormente questo quadro:

SBS_{ij} mostra segno positivo ma non significativo, mentre SBD_{ij} risulta negativo e debolmente significativo, indicando che una maggiore varietà interna al cluster scientifico non genera automaticamente complementarità cognitive favorevoli alla brevettazione, ma può piuttosto riflettere dispersione, frammentazione o insufficiente integrazione tra attività contigue.

L'unica variabile pienamente in linea con le attese teoriche è TO_i , che mantiene segno positivo e significatività elevata, mostrando come i settori caratterizzati da una maggiore opportunità tecnologica nazionale tendano anche a presentare una più alta intensità brevettuale a livello locale. Questo risultato è particolarmente importante, perché suggerisce che il modello colga correttamente almeno una dimensione fondamentale del fenomeno innovativo, vale a dire il ruolo decisivo della struttura tecnologica intrinseca dei settori.

Al contrario, $NORD_j$ e $UNIV_j$ non risultano statisticamente significativi nel modello completo: ciò non implica che la dimensione territoriale o la presenza universitaria siano irrilevanti in assoluto, ma piuttosto che, una volta controllato per specializzazione, diversità, struttura science-based, opportunità tecnologica e interazioni, il loro effetto medio residuo non emerge con sufficiente robustezza.

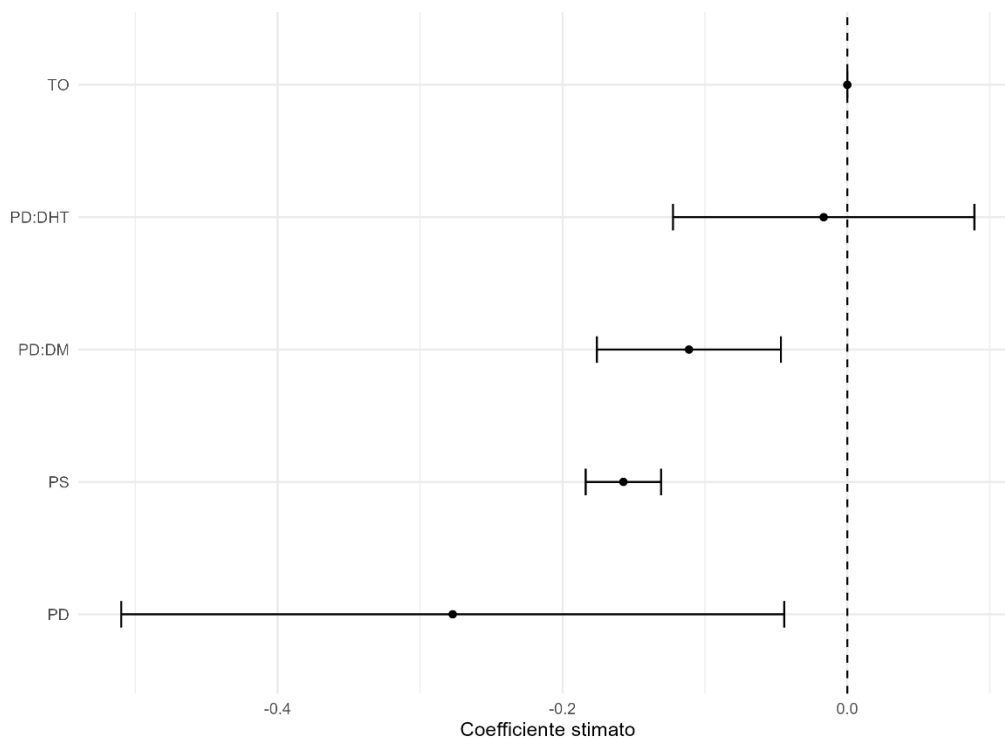


Figure 36: coefficient plot con errori robusti HC1

Dal grafico in figura 36, si vede bene che PS_{ij} è chiaramente negativo e significativo, $PD_j * DM_j$ è anch'esso negativo e significativo, SBD_{ij} è negativo e al limite della significatività, mentre PD_j , SBS_{ij} , $UNIV_j$, $NORD_j$ e $PD_j * DHT_i$ attraversano lo zero o vi si collocano molto vicino.

Il modello di regressione (1) restituisce un quadro sostanzialmente coerente con quello del modello di riferimento. Anche in questo caso la specializzazione produttiva è negativa e altamente significativa, mentre la diversità produttiva resta non significativa. Il confronto tra le due specificazioni mostra che

l'inclusione di controlli settoriali e science-based nel modello (2) non altera il risultato centrale, ma lo rende più informato sul piano economico. Questo rafforza l'idea che la relazione negativa tra specializzazione e brevetti EPO non sia il prodotto di un'omissione di variabili rilevanti, bensì un'evidenza da indagare.

Un ulteriore passaggio cruciale è rappresentato dal modello (3), che introduce le interazioni con la dummy Nord. Qui il dato più rilevante è il coefficiente positivo e significativo dell'interazione tra appartenenza al Nord Italia e specializzazione produttiva. Poiché il coefficiente diretto di PS_{ij} rimane negativo, il risultato non implica che nel Nord la specializzazione diventi in assoluto favorevole alla brevettazione EPO indica piuttosto che nel Nord l'effetto negativo della specializzazione si attenua. Questo è un passaggio molto importante alla luce della domanda di ricerca della tesi, perché suggerisce che la relazione tra struttura produttiva e innovazione non sia uniforme sul territorio nazionale. A parità di concentrazione settoriale, i sistemi locali del Nord sembrano in grado di assorbire meglio i costi potenziali della specializzazione, probabilmente grazie a un contesto istituzionale, infrastrutturale e relazionale più favorevole: reti di imprese più dense, maggiore prossimità a centri di conoscenza, migliori connessioni lungo le filiere e più elevata qualità complessiva dell'ecosistema innovativo. L'interazione tra Nord e diversità produttiva, invece, non risulta significativa, segnalando che il differenziale territoriale italiano incide soprattutto sulla capacità dei territori di rendere compatibile la specializzazione con l'innovazione, più che sul rendimento della varietà in quanto tale.

I modelli (4), (5) e (6) introducono rispettivamente gli effetti fissi di settore, gli effetti fissi regionali e la loro combinazione. Queste specificazioni svolgono una funzione essenziale di robustezza, poiché consentono di verificare se i risultati principali sopravvivano quando si controlla per eterogeneità non osservata a livello settoriale e territoriale. Il dato più importante è che la specializzazione produttiva continua a mantenere segno negativo e significatività elevata in tutte e tre le versioni, confermandosi come il risultato più robusto dell'intera analisi EPO. Ciò implica che l'associazione negativa tra specializzazione e brevettazione europea non dipende semplicemente dal fatto che alcuni settori siano più innovativi di altri o che alcune regioni presentino condizioni strutturali più favorevoli, ma continua a manifestarsi anche dopo aver assorbito tali componenti fisse.

Nel modello (4), che introduce le dummy settoriali, la diversità produttiva diventa negativa e significativa, mentre l'interazione tra diversità produttiva e settori high-tech assume segno positivo e significatività statistica. Questo risultato è teoricamente coerente con l'idea che gli effetti della varietà possano emergere in modo più nitido proprio nei comparti a maggiore intensità tecnologica, una volta depurati i dati dalle differenze medie tra settori. La lettura economica è che, all'interno dei settori high-tech, la diversità locale possa effettivamente favorire processi di ricombinazione cognitiva e complementarità tecnologiche. Tuttavia, poiché tale evidenza non risulta pienamente stabile nelle altre specificazioni, essa va interpretata con prudenza e non può essere considerata un risultato conclusivo.

Nel modello (5), che introduce le dummy regionali, emerge invece la significatività positiva della presenza di università. Questo suggerisce che il contributo delle istituzioni accademiche alla brevettazione EPO possa essere parzialmente nascosto nelle specificazioni prive di controlli territoriali più stringenti. Una volta tenuto fermo il contesto regionale, la presenza universitaria sembra associarsi a una maggiore intensità brevettuale, coerentemente con il ruolo delle università quali generatori e diffusori di conoscenza avanzata. Tuttavia, anche in questo caso la robustezza del risultato è limitata, poiché la significatività si attenua nella specificazione congiunta finale.

Il modello (6) rappresenta infine la verifica più severa, poiché combina effetti fissi di settore e di regione. In questa configurazione, la maggior parte dei regressori perde precisione o significatività, il che non è sorprendente: una quota importante della variabilità osservata viene assorbita dalle

eterogeneità strutturali che operano a livello settoriale e regionale. Proprio per questo, il fatto che PS_{ij} continui a risultare negativo e altamente significativo anche nel modello più restrittivo costituisce un'ulteriore conferma della solidità del risultato principale. Inoltre, l'interazione negativa tra diversità produttiva e aree metropolitane resta significativa, a testimonianza del fatto che la dimensione urbana non solo non compensa automaticamente gli effetti della varietà, ma può persino accentuarne il profilo sfavorevole rispetto alla brevettazione EPO.

Nel complesso, l'analisi econometrica dei brevetti EPO porta dunque a tre considerazioni di fondo.

- i. In primo luogo, la specializzazione produttiva locale non emerge come fattore favorevole alla produttività innovativa misurata tramite brevetti europei al contrario, la sua associazione è negativa, persistente e robusta a tutte le specificazioni considerate.
- ii. In secondo luogo, la diversità produttiva non presenta un effetto positivo generalizzato, ma sembra agire in modo contingente e condizionato dal contesto territoriale e settoriale.
- iii. In terzo luogo, l'eterogeneità geografica conta. L'interazione con il Nord mostra che gli effetti della specializzazione non sono uniformi nello spazio nazionale, mentre l'introduzione di effetti fissi di settore e di regione conferma che la geografia dell'innovazione è il risultato di una combinazione complessa di fattori locali, settoriali e istituzionali.

In questo senso, i risultati EPO sono coerenti con l'impostazione generale della tesi: la relazione tra struttura produttiva e innovazione non può essere interpretata in modo lineare e astratto, ma va letta alla luce delle specifiche configurazioni territoriali entro cui i processi innovativi prendono forma.

5.3 Modello di regressione OLS per i brevetti UIBM

Nel presente paragrafo si analizzano i risultati delle regressioni relative ai brevetti UIBM, preceduti da un esame preliminare della distribuzione delle variabili impiegate nel modello e della struttura delle correlazioni tra i regressori.

5.3.1 Le variabili della regressione: statistiche descrittive e correlazioni

L'analisi dei brevetti UIBM consente di osservare una dimensione dell'attività innovativa diversa da quella colta dai brevetti EPO. Se questi ultimi tendono a intercettare l'innovazione maggiormente formalizzata e orientata verso la frontiera tecnologica internazionale, i brevetti UIBM restituiscono una misura più ampia della capacità inventiva, maggiormente radicata nel contesto produttivo nazionale. Proprio per questo, prima di discutere i risultati delle regressioni, è utile soffermarsi sulle caratteristiche del campione e sulla struttura delle relazioni tra i regressori (*figura 37*).

Variabile	Tipo di variabile	Descrizione	Osservazioni	Media	Dev. Std	Min	Max
y_{ij}^{UIBM}	Variabile dipendente	Produttività dell'innovazione (UIBM) del settore i nel SLL j	3736	0,0543	0,1976	4,40E-05	5,25
PS_{ij}	Variabile indipendente	Specializzazione produttiva del settore i nel SLL j	3736	-0,1892	0,4663	-0,997	0,9768
PD_j	Variabile indipendente	Diversità produttiva nel SLL j	3998	0,2744	0,0471	0,1074	0,3852
TO_{ij}^{UIBM}	Variabile indipendente	Numero di brevetti del settore i nel SLL j	3998	29587	24383	582	77341
SBS_{ij}	Variabile di controllo	Specializzazione science - based del settore i nel SLL j	2233	0,6233	0,4976	-1	0,9925
SBD_{ij}	Variabile di controllo	Diversità science - based del settore i nel SLL j	2137	0,4018	0,1867	0,0044	1
$UNIV_j$	Variabile di controllo	Presenza di università nel SLL j	3998	0,2218	0,4155	0	1
$NORD_j$	Variabile di controllo	Appartenenza del SLL j al Nord Italia	3998	0,5067	0,5	0	1
DM_j	Variabile di controllo	Classificazione del SLL j in area metropolitana	3998	0,4	0,4901	0	1
DHT_i	Variabile di controllo	Classificazione di i come settore ad alta propensione innovativa	3648	0,9407	0,236	0	1

Figure 37: statistiche descrittive delle variabili utilizzate nella regressione OLS dei brevetti UIBM

Un primo elemento che emerge dalle statistiche descrittive riguarda la distribuzione della variabile dipendente. Anche in questo caso la distribuzione è asimmetrica, ma in misura meno accentuata rispetto a quella osservata per i brevetti EPO. L'attività brevettuale nazionale risulta dunque concentrata in un numero relativamente ristretto di combinazioni settore-territorio, ma il grado di concentrazione è inferiore rispetto a quello dei brevetti europei. Questa differenza è coerente con la natura stessa dell'indicatore: il brevetto UIBM cattura un insieme più ampio di attività inventive, comprendendo anche forme di innovazione meno selettive e più strettamente legate al tessuto produttivo interno.

