Finanziato dall'Unione europea
NextGenerationEU
Ministero dell'Università e della Ricerca
Italiadomani
PIANO NAZIONALE DI RIPRESA E RESILIENZA
GRINS FOUNDATION

GRINS FOUNDATION

Discussion Paper Series

Valutazione dei divari educativi tra le regioni italiane mediante la stima per piccole aree

Discussion paper n. 33/2025

Diego Battagliese (a), Alessio Pollice (a): (a) Università degli Studi di Bari Aldo Moro

Fondazione GRINS - Galleria Ugo Bassi 1, 40121, Bologna, IT - C.F/P.IVA 91451720378
info@grins.it | comunicazione@grinsfoundation.it | fondazionegrins@pec.grins.it | grins.it

ISSN 3035-5567

Valutazione dei divari educativi tra le regioni italiane mediante la stima per piccole aree

Diego Battagliese (a), Alessio Pollice (a): (a) Università degli Studi di Bari Aldo Moro

Discussion Paper DP n° 33/2025

Una delle principali sfide con cui si confrontano i decisori pubblici con riferimento al servizio scolastico è quella di ridurre i divari fra studenti all'interno di una stessa scuola, ma ancor di più fra tutti gli studenti su base nazionale. I dati dell'INVALSI (Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e Formazione) costituiscono la fonte informativa primaria per valutare tali divari. In questo lavoro, ci avvaliamo dei dati dell'indagine campionaria su scala regionale che vengono distribuiti liberamente dall'INVALSI. Il nostro obiettivo è quello di pervenire a stime più attendibili rispetto a quelle fornite dall'Istituto stesso. Per fare ciò, sfruttiamo i modelli small area di tipo area-level, ossia modelli lineari ad effetti misti, costruti su due livelli, in cui il dato INVALSI, che costituisce lo stimatore diretto, viene combinato con uno stimatore sintetico proveniente da modello, in modo da dare più vigore allo stimatore diretto per mezzo di variabili ausiliarie disponibili a livello d'area. Il nostro obiettivo è anche quello di identificare gli ostacoli che possano pregiudicare l'apprendimento da parte degli studenti. Inoltre, esploriamo l'inclusione di una componente che tiene conto della correlazione spaziale nei modelli di riferimento. I risultati della nostra analisi mostrano come le condizioni socioeconomiche e le dotazioni infrastrutturali abbiano un impatto significativo sui risultati conseguiti dagli studenti. Inoltre, mostriamo come l'applicazione dei modelli per piccole aree consenta di ridurre in maniera importante gli errori quadratici medi delle stime ottenute dall'INVALSI.

Keywords: Stima per piccole aree; INVALSI; Divario educativo; Modello di Fay-Herriot; Modello di Fay-Herriot spaziale.

Lo studio pubblicato è stato finanziato dall'Unione Europea – NextGenerationEU, nell'ambito del progetto GRINS – Growing Resilient Inclusive and Sustainable (GRINS PE00000018 – CUP H93C22000650001). I punti di vista e le opinioni espresse sono esclusivamente quelle degli autori e non riflettono necessariamente quelle dell'Unione Europea, né può l'Unione Europea essere ritenuta responsabile per esse.

Valutazione dei divari educativi tra le regioni italiane mediante la stima per piccole aree

Diego Battagliese[1]*, Alessio Pollice[1]

[1]*Dipartimento di Economia e Finanza, Università degli Studi di Bari Aldo Moro*

Riassunto: Una delle principali sfide con cui si confrontano i decisori pubblici con riferimento al servizio scolastico è quella di ridurre i divari fra studenti all'interno di una stessa scuola, ma ancor di più fra tutti gli studenti su base nazionale. I dati dell'INVALSI (Istituto Nazionale per la Valutazione del Sistema Educativo di Istruzione e Formazione) costituiscono la fonte informativa primaria per valutare tali divari. In questo lavoro, ci avvaliamo dei dati dell'indagine campionaria su scala regionale che vengono distribuiti liberamente dall'INVALSI. Il nostro obiettivo è quello di pervenire a stime più attendibili rispetto a quelle fornite dall'Istituto stesso. Per fare ciò, sfruttiamo i modelli *small area* di tipo *area-level*, ossia modelli lineari ad effetti misti, costruti su due livelli, in cui il dato INVALSI, che costituisce lo stimatore diretto, viene combinato con uno stimatore sintetico proveniente da modello, in modo da dare più vigore allo stimatore diretto per mezzo di variabili ausiliarie disponibili a livello d'area. Il nostro obiettivo è anche quello di identificare gli ostacoli che possano pregiudicare l'apprendimento da parte degli studenti. Inoltre, esploriamo l'inclusione di una componente che tiene conto della correlazione spaziale nei modelli di riferimento. I risultati della nostra analisi mostrano come le condizioni socioeconomiche e le dotazioni infrastrutturali abbiano un impatto significativo sui risultati conseguiti dagli studenti. Inoltre, mostriamo come l'applicazione dei modelli per piccole aree consenta di ridurre in maniera importante gli errori quadratici medi delle stime ottenute dall'INVALSI.

Keywords: Stima per piccole aree; INVALSI; Divario educativo; Modello di Fay-Herriot; Modello di Fay-Herriot spaziale.

* Autore corrispondente: diego.battagliese@uniba.it
Il lavoro qui descritto è stato svolto da D. Battagliese. La supervisione del lavoro è stata svolta da A. Pollice.