Le statistiche descrittive mostrano inoltre che le variabili non condividono lo stesso numero di osservazioni. In particolare, SBS_{ij} e SBD_{ij} risultano disponibili solo per un sottoinsieme più ristretto del campione. La riduzione del campione nelle specificazioni che includono le variabili science-based costituisce un punto metodologicamente rilevante, perché implica che il confronto tra modelli debba essere condotto tenendo presente non soltanto la diversa formulazione econometrica, ma anche il diverso insieme di osservazioni effettivamente utilizzato. Nel caso UIBM, questa considerazione è particolarmente importante, anche alla luce della perdita di copertura già discussa nella costruzione del dataset.

Dal lato dei regressori, il quadro è piuttosto articolato. La variabile di specializzazione produttiva PS_{ij} assume valori sia negativi sia positivi ed evidenzia una dispersione ampia, segnalando una marcata eterogeneità nei livelli di concentrazione settoriale locale. La diversità produttiva PD_j , al contrario, presenta una variabilità più contenuta. Le opportunità tecnologiche settoriali operano su una scala del tutto diversa rispetto agli altri regressori, con valori medi molto elevati. Ciò rende evidente che il coefficiente associato a tale variabile non possa essere confrontato direttamente, in termini di ampiezza numerica, con quelli delle altre covariate. Le dummy territoriali delineano, inoltre, un campione nel quale il 50,7% delle osservazioni appartiene a sistemi locali del Nord, il 40% ricade in aree metropolitane e il 22,2 per cento interessa sistemi locali con presenza universitaria. Anche in questo caso la quota di osservazioni relative a settori ad alta propensione innovativa è molto elevata, dal momento che DHT_i assume valore unitario nel 94,1% dei casi. Ne deriva che il campione UIBM, pur essendo meno selettivo di quello EPO, resta comunque fortemente concentrato nei comparti maggiormente innovativi.

La lettura della matrice di correlazione (*figura 38*) consente di integrare queste prime evidenze.

variabile	PS_ij	PD_j	TO_i_UIBM	SBS_ij	SBD_ij	DHT_i	DM_j	UNIV_j	NORD_j
PS_ij	1	0.25	-0.1501	0.0619	-0.1751	-0.0244	-0.0655	0.0395	0.2301
PD_j	0.25	1	-0.0079	0.0644	0.0336	-0.0883	0.487	0.4872	0.3135
TO_i_UIBM	-0.1501	-0.0079	1	-0.2	0.1088	0.2294	0.0068	0.0049	-0.0212
SBS_ij	0.0619	0.0644	-0.2	1	0.2596	-0.0571	-0.045	-0.0349	0.0546
SBD_ij	-0.1751	0.0336	0.1088	0.2596	1	0.016	0.0427	0.0307	-0.08
DHT_i	-0.0244	-0.0883	0.2294	-0.0571	0.016	1	-0.0692	-0.0446	-0.0274
DM_j	-0.0655	0.487	0.0068	-0.045	0.0427	-0.0692	1	0.5446	-0.081
UNIV_j	0.0395	0.4872	0.0049	-0.0349	0.0307	-0.0446	0.5446	1	0.2836
NORD_j	0.2301	0.3135	-0.0212	0.0546	-0.08	-0.0274	-0.081	0.2836	1

Figure 38: matrice di correlazione tra le variabili del modello di regressione OLS per i brevetti UIBM

Nel complesso, non emergono relazioni lineari di intensità tale da far presumere seri problemi di multicollinearità. Le correlazioni più elevate riguardano soprattutto alcune variabili di controllo territoriali. Al di fuori di queste associazioni, i coefficienti risultano generalmente contenuti. La correlazione tra specializzazione e diversità produttiva è positiva ma moderata, mentre le opportunità tecnologiche settoriali mostrano legami deboli con le altre covariate, a conferma del fatto che esse colgono una dimensione distinta, prevalentemente riconducibile alla struttura tecnologica dei comparti produttivi. In sintesi, la matrice suggerisce una parziale sovrapposizione tra alcune caratteristiche territoriali, ma non segnala un'intensità dei legami tale da compromettere l'identificazione dei singoli effetti nelle regressioni.

5.3.2. Risultati della regressione per i brevetti UIBM

Le sei specificazioni stimate per la regressione OLS (figura 39) consentono di valutare la robustezza dei coefficienti al variare dei controlli inclusi, delle interazioni territoriali e degli effetti fissi di settore e di regione. Va precisato sin d'ora che i coefficienti accompagnati da asterischi sono statisticamente significativi ai livelli indicati nella nota della tabella, mentre i coefficienti privi di asterischi non risultano statisticamente significativi e non permettono quindi di trarre conclusioni affidabili sull'esistenza di un effetto diverso da zero.

	(1)	(2)	(3)	(4)	(5)	(6)
Specializzazione produttiva	-0,1393*** (0,0101)	-0,1360*** (0,0131)	-0,1547*** (0,0174)	-0,1245*** (0,0121)	-0,1369*** (0,0134)	-0,1251*** (0,01226)
Diversità produttiva	-0,6387*** (0,1113)	-0,3365** (0,1292)	-0,3222** (0,1493)	-0,0395 (0,3457)	-0,4269** (0,1943)	-0,1986 (0,4359)
Opportunità tecnologiche settoriali	-8,7e-07 (1,95e-06)	8,5578e-06*** (2,0319e-06)	7,69e-06*** (2,03e-06)		8,54e-06*** (2e-06)	
Appartenenza al Nord Italia	0,0142** (0,0074)	0,0163* (0,0101)	0,0166 (0,0642)	0,0152 (0,0086)		
Presenza di università	-0,0022 (0,009)	0,0113 (0,0086)	0,0113 (0,0086)	0,0128 (0,0086)	0,0134* (0,0076)	0,0147* (0,0076)
Diversità produttiva * Aree metropolitane	0,0052 (0,0306)	-0,0425 (0,0269)	-0,0456 (0,0276)	-0,0334 (0,0273)	-0,0327 (0,0327)	-0,0241 (0,0332)
Diversità produttiva * Settori high tech	0,0117 (0,0296)	-0,1418** (0,0519)	-0,1546** (0,0529)	-0,5015 (0,3471)	-0,1299* (0,0711)	-0,41163 (0,3894)
Specializzazione science-based		0,0254*** (0,0077)	0,0244** (0,0080)	0,0213 (0,0178)	0,0245** (0,0079)	0,0216 (0,0174)
Diversità science-based		-0,1397*** (0,0246)	-0,1393*** (0,0245)	-0,1422*** (0,0262)	-0,1351*** (0,0248)	-0,1362*** (0,0266)
Appartenza al Nord Italia * Diversità produttiva			-0,0227 (0,2201)			
Appartenza al Nord Italia * Specializzazione produttiva			0,0357* (0,0252)			
Dummy Settore	NO	NO	NO	SI	NO	SI
Dummy Regione	NO	NO	NO	NO	SI	SI
Intercetta	0,1985*** (0,0282)	0,1733*** (0,0320)	0,1689*** (0,0392)	0,1802*** (0,0324)	0,1966*** (0,0538)	0,2051*** (0,0544)
Osservazioni	3381	2133	2133	2133	2133	2133
R-quadro corretto	0,1384	0,1682	0,1692	0,1893	0,1774	0,1986

Nota: errori standard robusti tra parentesi (*p < 0,10 **p < 0,05 ***p < 0,001). La variabile dipendente è il numero di brevetti UIBM per settore i e SLL j normalizzato rispetto al numero di addetti del settore i nel SLL j

Figure 39: esiti delle analisi di regressione OLS per i brevetti UIBM

In questa sezione, il modello (2) rappresenta la specificazione più informativa e può essere assunto come punto di riferimento dell'interpretazione. La ragione soprattutto nella possibilità di esaminare esplicitamente il ruolo dei cluster science-based attraverso le variabili SBS_{ij} e SBD_{ij} . Tale aspetto è centrale rispetto all'impostazione della tesi, che non si limita a verificare se continuo di più specializzazione o diversità produttiva, ma mira anche a comprendere in che modo la struttura delle attività a base scientifica influenzi la capacità innovativa dei territori.

Il modello presenta una capacità esplicativa moderata, ma comunque è più alta di quella osservata nel caso EPO. Il test F pari a 48,88 segnala inoltre che il modello è globalmente significativo.

La produttività brevettuale UIBM sembra essere spiegata un po' meglio dalla struttura territoriale e settoriale rispetto alla produttività brevettuale EPO. Questo è coerente con la natura dell'indicatore: i brevetti UIBM rappresentano una forma di tutela più ampia, meno selettiva e più vicina all'innovazione diffusa del tessuto produttivo locale, soprattutto manifatturiero e di PMI.

Anche in questo caso il modello deve essere letto con cautela. La *figura 40* mostra che il modello intercetta una parte della variabilità, ma la massa delle osservazioni resta fortemente concentrata vicino a zero, con pochi casi molto elevati. Questo è coerente con la distribuzione di $UIBM_weight_i$, già definita come fortemente sbilanciata a destra: quasi tutti gli SLL hanno valori bassi, mentre pochi casi estremi concentrano livelli di intensità molto alti. La stessa sezione sottolinea che il massimo è 2,427, il 99° percentile è 0,137 e il coefficiente di variazione è molto elevato, segnalando una distribuzione dominata da outlier.

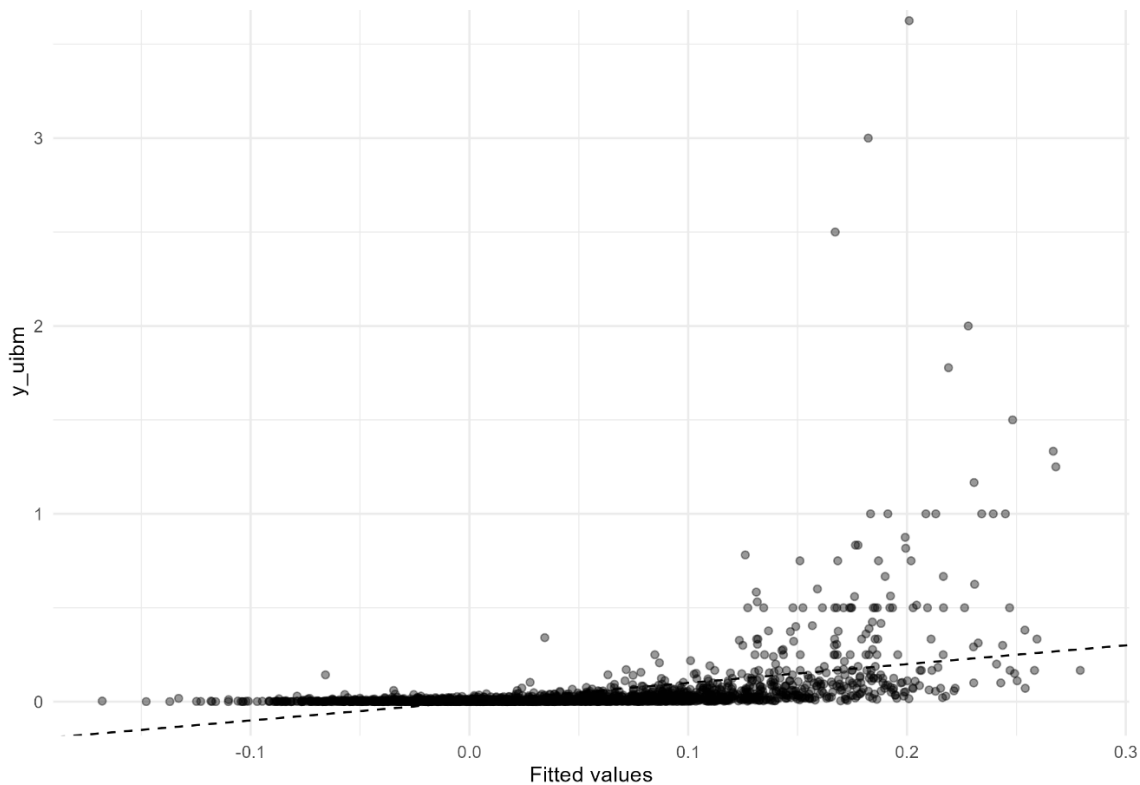


Figure 40: grafico di confronto tra valori osservati e valori stimati, modello 2

I test diagnostici (*tabella 18*) mostrano eteroschedasticità molto netta e una forte evidenza di miss specificazione funzionale. Dunque anche per UIBM l'interpretazione corretta è quella basata sugli errori robusti HC1, non sugli errori standard classici.

Table 18: risultati dei test diagnostici condotti sul modello di regressione OLS sui brevetti UIBM

Test diagnostico	Statistica	p-value
Breusch – Pagan	42,8486	2,3e-06
RESET	193,187	8,46e-78

Anche il grafico in *figura 41* conferma eteroschedasticità e non linearità residua, la curva smooth ha una chiara forma a U e la dispersione dei residui cresce quando aumentano i fitted.

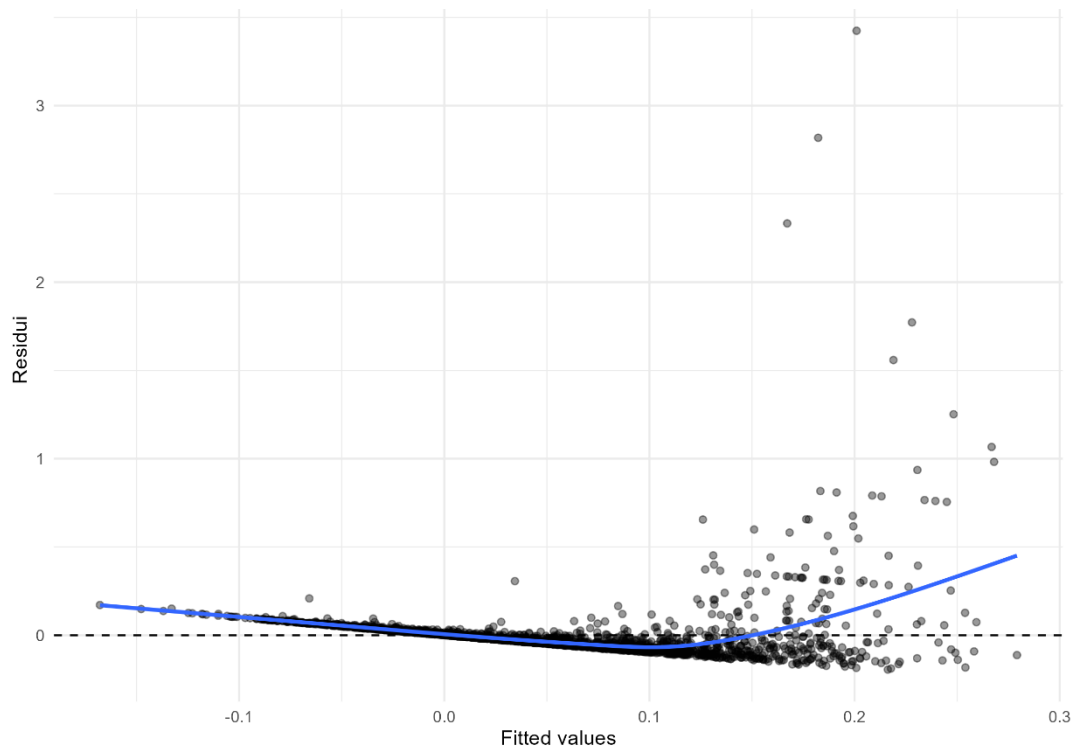


Figure 41: grafico di confronto tra valori residui e valori stimati per UIBM, modello 2

Dunque anche per UIBM l'interpretazione corretta è quella basata sugli errori robusti HC1, non sugli errori standard classici.

Il coefficiente di PS_{ij} è negativo e molto significativo, questo è il dato centrale del modello UIBM. Anche nel caso dei brevetti nazionali, quindi, una maggiore specializzazione produttiva locale nello stesso settore non si associa a una maggiore intensità brevettuale per addetto al contrario, la relazione stimata va in senso opposto.

Questo risultato conferma nel periodo e nella geografia analizzati la specializzazione occupazionale non coincide automaticamente con la propensione al deposito brevettuale per addetto.

I territori molto specializzati sono spesso territori ad alta densità produttiva, ma non necessariamente quelli in cui la formalizzazione della conoscenza tramite brevetto è più intensa in rapporto alla base occupazionale. Inoltre, anche qui la variabile dipendente è una misura di intensità: se gli addetti crescono molto nei territori fortemente specializzati, il denominatore può aumentare più rapidamente del numeratore brevettuale. Infine, nel caso UIBM conta anche il criterio di geolocalizzazione: l'indicatore è attribuito sulla base dell'applicant, cioè del depositante, non dell'inventore. Questo significa che la dimensione organizzativa e amministrativa della protezione brevettuale conta molto, e può non sovrapporsi perfettamente alla localizzazione della produzione o della conoscenza tecnica.

Nel caso UIBM il risultato su PD_j è ancora più netto che per EPO, il coefficiente è negativo e significativo. Questo significa che, mediamente, una maggiore diversificazione produttiva locale non favorisce la produttività brevettuale UIBM ma al contrario, vi si associa negativamente.

L'interazione $PD_j * DM_j$ ha segno negativo ma non è significativa, quindi non emerge un effetto specifico delle aree metropolitane. In altre parole, non c'è evidenza robusta che la diversità urbana rafforzi o modifichi in modo sistematico l'effetto medio di PD_j . Al contrario, l'interazione $PD_j * DHT_i$ è

negativa e significativa dunque nei settori ad alta tecnologia, la diversificazione produttiva locale appare ancora meno favorevole alla produttività brevettuale UIBM.

Nel complesso, non solo non emerge un effetto positivo della diversità locale, ma nel caso UIBM si osserva un'associazione negativa sia in media sia, con maggiore forza, nei settori high-tech. La diversificazione cattura territori più generalisti, nei quali convivono molte attività ma senza che ciò si traduca automaticamente in complementarità efficaci per la formalizzazione brevettuale. La varietà, insomma, non basta da sola a generare più brevetti.

Più articolate sono le evidenze che si possono trarre dalla lettura dei coefficienti delle variabili science-based.

SBS_{ij} è positivo e significativo, questo suggerisce che la specializzazione locale in attività appartenenti allo stesso cluster science-based ha un effetto favorevole sulla produttività brevettuale UIBM. Il risultato segnala che la prossimità scientifica/tecnologica, quando assume la forma di una maggiore concentrazione entro lo stesso cluster, può effettivamente sostenere l'attività brevettuale.

SBD_{ij} è invece negativo e molto significativo. La maggiore varietà interna al cluster science-based non aiuta anzi, si associa a minore produttività brevettuale.

In sintesi, conta di più la concentrazione di competenze affini che non la loro semplice compresenza eterogenea, per gli UIBM sembra funzionare meglio una base cognitiva coerente e addensata, piuttosto che una varietà interna troppo dispersa. Questo punto è interessante perché suggerisce una forma di spillover più "ristretta": la prossimità scientifica sembra essere utile quando produce massa critica, non quando si traduce in una diversità interna troppo ampia.

TO_i è positivo e altamente significativo. Questo conferma che l'opportunità tecnologica settoriale continua a essere una delle dimensioni più solide del modello. I settori che, a livello nazionale, dispongono di una base tecnologica più ricca tendono anche a mostrare una maggiore produttività brevettuale locale.

Questo è un risultato prezioso perché dà coerenza all'intero impianto empirico: anche quando i risultati sulle esternalità territoriali tradizionali sono deboli o controintuitivi, il modello continua a cogliere bene il fatto che la propensione a brevettare dipende in misura importante dalle caratteristiche intrinseche del settore.

Le dummy $NORD_j$ e $UNIV_j$ risultano positive ma non significative. Quindi, una volta controllato per specializzazione, diversità, cluster science-based, opportunità tecnologica e interazioni, il loro effetto medio residuo non emerge in modo robusto.

Questo non significa che il Nord o la presenza universitaria non contino in assoluto. Anzi, nella parte descrittiva del capitolo UIBM emerge chiaramente un vantaggio del Nord sia nei conteggi sia, in misura più strutturale, anche nell'intensità per addetto, e la mappa in quantili suggerisce una presenza più estesa di SLL del Nord nelle classi alte. Tuttavia, nel modello di regressione quell'effetto generale viene in larga misura assorbito dalle altre covariate, più vicine al meccanismo economico che si vuole spiegare.

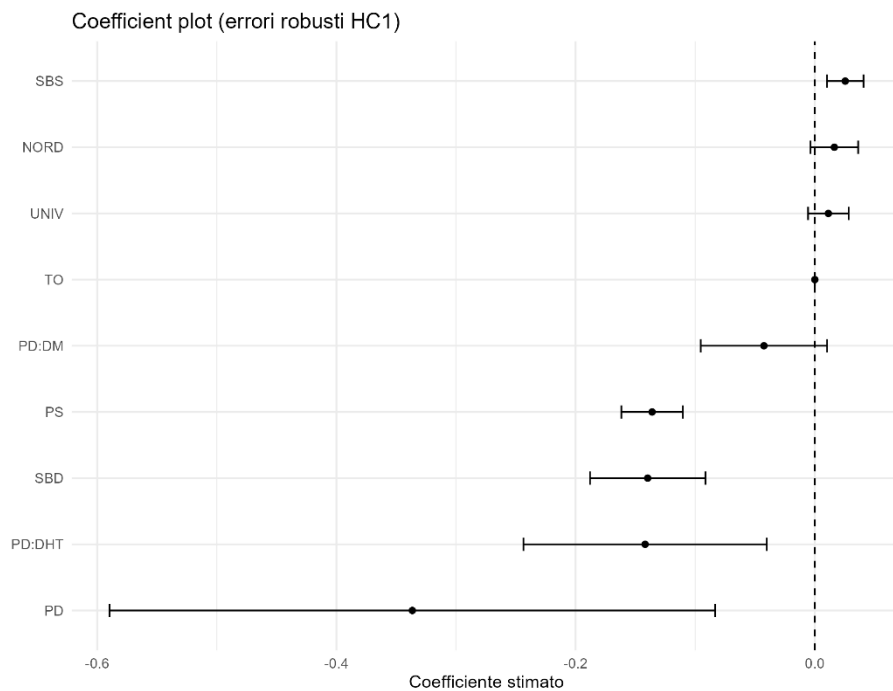


Figura 36: Boxplot dei coefficienti stimati nel modello UIBM

Il grafico in figura 36 mostra con chiarezza tre gruppi di risultati:

- coefficienti robustamente negativi: PS_{ij} , PD_{ij} , SBD_{ij} , $PD_j * DHT_i$;
- coefficienti robustamente positivi: TO_i e SBS_{ij} ;
- coefficienti non robusti: $NORD_j$ e $UNIV_j$, $PD_j * DM_j$;

Il boxplot mostra che il modello UIBM non conferma pienamente la formulazione classica delle esternalità di specializzazione e diversità. Al contrario, i risultati suggeriscono una configurazione più selettiva: la specializzazione produttiva e la diversità generale risultano associate negativamente alla variabile dipendente, mentre la specializzazione interna ai cluster science-based e l'opportunità tecnologica mostrano un'associazione positiva. Ne emerge dunque un quadro in cui gli effetti territoriali e settoriali non agiscono in modo uniforme, ma dipendono dalla specifica natura delle complementarità tecnologiche considerate.

Il passaggio ai modelli successivi consente di capire se questi risultati resistano all'introduzione di ulteriori controlli. Il modello (1), più parsimonioso, restituisce già una configurazione simile: sia specializzazione sia diversità produttiva sono negative e statisticamente significative. Tuttavia, il confronto con il modello (2) mostra che l'inclusione delle variabili science-based e delle opportunità tecnologiche riduce l'ampiezza assoluta del coefficiente di PD_j , pur senza modificarne il segno. Questo suggerisce che una parte dell'associazione negativa inizialmente osservata sia mediata dalla composizione tecnologica del territorio e dalla struttura della sua base scientifica.

Il modello (3) introduce le interazioni con la dummy Nord, aggiungendo una dimensione particolarmente rilevante rispetto alla domanda di ricerca della tesi. Il risultato più interessante riguarda l'interazione tra appartenenza al Nord e specializzazione produttiva, che è positiva e significativa. Poiché il coefficiente diretto di PS_{ij} resta negativo, il dato non implica che nel Nord la specializzazione diventi di per sé favorevole, ma segnala che il suo effetto negativo risulta meno intenso nei sistemi locali settentrionali. L'interpretazione è coerente con l'idea che la stessa struttura produttiva possa produrre effetti differenti a seconda del contesto territoriale in cui si inserisce. Nel

Nord, la presenza di reti industriali più dense, infrastrutture migliori, maggiore qualità istituzionale e una più stretta integrazione tra imprese e fonti di conoscenza possono attenuare gli effetti sfavorevoli di una forte concentrazione settoriale. L'interazione tra Nord e diversità produttiva non risulta invece significativa, suggerendo che il differenziale territoriale italiano non incida in maniera apprezzabile sul rendimento della varietà locale.

Le specificazioni (4), (5) e (6) introducono rispettivamente effetti fissi di settore, effetti fissi regionali e la loro combinazione. L'utilità di questi modelli è soprattutto quella di sottoporre a verifica la robustezza dei coefficienti principali rispetto a forme di eterogeneità non osservata. Il primo elemento da sottolineare è che la specializzazione produttiva conserva in tutte le specificazioni segno negativo e significatività elevata. Questo è il risultato più robusto dell'intera analisi UIBM. Esso indica che l'associazione negativa tra concentrazione produttiva locale e brevetti nazionali non dipende soltanto dal fatto che alcuni settori o alcune regioni siano mediamente più innovativi di altri, ma continua a manifestarsi anche quando tali differenze vengono assorbite dagli effetti fissi.

La diversità produttiva, invece, mostra una minore stabilità. Nei modelli senza effetti fissi settoriali essa tende ad assumere un coefficiente negativo e significativo; quando però vengono introdotte le dummy di settore, come nei modelli (4) e (6), la significatività scompare. Questo passaggio è importante, perché suggerisce che una parte dell'effetto attribuito alla varietà produttiva sia in realtà legata a differenze medie tra comparti. In questo senso, il risultato invita a una lettura più cauta: nel caso UIBM, la diversità non sembra agire come determinante autonomo e robusto dell'attività brevettuale, ma appare fortemente condizionata dalla composizione settoriale del territorio.

Le variabili science-based mostrano, a loro volta, una configurazione interessante. La specializzazione science-based resta positiva e significativa nei modelli (2), (3) e (5), ma perde significatività nelle specificazioni con effetti fissi di settore, segnalando che una parte del suo effetto positivo è assorbita dall'eterogeneità strutturale dei comparti produttivi. La diversità science-based, invece, mantiene segno negativo e significatività elevata in tutte le versioni più ricche del modello, confermandosi come un risultato molto stabile. Questo rafforza l'idea che la dispersione delle attività a base scientifica costituisca, nel caso UIBM, un elemento sfavorevole alla capacità innovativa.

Un cenno a parte merita la presenza di università. Nei modelli con effetti fissi regionali, la variabile assume coefficiente positivo e significativo. Il risultato suggerisce che il ruolo delle istituzioni accademiche emerga con maggiore chiarezza una volta tenuto costante il contesto regionale, cioè una volta isolato l'effetto di differenze territoriali più ampie. Anche se la significatività non è uniforme in tutte le specificazioni, il segnale che proviene dai modelli (5) e (6) è coerente con l'idea che le università possano contribuire positivamente alla diffusione di conoscenza e al rafforzamento della capacità inventiva locale.

Nel complesso, i risultati UIBM delineano un quadro abbastanza chiaro. La specializzazione produttiva locale emerge come fattore sistematicamente associato in modo negativo alla performance innovativa, e questo risultato appare robusto a tutte le principali verifiche. La diversità produttiva non fornisce invece un sostegno convincente all'ipotesi di un effetto favorevole al contrario, il suo segno è spesso negativo e la sua significatività dipende dalla specificazione adottata. Infine, la struttura dei cluster science-based si rivela cruciale: la concentrazione delle attività a base scientifica appare favorevole, mentre la loro dispersione risulta persistentemente sfavorevole.

Anche nel caso dei brevetti UIBM, dunque, la relazione tra struttura produttiva e innovazione non può essere letta in modo lineare. Essa dipende dall'interazione tra caratteristiche settoriali, organizzazione della base conoscitiva e differenze territoriali, confermando che i meccanismi locali dell'innovazione operano in modo condizionato e non uniforme nello spazio nazionale.

5.4 Risultati della regressione per le Start Up Innovative

Nel presente paragrafo si analizzano i risultati delle regressioni relative alle start up, preceduti da un esame preliminare della distribuzione delle variabili impiegate nel modello e della struttura delle correlazioni tra i regressori.

5.3.1 Le variabili della regressione: statistiche descrittive e correlazioni

Nel caso delle start-up innovative, l'analisi empirica intercetta una dimensione dell'innovazione diversa rispetto a quella osservata attraverso i brevetti. Qui l'oggetto di studio non è l'output inventivo formalizzato, bensì la capacità dei territori di generare nuova imprenditorialità innovativa. Proprio per questa ragione, prima di discutere le regressioni è utile soffermarsi sulle caratteristiche del campione e sulla struttura delle relazioni tra i regressori (*figura 42*), così da inquadrare meglio il significato economico dei risultati.

Variabile	Tipo di variabile	Descrizione	Osservazioni	Media	Dev. Std	Min	Max
$y_{ij}^{StartUp}$	Variabile dipendente	Produttività dell'innovazione (StartUp) del settore i nel SLL j	35713	8,21E-04	0,0198	0,00E+00	1,5
PS_{ij}	Variabile indipendente	Specializzazione produttiva del settore i nel SLL j	35713	-0,1964	0,4096	-0,997	0,9988
PD_j	Variabile indipendente	Diversità produttiva nel SLL j	35713	0,2452	0,0527	0,1074	0,3852
$TO_{ij}^{StartUp}$	Variabile indipendente	Numero di brevetti del settore i nel SLL j	30300	19,204	58,498	1	418
SBS_{ij}	Variabile di controllo	Specializzazione scienze - based del settore i nel SLL j	8006	0,5139	0,5707	-1	0,9925
SBD_{ij}	Variabile di controllo	Diversità scienze - based del settore i nel SLL j	7596	0,3869	0,2015	0,0017	1
$UNIV_j$	Variabile di controllo	Presenza di università nel SLL j	35713	0,1226	0,328	0	1
$NORD_j$	Variabile di controllo	Appartenenza del SLL j al Nord Italia	35713	0,3625	0,4807	0	1
DM_j	Variabile di controllo	Classificazione del SLL j in area metropolitana	35713	0,2439	0,4294	0	1
DHT_i	Variabile di controllo	Classificazione di i come settore ad alta propensione innovativa	22636	0,6779	0,4673	0	1

Figure 42: statistiche descrittive delle variabili del modello di regressione OLS per le Start Up

Un primo elemento di rilievo emerge già dalla distribuzione della variabile dipendente. La produttività dell'innovazione misurata tramite start-up innovative presenta una media estremamente bassa con una deviazione standard di 0,0198, un minimo pari a zero e un massimo di 1,5. Ciò segnala con evidenza che il fenomeno osservato è fortemente rarefatto: nella grande maggioranza delle combinazioni settore-territorio la presenza di start-up innovative è nulla o molto contenuta, mentre soltanto in un numero limitato di casi si osservano valori relativamente più elevati. In termini sostanziali, la nascita di start-up innovative appare dunque come un evento raro, molto più di quanto non accada per le misure brevettuali. Questo aspetto è importante anche ai fini dell'interpretazione econometrica, perché suggerisce fin dall'inizio che la componente idiosincratICA del fenomeno sia elevata e che una quota non trascurabile della variabilità dipenda da fattori difficilmente catturabili con sole variabili strutturali di settore e territorio.

Le statistiche descrittive mostrano inoltre una forte differenza nel numero di osservazioni disponibili a seconda delle variabili considerate, questo significa che il confronto tra le diverse specificazioni di regressione deve essere condotto con particolare cautela: i modelli più ricchi non solo introducono più controlli, ma si basano anche su un sottocampione molto più ristretto. Nel caso delle start-up, tale aspetto è particolarmente rilevante, perché il passaggio dal modello base alle specificazioni più articolate comporta una perdita molto consistente di osservazioni.

Dal lato dei regressori, il quadro restituisce una configurazione piuttosto eterogenea. La variabile di specializzazione produttiva PS_{ij} assume una notevole dispersione dei livelli di concentrazione settoriale locale. La diversità produttiva PD_j , invece presenta una variabilità più contenuta. Le opportunità tecnologiche settoriali $TO_i^{StartUp}$ mostrano, come già accaduto negli altri modelli, una scala numerica molto diversa dalle altre covariate, di conseguenza l'ampiezza del coefficiente associato a tale variabile non può essere confrontata direttamente con quella degli altri regressori. Le dummy territoriali delineano inoltre un campione in cui il 36,3% delle osservazioni appartiene a sistemi locali del Nord, il 24,4% a contesti metropolitani e il 12,3% a sistemi locali con presenza universitaria. Meno elevata rispetto ai casi EPO e UIBM è invece la quota di osservazioni appartenenti a settori ad alta propensione innovativa: DHT_i assume valore unitario nel 67,8% dei casi. Questo suggerisce che, nel caso delle start-up, la base empirica risulta più ampia e meno rigidamente concentrata nei soli comparti tecnologicamente avanzati.

La matrice di correlazione (figura 43) conferma che non emergono problemi particolarmente gravi di multicollinearità.

variabile	PS_ij	PD_j	TO_i_StartUp	SBS_ij	SBD_ij	DHT_i	DM_j	UNIV_j	NORD_j
PS_ij	1	0.1042	-0.1493	0.0202	-0.0023	-0.1708	0.0559	0.0509	0.0232
PD_j	0.1042	1	-0.0057	0.2513	0.0321	0.0064	0.5047	0.4537	0.3776
TO_i_StartUp	-0.1493	-0.0057	1	0.0994	0.0904	0.1816	-0.0052	-0.0035	-0.0017
SBS_ij	0.0202	0.2513	0.0994	1	0.155	0.0073	0.0972	0.0834	0.1071
SBD_ij	-0.0023	0.0321	0.0904	0.155	1	-0.0942	0.0694	0.0651	-0.053
DHT_i	-0.1708	0.0064	0.1816	0.0073	-0.0942	1	-0.008	-0.0083	0.0034
DM_j	0.0559	0.5047	-0.0052	0.0972	0.0694	-0.008	1	0.5255	0.021
UNIV_j	0.0509	0.4537	-0.0035	0.0834	0.0651	-0.0083	0.5255	1	0.3034
NORD_j	0.0232	0.3776	-0.0017	0.1071	-0.053	0.0034	0.021	0.3034	1

Figure 43: matrice di correlazione tra le variabili della regressione OLS delle Start Up

Le associazioni più elevate si osservano, anche qui, tra alcune variabili territoriali: la correlazione tra area metropolitana e presenza universitaria è pari a 0,5255, mentre la diversità produttiva è positivamente correlata sia con DM_j sia con $UNIV_j$. Si tratta di un risultato coerente con l'idea che i sistemi locali più diversificati coincidano più spesso con contesti urbani di maggiore dimensione e con territori dotati di infrastrutture di conoscenza. Anche la correlazione tra PD_j e $NORD_j$ è positiva, sebbene più contenuta. Al di fuori di queste relazioni, i coefficienti risultano generalmente modesti: la correlazione tra specializzazione e diversità produttiva è bassa, mentre le opportunità tecnologiche settoriali mostrano legami contenuti con il resto delle covariate. Nel complesso, la matrice suggerisce dunque una parziale sovrapposizione tra alcune caratteristiche territoriali, ma non tale da compromettere in modo evidente la leggibilità dei coefficienti di regressione.

5.4.2. Risultati della regressione per le Start Up

Su queste basi si può passare alla discussione delle stime econometriche (figura 44). Le sei specificazioni considerate permettono di verificare quanto i risultati siano sensibili all'inclusione di variabili aggiuntive, interazioni territoriali ed effetti fissi di settore e di regione. Come nelle sezioni precedenti, i coefficienti accompagnati da asterischi sono statisticamente significativi ai livelli riportati nella nota della tabella, mentre i coefficienti privi di asterischi non sono statisticamente significativi e pertanto, non consentono di formulare inferenze affidabili su un effetto diverso da zero.

Il tratto più netto che emerge dall'intera analisi riguarda la specializzazione produttiva. In tutte e sei le specificazioni, PS_{ij} presenta coefficiente negativo e altamente significativo. Si tratta del risultato più robusto del modello sulle start-up innovative. L'evidenza suggerisce che, a parità delle altre condizioni, i territori maggiormente concentrati in uno specifico settore non siano quelli in cui la nascita di nuove imprese innovative risulta più intensa. Dal punto di vista interpretativo, questo risultato appare coerente con l'idea che le start-up innovative traggano vantaggio non tanto da contesti produttivi già fortemente consolidati e specializzati, quanto piuttosto da ambienti più aperti, in cui sia possibile combinare competenze differenti, intercettare conoscenze eterogenee e sfruttare occasioni di ibridazione tra attività diverse. In questa prospettiva, l'eccessiva specializzazione può rappresentare un vincolo, più che una risorsa, perché tende a riprodurre traiettorie produttive esistenti anziché favorire la sperimentazione imprenditoriale.

	(1)	(2)	(3)	(4)	(5)	(6)
Specializzazione produttiva	-0,0023*** (0,0004)	-0,0025*** (0,0006)	-0,0047*** (0,0012)	-0,0021*** (0,0006)	-0,0026*** (0,0006)	-0,0021*** (0,0006)
Diversità produttiva	-0,0052 (0,0036)	-0,0092 (0,0056)	-0,0144* (0,0081)	0,010* (0,0048)	0,0036 (0,0060)	0,0225** (0,0072)
Opportunità tecnologiche settoriali	4,75e-05*** (8,87e-06)	3,05e-05*** (9,96e-06)	3,04e-05*** (9,02e-06)		3,05e-05*** (9,04e-06)	
Appartenenza al Nord Italia	-0,0015*** (0,0003)	-0,0017*** (0,0004)	-0,0039* (0,0021)	-0,0015*** (0,0003)		
Presenza di università	0,0006* (0,000939)	0,0008* (0,0005)	0,0003 (0,0004)	0,0006* (0,0004)	0,0006 (0,0005)	0,0005 (0,0004)
Diversità produttiva * Aree metropolitane	-0,0017* (0,0009)	-0,0011 (0,0016)	-0,0006 (0,0016)	-0,0011 (0,0014)	-0,0024 (0,0018)	-0,0022 (0,0016)
Diversità produttiva * Settori high tech	-0,0002 (0,000606)	-0,0002 (0,0012)	-0,0007 (0,0012)	-0,0198*** (0,0059)	-0,0002 (0,0012)	-0,0207** (0,0064)
Specializzazione science-based		-2,4e-05 (0,0003)	-6,9e-05 (0,0003)	-0,0007 (0,0006)	-3,7e-05 (0,0003)	-0,0008 (0,00062)
Diversità science-based		0,0008 (0,0009)	0,0004 (0,0009)	0,0016 (0,0011)	0,0007 (0,0009)	0,0015 (0,0011)
Appartenza al Nord Italia * Diversità produttiva			0,0141 (0,0086)			
Appartenenza al Nord Italia * Specializzazione produttiva			0,0054 (0,0015)			
Dummy Settore	NO	NO	NO	SI	NO	SI
Dummy Regione	NO	NO	NO	NO	SI	SI
Intercetta	0,0014 (0,0008)	0,0023 (0,0015)	0,0030 (0,0021)	0,0045 (0,0018)	-0,0029 (0,0018)	-0,0071*** (0,0019)
Osservazioni	18937	6580	6580	7577	6580	7577
R-quadro corretto	0,0216	0,0196	0,0215	0,0183	0,0219	0,0203

Nota: errori standard robusti tra parentesi (*p < 0,10 **p < 0,05 ***p < 0,001). La variabile dipendente è il numero di Start Up per settore i e SLL j normalizzato rispetto al numero di addetti del settore i nel SLL j

Figure 44: analisi di regressione OLS su Start Up

Il *modello 2* è considerato la specificazione di riferimento, perché consente di verificare in maniera esplicita il ruolo della componente science-based mantenendo una struttura ancora sufficientemente leggibile dei coefficienti.

Il modello presenta una capacità esplicativa molto contenuta. L'insieme delle variabili esplicative intercetta soltanto una quota molto ridotta della variabilità della propensione locale alla creazione di start-up innovative. Questo dato, di per sé, non invalida l'analisi, ma segnala che il fenomeno osservato è solo parzialmente spiegabile tramite la struttura produttiva e territoriale modellata sulla falsariga di Paci e Usai (2000) e che una parte molto rilevante della variabilità dipende verosimilmente da fattori non osservati, da forti eterogeneità locali o dalla natura stessa della variabile dipendente.

Table 19: risultati dei test diagnostici condotti sul modello di regressione OLS sulle Start Up, modello 2

Test diagnostico	Statistica	p-value
Breusch – Pagan	32,8260	0,000143
RESET	45,3774	2,68e-20

Dal punto di vista diagnostico (Tabella 19), i risultati confermano che la specificazione OLS soffre di problemi rilevanti. Il test di Breusch - Pagan risulta statisticamente significativo, indicando la presenza di eteroschedasticità; inoltre, il test RESET è anch'esso fortemente significativo, suggerendo una possibile miss specificazione funzionale del modello.

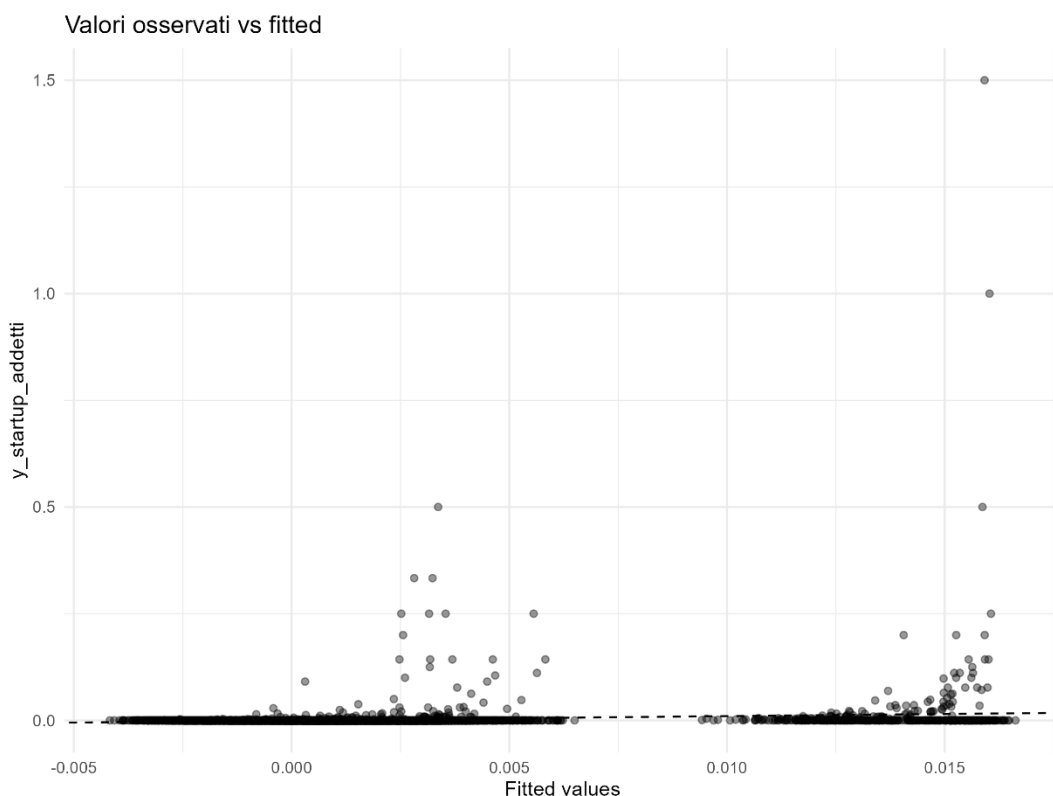


Figure 45: grafico di confronto tra valori osservati e valori stimati per le Start Up, modello 2

Questo quadro è coerente con la natura della distribuzione di $y_{ij}^{Start Up}$ che risulta fortemente zero-inflated, con una massa molto elevata di osservazioni pari o prossime a zero e una coda destra lunga, dovuta a pochi casi estremi. Già nel capitolo descrittivo emerge infatti che 292 SLL su 515 non registrano alcuna start-up innovativa e che, anche tra gli SLL positivi, l'intensità è in genere molto bassa, con pochi outlier estremi.

Nel grafico in *figura 45* i valori stimati risultano compressi in un intervallo molto ristretto, vicino allo zero, mentre i valori osservati presentano una forte concentrazione sulla linea dello zero e pochi punti isolati molto più elevati. Questo significa che il modello tende a schiacciare la previsione verso il valore medio e fatica a riprodurre la forte asimmetria empirica della distribuzione, la forma distributiva di $y_{ij}^{Start Up}$ è poco compatibile con l'ipotesi lineare classica.

Sul piano dei coefficienti della regressione, l'aspetto forse più interessante riguarda il segno assunto da PS_{ij} , che risulta negativo e significativo. Mantenendo la chiave interpretativa proposta da Paci e Usai, tale evidenza suggerisce che una maggiore specializzazione produttiva locale nel settore non favorisca l'intensità della presenza di start-up innovative, ma tenda piuttosto ad associarsi a una sua riduzione. L'innovazione imprenditoriale veicolata dalle start-up non necessariamente si sviluppa nei contesti caratterizzati da una più elevata specializzazione produttiva tradizionale. Al contrario, essa sembra emergere più facilmente in ambienti nei quali la nascita di nuove imprese innovative è meno legata alla concentrazione settoriale e maggiormente connessa alla combinazione di risorse, competenze e opportunità che gli indicatori classici di specializzazione non sono in grado di cogliere pienamente.

Anche il coefficiente positivo e significativo di TO_i m è coerente con l'idea che la nascita di start-up innovative sia favorita da settori che, a livello nazionale, presentano un più alto contenuto di opportunità tecnologica.

Il coefficiente negativo di $NORD_j$ merita una riflessione specifica. Condizionatamente alle altre variabili incluse nel modello, gli SLL del Nord mostrano una minore intensità di start-up per addetto rispetto al resto del Paese. Questo risultato, a prima vista controintuitivo se guardato in termini assoluti, è però in realtà coerente con quanto emerge già dalle mappe descrittive: quando si passa dai conteggi assoluti alla misura normalizzata per addetti, diversi hotspot si spostano dal Nord verso il Mezzogiorno e le Isole, mentre molte aree settentrionali perdono centralità relativa proprio perché l'ampia base occupazionale diluisce l'intensità.

Molto debole, invece, è l'evidenza a favore delle variabili che dovrebbero catturare le esternalità di diversità e di base scientifica.

- PD_j ha segno negativo ma non significativo;
- SBS_{ij} è sostanzialmente nullo mentre SBD_{ij} è positivo ma ampiamente non significativo;
- Le due interazioni $PD_j * DM_j$ e $PD_j * DHT_i$ non mostrano alcun effetto statisticamente apprezzabile;

Se nel modello originale di Paci e Usai, la diversità produttiva del sistema locale ha un effetto positivo e significativo sull'innovazione, e tale effetto cresce ulteriormente nei contesti metropolitani e nei settori high-tech. Nel caso delle start-up innovative, invece, il modello non restituisce alcuna evidenza robusta di questo meccanismo.

Questa divergenza non è sorprendente se si considera la natura diversa dell'output. Le start-up innovative rappresentano una forma di innovazione diversa dai brevetti: non misurano soltanto produzione di conoscenza codificata, ma il suo sfruttamento imprenditoriale. È quindi plausibile che i determinanti della loro localizzazione siano meno legati alle esternalità produttive tradizionali e più a fattori come reti finanziarie, qualità istituzionale locale, incubatori, capitale umano imprenditoriale, ecosistemi universitari specifici, servizi avanzati e accesso al mercato. Il fatto che $UNIV_j$ compaia con segno positivo, pur solo marginalmente significativo, va proprio in questa direzione: la presenza universitaria potrebbe avere un ruolo favorevole, ma il suo effetto non emerge in modo abbastanza forte da compensare la generale debolezza della specificazione.

Le specificazioni con interazioni aiutano a precisare ulteriormente questo quadro. Nel modello (3), l'introduzione delle interazioni con la dummy Nord non modifica il risultato principale sulla specializzazione, che rimane negativa e altamente significativa. Le interazioni tra Nord e specializzazione produttiva, da un lato, e tra Nord e diversità produttiva, dall'altro, non risultano però significative. Ne consegue che, nel caso delle start-up innovative, il differenziale territoriale Nord-Centro-Sud non sembra alterare in modo sistematico il modo in cui specializzazione e diversità incidono sulla variabile dipendente. Questo dato è interessante, soprattutto se confrontato con il caso dei brevetti, dove il Nord attenuava parzialmente l'effetto negativo della specializzazione. Qui, invece, la relazione appare più uniforme sul territorio nazionale, segnalando che i meccanismi che governano la nascita di start-up innovative sono meno sensibili alla tradizionale frattura geografica italiana, almeno per come essa è catturata dalla dummy Nord.

Più complesso è il comportamento della diversità produttiva nelle specificazioni con effetti fissi di settore. Nei modelli (4) e (6), PD_j assume coefficiente positivo e statisticamente significativo, mentre l'interazione tra diversità produttiva e settori high-tech è negativa e significativa. Questo è il risultato più interessante dell'intera analisi sulle start-up, perché suggerisce che l'effetto della diversità non sia univoco, ma dipenda dalla natura dei settori considerati. Una volta controllato per l'eterogeneità settoriale media, la diversità produttiva sembra associarsi positivamente alla creazione di start-up nei settori di riferimento; questo effetto si riduce sensibilmente (e in alcuni casi si annulla o si inverte) nei settori ad alta tecnologia. L'interpretazione plausibile è che, fuori dai comparti high-tech, la varietà del contesto locale possa effettivamente favorire la sperimentazione imprenditoriale, offrendo una base più ampia di competenze, relazioni e opportunità. Nei settori high-tech, al contrario, la creazione di nuove imprese innovative potrebbe richiedere ecosistemi più focalizzati, competenze più specialistiche e una maggiore coerenza tecnologica, così che l'eterogeneità del contesto locale non si traduce necessariamente in un vantaggio.

Al contrario, l'interazione tra diversità produttiva e area metropolitana mostra un effetto debole e poco stabile: compare con segno negativo e significatività soltanto nella specificazione più semplice, mentre perde rilevanza nelle altre. Ne consegue che non vi è evidenza robusta del fatto che la condizione metropolitana amplifichi o modifichi in maniera sistematica il legame tra varietà produttiva e start-up innovative. Anche la presenza di università presenta un profilo non pienamente stabile: nei modelli più semplici il coefficiente è positivo e debolmente significativo, ma perde significatività una volta introdotti controlli più stringenti. Ciò suggerisce che il ruolo delle università, pur plausibilmente importante sul piano sostanziale, non emerga in modo robusto in questa specificazione lineare, forse perché già assorbito da altre caratteristiche territoriali o perché opera attraverso canali indiretti non pienamente catturati dal modello.

Un ulteriore aspetto da sottolineare riguarda la dummy Nord, che nei modelli in cui è inclusa direttamente entra con coefficiente negativo e significativo. Questo risultato segnala che, a parità delle altre condizioni, i sistemi locali del Nord mostrano una minore intensità di start-up innovative rispetto alla categoria di riferimento. Si tratta di un esito che può apparire controintuitivo, ma che non va letto come indicazione di una minore capacità innovativa del Nord in senso assoluto. Piuttosto, esso suggerisce che la creazione di start-up innovative non coincide necessariamente con i territori tradizionalmente più forti sul piano industriale e brevettuale. In altri termini, la geografia dell'imprenditorialità innovativa può differire da quella dell'innovazione brevettata, riflettendo logiche differenti, in cui contano non solo le basi produttive esistenti, ma anche fattori istituzionali, incentivi specifici, condizioni di accesso al mercato e traiettorie più recenti di sviluppo locale.

Nel complesso, i risultati relativi alle start-up innovative conducono a una conclusione piuttosto chiara. La specializzazione produttiva locale non appare come un ambiente favorevole alla nascita di nuova imprenditorialità innovativa al contrario, il suo effetto è sistematicamente negativo e molto robusto. La

diversità produttiva, invece, non esercita un'influenza univoca: il suo impatto dipende dalla specificazione adottata e, soprattutto, dalla natura dei settori considerati, diventando più favorevole soltanto in alcune configurazioni e al di fuori dei comparti high-tech.

Le variabili science-based non offrono evidenze robuste, mentre la geografia Nord-Sud sembra incidere meno sui meccanismi strutturali dell'innovazione imprenditoriale rispetto a quanto accade nel caso dei brevetti. Ne emerge un quadro coerente con l'idea che le start-up innovative rispondano a logiche parzialmente diverse da quelle che governano l'attività brevettuale: più che radicarsi nei contesti maggiormente specializzati, esse sembrano trovare spazio in ambienti in cui l'innovazione nasce da combinazioni meno standardizzate di risorse, opportunità e competenze.

6. Conclusioni

L'analisi svolta in questa tesi muove da un presupposto teorico preciso: l'innovazione non costituisce un fenomeno unitario e, di conseguenza, non può essere osservata in modo soddisfacente attraverso un solo indicatore.

Per questa ragione il lavoro ha adottato una strategia di misurazione plurima, costruendo tre indicatori distinti di produttività dell'innovazione, basati rispettivamente sui brevetti UIBM, sui brevetti EPO e sulle start up innovative, tutti normalizzati rispetto al numero di addetti e attribuiti alla scala dei Sistemi Locali del Lavoro. Nel capitolo 3 si è mostrato come tale scelta non risponda soltanto a un'esigenza descrittiva, ma rifletta una distinzione sostanziale tra tre forme di output innovative solo parzialmente sovrapponibili: i brevetti UIBM colgono una capacità inventiva più diffusa e maggiormente radicata nel tessuto produttivo nazionale, i brevetti EPO intercettano attività inventive più selettive e più vicine alla frontiera tecnologica, mentre le start up innovative restituiscono la dimensione imprenditoriale dell'innovazione, cioè la sua traduzione in iniziativa economica. La scelta degli SLL, inoltre, è coerente con l'ipotesi che gli spillover di conoscenza operino entro mercati del lavoro locali integrati e non semplicemente entro confini amministrativi.

La prima conclusione che emerge dal lavoro riguarda la geografia dell'innovazione italiana. L'evidenza cartografica e quella spaziale suggeriscono che l'innovazione, quando è misurata in termini di intensità rispetto alla base occupazionale, non si distribuisce secondo un unico pattern territoriale compatto. Il capitolo 3 mostra infatti che la normalizzazione per addetti modifica in misura sostanziale la geografia dei fenomeni innovativi, attenuando gli effetti di scala e facendo emergere anche sistemi locali di dimensione intermedia o periferica. Nel caso dei brevetti EPO l'intensità innovativa risulta fortemente concentrata e asimmetrica, con pochi territori che si collocano nella coda superiore della distribuzione e numerosi SLL con valori molto bassi o nulli; nel caso delle start up, la distribuzione è ancora più rarefatta, con molti zeri e pochi nuclei locali. In entrambi i casi, e in misura plausibilmente analoga anche per UIBM, il quadro che emerge non è quello di grandi blocchi territoriali continui, ma di una geografia fatta di poli, discontinuità, cluster locali e outlier. Ciò porta a leggere la distribuzione dell'innovazione italiana come più policentrica e frammentata di quanto una semplice analisi per conteggi assoluti lascerebbe intendere.

Su questo sfondo, il risultato empirico più robusto riguarda il ruolo della specializzazione produttiva locale. In tutte e tre le famiglie di modelli stimati, la specializzazione presenta un coefficiente negativo e statisticamente significativo nelle principali specificazioni. Il dato è particolarmente rilevante perché si discosta dall'aspettativa più strettamente marshalliana, secondo cui la concentrazione territoriale di attività appartenenti allo stesso settore dovrebbe favorire lo scambio di conoscenze, la circolazione di competenze e, quindi, l'innovazione. I risultati di questa tesi suggeriscono invece che la co-localizzazione produttiva, almeno quando l'output innovativo è misurato in termini di intensità per addetto, non coincide automaticamente con le condizioni più favorevoli alla generazione di innovazione. Una possibile interpretazione è che i territori fortemente specializzati possano essere molto efficienti sul piano produttivo o manifatturiero, ma non per questo particolarmente favorevoli alla ricombinazione di conoscenze, alla varietà degli stimoli cognitivi o all'emersione di traiettorie innovative nuove. Nel caso delle variabili normalizzate, inoltre, è plausibile che nei contesti a forte specializzazione il denominatore occupazionale cresca più rapidamente del numeratore innovativo, contribuendo così al segno negativo stimato. Si tratta, in ogni caso, del risultato più stabile dell'intero impianto empirico.

Più complessa è, invece, la lettura della diversità produttiva. L'analisi non fornisce una conferma generalizzata dell'ipotesi jacobsoniana nella sua formulazione più lineare. Nei modelli EPO la diversità

non emerge come un determinante positivo stabile e generalizzabile. Quando compaiono segnali favorevoli, essi sono circoscritti a specificazioni particolari e non sufficientemente robusti da consentire conclusioni forti. Nei modelli UIBM la diversità tende più spesso ad assumere segno negativo, o comunque a non esprimere un contributo univocamente favorevole. Nelle start up innovative il quadro è ancora più selettivo: la diversità risulta positiva solo in alcune specificazioni con effetti fissi settoriali, mentre nei settori high tech la relativa interazione assume segno negativo.

Nel complesso, l'evidenza induce a ritenere che la varietà produttiva non operi come meccanismo automatico di generazione dell'innovazione. Più che la diversità in sé, sembra contare la qualità economica e cognitiva delle combinazioni tra attività, cioè la misura in cui la compresenza di settori diversi si traduce in complementarità effettive e non in semplice eterogeneità. In questo senso, i risultati sono più vicini alle letture successive della letteratura, che insistono sul carattere condizionato degli effetti di specializzazione e diversificazione, che non a una contrapposizione rigida tra modello MAR e modello Jacobs.

La comparazione tra i tre indicatori di output permette di qualificare ulteriormente questa conclusione generale. I brevetti EPO, per la loro maggiore selettività e per il loro più stretto legame con attività inventive ad alto contenuto tecnologico, mostrano una relazione particolarmente netta con le opportunità tecnologiche settoriali e una minore aderenza alle forme classiche delle esternalità territoriali. In questo caso, il territorio continua a contare, ma in modo più debole, più selettivo e meno lineare. I brevetti UIBM, invece, sembrano essere spiegati meglio dalla struttura settoriale e territoriale del sistema produttivo locale. Qui emerge con particolare chiarezza il ruolo della componente science based. La specializzazione science based tende ad associarsi positivamente alla produttività brevettuale, mentre la diversità interna ai cluster science based mostra un'associazione negativa relativamente stabile. Il risultato suggerisce che, per la brevettazione nazionale, una base cognitiva coerente e addensata possa risultare più favorevole di una semplice dispersione di attività scientifiche e tecnologiche. Le start up innovative, infine, seguono una logica in parte diversa. Pur mostrando anch'esse un coefficiente negativo della specializzazione produttiva, presentano una capacità esplicativa complessiva molto più bassa e non restituiscono evidenze altrettanto solide per la componente science based. Questo dato è coerente con la natura dell'indicatore: la nascita di start up innovative non misura soltanto la produzione di conoscenza, ma la sua trasformazione in iniziativa economica, e può quindi dipendere da una costellazione di fattori che vanno oltre la struttura produttiva locale in senso stretto.

Tra le variabili esplicative, le opportunità tecnologiche settoriali rappresentano la dimensione più coerente con le attese teoriche. Nei modelli brevettuali, e in misura più debole ma comunque leggibile anche nelle start up, esse mostrano un'associazione positiva e robusta con l'output innovativo. Si tratta di un risultato importante, perché suggerisce che la geografia dell'innovazione non possa essere letta indipendentemente dalla struttura tecnologica dei settori che compongono i territori. Alcuni comparti sono intrinsecamente più fertili di altri sotto il profilo della produzione di conoscenza brevettabile o della generazione di nuova imprenditorialità innovativa. La concentrazione geografica dell'innovazione non dipende soltanto dalla localizzazione della produzione, ma anche dalla natura tacita, cumulativa e knowledge intensive delle attività economiche. In questo senso, i risultati della tesi sono coerenti con quanto osservato da Audretsch e Feldman e, più in generale, con l'idea che il territorio conti soprattutto quando ospita attività inserite in settori tecnologicamente più promettenti.

Il confronto con Paci e Usai è particolarmente istruttivo il loro lavoro infatti rappresenta un riferimento fondamentale per la geografia dell'innovazione italiana. Essi mostrano che l'attività innovativa tende a concentrarsi in specifiche aree del Paese seguendo pattern spaziali non casuali e che sia la specializzazione sia la diversificazione possono esercitare un'influenza positiva sull'innovazione locale, con una maggiore rilevanza della varietà nei settori high tech e nelle aree metropolitane. I risultati

ottenuti in questa tesi si collocano in un punto di confronto critico rispetto a quel quadro. La centralità della dimensione territoriale non viene smentita, ma la sua traduzione empirica risulta meno lineare. Né la specializzazione né la diversità emergono qui come meccanismi generalizzati e stabilmente favorevoli; la componente spaziale appare più frammentata e meno uniforme; l'effetto territoriale, quando presente, sembra essere filtrato dalla natura dell'output osservato, dalla struttura tecnologica dei settori e dalla configurazione della base conoscitiva locale. Questa divergenza deve essere però inquadrata nel contesto di questa analisi che è stata condotta lavorando su una geografia più aggiornata, su un quadro informativo più articolato e su tre proxy differenti dell'output innovativo, calcolate in modo diverso rispetto a quanto prodotto da Paci ed Usai.

Anche l'eterogeneità territoriale italiana richiede una lettura più sfumata di quella fornita dalla semplice opposizione Nord Sud. Il capitolo 3 mostra con chiarezza che la struttura occupazionale resta fortemente polarizzata verso il Nord e che, in termini di massa innovativa, i sistemi locali settentrionali mantengono un vantaggio evidente. Tuttavia, quando si passa agli indicatori normalizzati e ai modelli di regressione, il quadro si fa più complesso. Nei modelli EPO e UIBM, l'interazione tra Nord e specializzazione produttiva è positiva e significativa, il che suggerisce che nei sistemi locali settentrionali l'effetto negativo della specializzazione si attenua. Nelle start up, invece, questa modulazione territoriale è molto meno chiara. Ciò induce a ritenere che il Nord non sia semplicemente più innovativo in senso assoluto, ma rappresenti un contesto in cui alcune configurazioni produttive risultano meno penalizzanti, probabilmente per la presenza di reti industriali più dense, infrastrutture migliori, maggiore prossimità a fonti di conoscenza e più elevata qualità istituzionale. Allo stesso tempo, l'analisi spaziale mostra che l'innovazione relativa può emergere anche in sistemi locali non appartenenti ai poli tradizionali, compresi alcuni territori del Mezzogiorno e delle Isole. Ne deriva una geografia meno dicotomica e più articolata, nella quale gerarchie consolidate e nuclei locali dinamici coesistono senza necessariamente trasformarsi in processi diffusi di convergenza territoriale.

Nel loro insieme, i risultati della tesi suggeriscono dunque che la relazione tra struttura produttiva e innovazione non possa essere interpretata in modo meccanico. Non emerge un unico modello territoriale dell'innovazione valido per tutti gli output considerati. La specializzazione appare sistematicamente associata in modo negativo all'intensità innovativa; la diversità non costituisce una garanzia generalizzata di maggiore innovazione; la struttura science based mostra effetti selettivi, più leggibili nella brevettazione UIBM che non nei brevetti EPO o nelle start up; la dimensione territoriale, infine, continua a essere rilevante, ma attraverso meccanismi che dipendono dalla natura dell'output osservato, dalla composizione settoriale e dalla qualità delle complementarità cognitive e organizzative. Da questo punto di vista, la tesi si colloca in continuità con gli sviluppi più recenti della letteratura, che insistono sulla relatedness, sulla coerenza cognitiva e sulla natura cumulativa e path dependent dell'innovazione, più che su una semplice contrapposizione tra specializzazione marshalliana e varietà jacobsoniana.

6.1 Limiti dello studio e sviluppi futuri

I risultati ottenuti devono essere interpretati tenendo conto di alcuni limiti che definiscono il perimetro inferenziale del lavoro. Il primo riguarda la natura degli indicatori di output. Brevetti EPO, brevetti UIBM e start up innovative non rappresentano tre misure omogenee dello stesso fenomeno, ma tre manifestazioni diverse dell'innovazione. Questo costituisce, da un lato, un punto di forza dell'analisi, perché consente un confronto più ricco; dall'altro, impone cautela, poiché le differenze tra modelli non possono essere lette come semplici variazioni quantitative di uno stesso oggetto empirico.

A ciò si aggiunge una differenza non trascurabile nei criteri di georeferenziazione: i brevetti EPO sono attribuiti all'indirizzo dell'inventore, i brevetti UIBM all'indirizzo dell'applicant o alla sede legale, le

start up alla sede legale. Si tratta di scelte coerenti con la qualità delle fonti disponibili, ma che introducono inevitabilmente differenze tra luogo osservato e luogo effettivo di generazione o internalizzazione dell'innovazione.

Un secondo limite riguarda la copertura del campione e la costruzione delle variabili. La normalizzazione per addetti è teoricamente appropriata, perché consente confronti territoriali depurati dagli effetti dimensionali, ma rende l'analisi sensibile alla qualità del denominatore occupazionale e ai processi di armonizzazione territoriale. Inoltre, le variabili science based restringono in misura significativa il campione disponibile e il confronto tra modelli non avviene sempre su uno stesso insieme di osservazioni. Anche la costruzione delle opportunità tecnologiche settoriali per i brevetti richiede procedure di raccordo tra classificazioni tecnologiche e classificazioni settoriali, con perdite informative contenute ma non nulle. Di conseguenza, la variazione dei coefficienti tra una specificazione e l'altra non va interpretata soltanto come effetto della diversa formulazione econometrica, ma anche come conseguenza del diverso campione effettivamente utilizzato.

Un terzo limite è di natura propriamente econometrica. Le stime utilizzano modelli lineari OLS con errori robusti, scelta che garantisce trasparenza interpretativa e confrontabilità tra specificazioni, ma che non esaurisce la complessità dei fenomeni osservati. In particolare, nel caso delle start up il basso \widehat{R}^2 segnala che una quota molto rilevante della variabilità dipende verosimilmente da fattori non osservati, quali qualità dell'ecosistema imprenditoriale, accesso al capitale di rischio, ruolo degli incubatori, caratteristiche dei fondatori, servizi avanzati e governance locale. Anche nei modelli brevettuali, i coefficienti devono essere letti come associazioni parziali e non come relazioni causali pienamente identificate. Per questa ragione, i risultati ottenuti definiscono una geografia economica plausibile dell'innovazione, ma non autorizzano inferenze causali forti sui meccanismi sottostanti.

Da questi limiti discendono alcune linee di sviluppo particolarmente promettenti. Una prima direzione consiste nell'estendere l'analisi in chiave dinamica, attraverso dati panel o finestre temporali più lunghe, così da distinguere meglio i fattori strutturali persistenti dai mutamenti nelle traiettorie territoriali dell'innovazione. Una seconda riguarda l'introduzione di misure più dirette di relatedness produttiva e tecnologica. I risultati della tesi suggeriscono infatti che non sia la varietà in sé a contare, ma la qualità delle complementarità tra attività locali. In questo senso, una misura di diversità correlata potrebbe risultare più informativa della sola diversità generale. Una terza direzione di ricerca appare particolarmente rilevante per il caso delle start up, dove l'integrazione di informazioni su università, incubatori, venture capital, servizi digitali e qualità istituzionale locale potrebbe migliorare in modo sostanziale la capacità esplicativa dei modelli. Infine, meriterebbe un approfondimento specifico il caso dei territori meridionali e insulari che emergono come hotspot relativi nelle mappe di intensità. In questi contesti, la questione non riguarda soltanto l'ampiezza del ritardo innovativo, ma anche le condizioni in cui nuclei locali di capacità innovativa riescono a emergere pur in assenza di un diffuso processo di trascinamento territoriale.

6.2 Implicazione di policy

I risultati invitano a una riflessione particolarmente rilevante per il Mezzogiorno. L'evidenza empirica non suggerisce che il rafforzamento della sola specializzazione produttiva costituisca, di per sé, una strategia sufficiente a sostenere l'innovazione territoriale. Al contrario, il lavoro mostra come né la concentrazione settoriale né una diversificazione intesa in senso generico producano automaticamente esiti favorevoli. Ciò vale a maggior ragione in contesti, come quelli meridionali, caratterizzati da una minore densità produttiva media, da reti innovative più fragili e da una più debole capacità di trasformare episodi locali di dinamismo in processi cumulativi di sviluppo. In questa prospettiva, la questione non sembra essere quella di replicare i modelli dei territori più forti, né quella di inseguire

indiscriminatamente una maggiore specializzazione, ma piuttosto di costruire le condizioni affinché le capacità innovative già presenti possano consolidarsi, coordinarsi e diffondersi. Per il Mezzogiorno, ciò implica verosimilmente una politica territoriale più selettiva e maggiormente orientata alla qualità delle connessioni: rafforzare i legami tra università, imprese e servizi avanzati, sostenere i nodi locali che mostrano potenzialità innovative, soprattutto quando emergono in termini di intensità e non di sola massa, migliorare l'accesso a infrastrutture, competenze manageriali e strumenti finanziari coerenti con la crescita di attività innovative. In altri termini, se il Mezzogiorno non appare privo di capacità innovative, ma piuttosto segnato da una loro distribuzione discontinua e spesso poco connessa, l'obiettivo dovrebbe essere meno quello di creare ex novo specializzazioni e più quello di rendere persistenti e territorialmente trasmissibili i processi innovativi già esistenti. È probabilmente in questa direzione che una politica place-based può risultare più efficace: non limitandosi a compensare un divario, ma cercando di trasformare nuclei locali ancora episodici in strutture più stabili di produzione, circolazione e valorizzazione della conoscenza.

Bibliografia

- Acs, Z. J., & Audretsch, D. B. (1989). Patents as a measure of innovative activity. *Kyklos*, 42(2), 171–180.
- Acs, Z. J., Anselin, L., & Varga, A. (2002). Patents and innovation counts as measures of regional production of new knowledge. *Research Policy*, 31(7), 1069–1085.
- Acs, Z. J., Audretsch, D. B., & Feldman, M. P. (1992). Real effects of academic research: Comment. *American Economic Review*, 82(1), 363–367.
- Almeida, P., & Kogut, B. (1999). Localization of knowledge and the mobility of engineers in regional networks. *Management Science*, 45(7), 905–917.
- Anselin, L. (1988). *Spatial econometrics: Methods and models*. Dordrecht: Kluwer Academic Publishers.
- Anselin, L. (1995). Local indicators of spatial association - LISA. *Geographical Analysis*, 27(2), 93–115.
- Anselin, L., Varga, A., & Acs, Z. J. (1997). Local geographic spillovers between university research and high technology innovations. *Journal of Urban Economics*, 42(3), 422–448.
- Anselin, L., Varga, A., & Acs, Z. J. (2000). Geographic and sectoral characteristics of academic knowledge externalities. *Papers in Regional Science*, 79(4), 435–443.
- Arrow, K. J. (1962). The economic implications of learning by doing. *Review of Economic Studies*, 29(3), 155–173.
- Arthur, W. B. (1988). Self-reinforcing mechanisms in economics. In P. W. Anderson, K. J. Arrow, & D. Pines (Eds.), *The economy as an evolving complex system* (pp. 9–31). Redwood City, CA: Addison-Wesley.
- Asheim, B. T., & Gertler, M. S. (2005). The geography of innovation: Regional innovation systems. In J. Fagerberg, D. C. Mowery, & R. R. Nelson (Eds.), *The Oxford handbook of innovation* (pp. 291–317). Oxford: Oxford University Press.
- Audretsch, D. B. (1995). *Innovation and industry evolution*. Cambridge, MA: MIT Press.
- Audretsch, D. B., & Feldman, M. P. (1996). R&D spillovers and the geography of innovation and production. *American Economic Review*, 86(3), 630–640.
- Audretsch, D. B., & Feldman, M. P. (1999). Innovation in cities: Science-based diversity, specialization, and localized competition. *European Economic Review*, 43(2), 409–429.
- Audretsch, D. B., & Feldman, M. P. (2004). Knowledge spillovers and the geography of innovation. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2713–2739). Amsterdam: Elsevier.
- Audretsch, D. B., & Stephan, P. E. (1996). Company-scientist locational links: The case of biotechnology. *American Economic Review*, 86(3), 641–652.
- Balland, P.-A. (2012). Proximity and the evolution of collaboration networks: Evidence from research and development projects within the global navigation satellite system industry. *Regional Studies*, 46(6), 741–756.
- Balland, P.-A. (2016). Relatedness and the geography of innovation. In R. Shearmur, C. Carrincazeaux, & D. Doloreux (Eds.), *Handbook on the geographies of innovation* (pp. 127–149). Cheltenham: Edward Elgar.

- Balland, P.-A., & Rigby, D. L. (2017). The geography of complex knowledge. *Economic Geography*, 93(1), 1–23.
- Balland, P.-A., de Vaan, M., & Boschma, R. (2013). The dynamics of interfirm networks along the industry life cycle: The case of the global video game industry, 1987–2007. *Journal of Economic Geography*, 13(5), 741–765.
- Balland, P.-A., Rigby, D., & Boschma, R. (2015). The technological resilience of U.S. cities. *Cambridge Journal of Regions, Economy and Society*, 8(2), 167–184.
- Baptista, R., & Swann, P. (1998). Do firms in clusters innovate more? *Research Policy*, 27(5), 525–540.
- Baptista, R., & Swann, P. (1999). A comparison of clustering dynamics in the U.S. and U.K. computer industries. *Journal of Evolutionary Economics*, 9(3), 373–399.
- Barca, F., McCann, P., & Rodríguez-Pose, A. (2012). The case for regional development intervention: Place-based versus place-neutral approaches. *Journal of Regional Science*, 52(1), 134–152.
- Bartelsman, E. J., Caballero, R. J., & Lyons, R. K. (1994). Customer- and supplier-driven externalities. *American Economic Review*, 84(4), 1075–1084.
- Beaudry, C., & Schiffauerova, A. (2009). Who's right, Marshall or Jacobs? The localization versus urbanization debate. *Research Policy*, 38(2), 318–337.
- Becattini, G. (Ed.). (1987). *Mercato e forze locali: Il distretto industriale*. Bologna: Il Mulino.
- Boschma, R. A. (2004). Competitiveness of regions from an evolutionary perspective. *Regional Studies*, 38(9), 1001–1014.
- Boschma, R. A. (2005). Proximity and innovation: A critical assessment. *Regional Studies*, 39(1), 61–74.
- Boschma, R. A., & Frenken, K. (2006). Why is economic geography not an evolutionary science? Towards an evolutionary economic geography. *Journal of Economic Geography*, 6(3), 273–302.
- Boschma, R., Balland, P.-A., & Kogler, D. F. (2015). Relatedness and technological change in cities: The rise and fall of technological knowledge in U.S. metropolitan areas from 1981 to 2010. *Industrial and Corporate Change*, 24(1), 223–250.
- Boschma, R., Minondo, A., & Navarro, M. (2013). The emergence of new industries at the regional level in Spain: A proximity approach based on product relatedness. *Economic Geography*, 89(1), 29–51.
- Bottazzi, L., & Peri, G. (2003). Innovation and spillovers in regions: Evidence from European patent data. *European Economic Review*, 47(4), 687–710.
- Breschi, S. (1997). *The geography of innovation: A cross-section analysis*. CESPRI Working Paper No. 95. Milano: Università Bocconi.
- Breschi, S., & Lissoni, F. (2002). Mobility and social networks: Localized knowledge spillovers revisited. Paper presented at the workshop *Clusters in High-Technology: Aerospace, Biotechnology and Software Compared*, Montréal.
- Breschi, S., & Lissoni, F. (2009). Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography*, 9(4), 439–468.
- Brouwer, E., Budil-Nadvornikova, H., & Kleinknecht, A. (1999). Are urban agglomerations a better breeding place for product innovation? *Regional Studies*, 33(6), 541–549.
- Brusco, S. (1982). The Emilian model: Productive decentralisation and social integration. *Cambridge Journal of Economics*, 6(2), 167–184.
- Brühlhart, M. (1998). Economic geography, industry location and trade: The evidence. *World Economy*, 21(6), 775–801.

- Buzzacchi, L., De Marco, A., & Pagnini, M. (2024). Agglomeration and the Italian North–South divide. *Journal of Economic Geography*, 24(5), 707–728
- Camagni, R. (Ed.). (1991). *Innovation networks: Spatial perspectives*. London: Belhaven Press.
- Capello, R. (1999). Spatial transfer of knowledge in high technology milieux: Learning versus collective learning processes. *Regional Studies*, 33(4), 353–365.
- Carlino, G. A., Chatterjee, S., & Hunt, R. M. (2007). Urban density and the rate of invention. *Journal of Urban Economics*, 61(3), 389–419.
- Castaldi, C., Frenken, K., & Los, B. (2015). Related variety, unrelated variety and technological breakthroughs: An analysis of U.S. state-level patenting. *Regional Studies*, 49(5), 767–781.
- Castellano, R., Punzo, G., Scandurra, G., & Thomas, A. (2023). Agglomeration externalities and innovation in Europe: The role of migration and human capital. *Journal of the Knowledge Economy*, 14, 1–29.
- Cingano, F., & Schivardi, F. (2004). Identifying the sources of local productivity growth. *Journal of the European Economic Association*, 2(4), 720–742.
- Coe, D. T., & Helpman, E. (1995). International R&D spillovers. *European Economic Review*, 39(5), 859–887.
- Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, 35(1), 128–152.
- Combes, P.-P. (2000). Economic structure and local growth: France, 1984–1993. *Journal of Urban Economics*, 47(3), 329–355.
- Combes, P.-P., Magnac, T., & Robin, J.-M. (2004). The dynamics of local employment in France. *Journal of Urban Economics*, 56(2), 217–243.
- Cooke, P. (1991). Regional innovation systems, clusters, and the knowledge economy. *Industrial and Corporate Change*, 10(4), 945–974.
- Cooke, P., & Morgan, K. (1998). *The associational economy: Firms, regions, and innovation*. Oxford: Oxford University Press.
- Deidda, S., Paci, R., & Usai, S. (2006). Spatial externalities and local economic growth. *Contributi di Ricerca CRENoS*, 02/06.
- Duranton, G., & Overman, H. G. (2005). Testing for localization using micro-geographic data. *Review of Economic Studies*, 72(4), 1077–1106.
- Duranton, G., & Puga, D. (2000). Diversity and specialisation in cities: Why, where and when does it matter? *Urban Studies*, 37(3), 533–555.
- Duranton, G., & Puga, D. (2001). Nursery cities. *American Economic Review*, 91(5), 1454–1477.
- Duranton, G., & Puga, D. (2004). Micro-foundations of urban agglomeration economies. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2063–2117). Amsterdam: Elsevier.
- Ejeremo, O. (2005). Technological diversity and Jacobs' externality hypothesis revisited. *Growth and Change*, 36(2), 167–195.
- Ellison, G., & Glaeser, E. L. (1999). The geographic concentration of industry: Does natural advantage explain agglomeration? *American Economic Review*, 89(2), 311–316.
- Evenson, R. E. (1993). Patents, R&D and invention potential: International evidence. *American Economic Review Papers and Proceedings*, 83(2), 463–468.
- Feldman, M. P. (1994). *The geography of innovation*. Dordrecht: Kluwer Academic Publishers.

- Feldman, M. P., & Audretsch, D. B. (1999). Innovation in cities: Science-based diversity, specialization and localized competition. *European Economic Review*, 43(2), 409–429.
- Figueiredo, O., Guimarães, P., & Woodward, D. (2015). Spatial proximity and knowledge spillovers: Evidence from patent data. *Journal of Urban Economics*, 89, 21–31.
- Forni, M., & Paba, S. (2002). Spillovers and the growth of local industries. *Journal of Industrial Economics*, 50(2), 151–171.
- Frenken, K., van Oort, F., & Verburg, T. (2007). Related variety, unrelated variety and regional economic growth. *Regional Studies*, 41(5), 685–697.
- Glaeser, E. L., Kallal, H. D., Scheinkman, J. A., & Shleifer, A. (1992). Growth in cities. *Journal of Political Economy*, 100(6), 1126–1152.
- Griliches, Z. (1979). Issues in assessing the contribution of research and development to productivity growth. *Bell Journal of Economics*, 10(1), 92–116.
- Griliches, Z. (1992). The search for R&D spillovers. *Scandinavian Journal of Economics*, 94, S29–S47.
- Guimarães, P., & Portugal, P. (2010). A simple feasible procedure to fit models with high-dimensional fixed effects. *Stata Journal*, 10(4), 628–649.
- Hall, B. H., Jaffe, A. B., & Trajtenberg, M. (2001). The NBER patent citations data file: Lessons, insights and methodological tools. NBER Working Paper No. 8498.
- Harrison, B., Kelley, M. R., & Gant, J. (1996). Specialization versus diversity in local economies: The implications for innovative private-sector behavior. *Cityscape*, 2(2), 61–93.
- Harrison, B., Kelley, M. R., & Gant, J. (1997). Innovative firm behavior and local milieu: Exploring the intersection of agglomeration, firm effects, and technological change. *Economic Geography*, 72(3), 233–258.
- Henderson, V. (1986). Efficiency of resource usage and city size. *Journal of Urban Economics*, 19(1), 47–70.
- Henderson, V. (1997). Externalities and industrial development. *Journal of Urban Economics*, 42(3), 449–470.
- Henderson, V. (2003). Marshall's scale economies. *Journal of Urban Economics*, 53(1), 1–28.
- Henderson, V., Kuncoro, A., & Turner, M. (1995). Industrial development in cities. *Journal of Political Economy*, 103(5), 1067–1090.
- Hidalgo, C. A., Klinger, B., Barabási, A.-L., & Hausmann, R. (2007). The product space conditions the development of nations. *Science*, 317(5837), 482–487.
- Jacobs, J. (1969). *The economy of cities*. New York: Random House.
- Jaffe, A. B. (1989). Real effects of academic research. *American Economic Review*, 79(5), 957–970.
- Jaffe, A. B., & Trajtenberg, M. (2002). *Patents, citations, and innovations: A window on the knowledge economy*. Cambridge, MA: MIT Press.
- Jaffe, A. B., Trajtenberg, M., & Fogarty, M. S. (2000). Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, 90(2), 215–218.
- Jaffe, A. B., Trajtenberg, M., & Henderson, R. (1993). Geographic localization of knowledge spillovers as evidenced by patent citations. *Quarterly Journal of Economics*, 108(3), 577–598.
- Kelley, M. R., & Hageman, A. M. (1999). Marshallian externalities in innovation. *Journal of Economic Growth*, 4(1), 39–54.
- Kogler, D. F., Rigby, D. L., & Tucker, I. (2013). Mapping knowledge space and technological relatedness in U.S. cities. *European Planning Studies*, 21(9), 1374–1391.

- Krugman, P. (1991). *Geography and trade*. Cambridge, MA: MIT Press.
- Lamorgese, A. R. (1997). *Externalities, economic geography and growth: A cross-section analysis*. CESPRI Working Paper No. 100.
- Lawson, C., & Lorenz, E. (1999). Collective learning, tacit knowledge and regional innovative capacity. *Regional Studies*, 33(4), 305–317.
- Lucas, R. E. (1988). On the mechanics of economic development. *Journal of Monetary Economics*, 22(1), 3–42.
- Marshall, A. (1890). *Principles of economics*. London: Macmillan.
- Maskell, P., & Malmberg, A. (1999). Localised learning and industrial competitiveness. *Cambridge Journal of Economics*, 23(2), 167–185.
- Neffke, F., Henning, M., & Boschma, R. (2011). How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography*, 87(3), 237–265.
- Paci, R., & Usai, S. (1999). The role of specialisation and diversity externalities in the agglomeration of innovative activities. Working Paper, University of Cagliari and CRENoS.
- Paci, R., & Usai, S. (2000). Externalities, knowledge spillovers and the spatial distribution of innovation. *GeoJournal*, 49(4), 381–390.
- Paci, R., & Usai, S. (2005). Agglomeration economies and growth in Italian local labour systems, 1991–2001. Working Paper, CRENoS.
- Porter, M. E. (1990). *The competitive advantage of nations*. London: Macmillan.
- Porter, M. E. (2003). The economic performance of regions. *Regional Studies*, 37(6–7), 549–578.
- Rigby, D. L. (2015). Technological relatedness and knowledge space: Entry and exit of U.S. cities from patent classes. *Regional Studies*, 49(11), 1922–1937.
- Rigby, D. L., & Essletzbichler, J. (2002). Agglomeration economies and productivity differences in U.S. cities. *Journal of Economic Geography*, 2(4), 407–432.
- Romer, P. M. (1986). Increasing returns and long-run growth. *Journal of Political Economy*, 94(5), 1002–1037.
- Romer, P. M. (1990). Endogenous technological change. *Journal of Political Economy*, 98(5), S71–S102.
- Rosenthal, S. S., & Strange, W. C. (2003). Geography, industrial organization, and agglomeration. *Review of Economics and Statistics*, 85(2), 377–393.
- Rosenthal, S. S., & Strange, W. C. (2004). Evidence on the nature and sources of agglomeration economies. In J. V. Henderson & J.-F. Thisse (Eds.), *Handbook of regional and urban economics* (Vol. 4, pp. 2119–2171). Amsterdam: Elsevier.
- Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics*, 88(4), 641–658.
- Saxenian, A. (1994). *Regional advantage: Culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.
- Shefer, D., & Frenkel, A. (1998). Local milieu and innovations: Some empirical results. *Annals of Regional Science*, 32(1), 185–200.
- Swann, G. M. P., & Prevezer, M. (1996). A comparison of the dynamics of industrial clustering in computing and biotechnology. *Research Policy*, 25(7), 1139–1157.

- Swann, G. M. P., Prevezer, M., & Stout, D. (Eds.). (1998). *The dynamics of industrial clusters: International comparisons in computing and biotechnology*. Oxford: Oxford University Press.
- Torre, A., & Rallet, A. (2005). Proximity and localization. *Regional Studies*, 39(1), 47–59.
- Usai, S., & Paci, R. (2003). Externalities and local economic growth in manufacturing industries. In B. Fingleton (Ed.), *European regional growth* (pp. 229–251). Berlin: Springer.
- van Oort, F. G. (2002). *Agglomeration, economic growth and innovation: Spatial analysis of growth and R&D externalities in the Netherlands*. Rotterdam: Tinbergen Institute Research Series.
- Wang, M.-C., Chen, P.-C., & Fang, S.-C. (2018). A critical view of knowledge networks and innovation performance: The mediation role of firms' knowledge integration capability. *Journal of Business Research*, 88, 222–233.