1. Introduzione

La stima per piccole aree (Small area estimation, SAE) facilita la produzione di indicatori disaggregati su scale territoriali in cui la dimensione del campione di riferimento non è tale da offrire stime affidabili. La stima per piccole aree si avvale di due classi di modelli: modelli *unit-level*, e modelli *area-level*. I primi sottintendono la disponibilità di dati individuali o microdati, mentre i secondi richiedono dati aggregati su un certo dominio territoriale. Esiste una pletora di letteratura scientifica sui modelli di stima per piccole aree. Sebbene numerose estensioni ai modelli preesistenti siano state proposte nel corso degli anni, l'attività di ricerca è a tutt'oggi intensa, soprattutto per quanto concerne il problema della *model selection*.

Le agenzie governative, usualmente, stimano delle quantità d'interesse su scala nazionale, quali possono essere, ad esempio, il tasso di disoccupazione, il reddito, l'indice di povertà, ecc. Ad ogni modo, la stima su scala nazionale non è indicativa delle differenze fra aree geografiche o sottopopolazioni, pertanto stime su scala ridotta sono necessarie affinché si possa intervenire nei territori che evidenziano situazioni di svantaggio. Le quantità che vengono prodotte mediante stimatori diretti *design-based* sono, di solito, meno affidabili, proprio alla luce della dimensione ridotta dell'area di riferimento. Tuttavia, esistono altri motivi che possono ostacolare la raccolta di un campione adeguato, che prescindono dalla dimensione del dominio geografico o demografico di riferimento, come, ad esempio, la reticenza ad esternare informazioni personali, soprattutto di matrice economica, ed in particolare in specifiche sottopopolazioni. SAE si avvale di un approccio di tipo *model-based* per produrre stimatori indiretti che vanno ad integrare l'informazione proveniente dallo stimatore diretto, in modo da supplire alla poca accuratezza che può caratterizzare quest'ultimo. Gli stimatori di *small area* nei modelli lineari sono, pertanto, tipicamente espressi come combinazione lineare convessa di stimatori diretti e stimatori sintetici desunti da un certo modello.

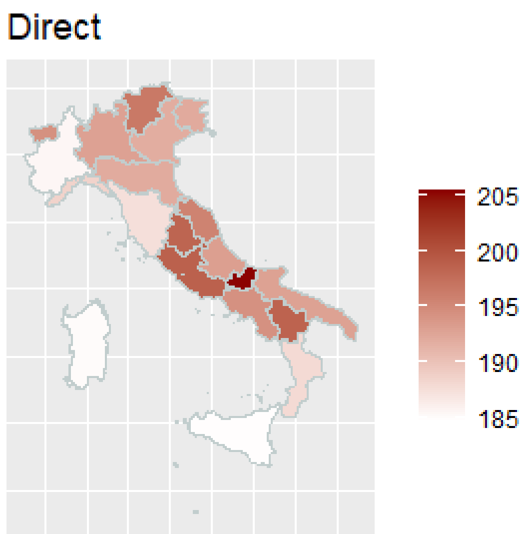
Il modello di Fay & Herriot (1979) è il modello più popolare per dati disponibili a livello di area. Nella fattispecie, esso sfrutta lo stimatore diretto, assumendo una struttura gerarchica per quest'ultimo, e utilizza l'informazione di variabili ausiliarie che può provenire da dati di registro o censuari. Numerose estensioni del modello di Fay & Herriot sono state proposte, per tenere conto di particolari specifiche dei dati. Ad esempio, alcune covariate potrebbero essere affette da errore di misurazione. Ybarra & Lohr (2008) hanno proposto una modifica del modello di Fay & Herriot in cui alcune covariate sono misurate con errore, usando un approccio frequentista. Arima et al. (2015) hanno rivisto in chiave Bayesiana gli stimatori di *small area* proposti da Ybarra & Lohr. Datta et al. (2011) hanno dimostrato che l'errore di misurazione può essere ridondante se considerato insieme ad effetti casuali riferiti alle singole aree; pertanto, si pone un problema di selezione (delle componenti rilevanti) del modello; si veda anche Datta & Mandal (2015). In molte situazioni, ad esempio, si ha a che fare con dati che deviano dalla normalità. Solitamente, per il reddito e per alcuni dati aziendali si osserva un'asimmetria positiva; ne consegue che la relazione lineare tra la variabile risposta e le covariate e le assunzioni di normalità dell'effetto casuale e dell'errore di campionamento, insite nel modello standard di Fay & Herriot, potrebbero essere violate. A tal proposito, sono state proposte soluzioni per ovviare a questo problema, che risiedono nell'applicazione di opportune trasformazioni degli stimatori di *small area* (Neves *et al.*, 2013; Schmid *et al.*, 2017). Il modello di Fay & Herriot standard assume indipendenza fra gli effetti casuali delle aree; d'altro canto, introdurre la correlazione spaziale fra aree geografiche può comportare un valore aggiunto in termini di stime (Petrucci e Salvati, 2006). Per una panoramica generale sui modelli di *small area* si veda Rao (2003).

Il presente lavoro ha ad oggetto la valutazione della qualità del sistema educativo e le determinanti di quest'ultima a livello regionale, in modo da rilevare eventuali gap fra regioni, e dunque permettere agli agenti politici di intervenire con azioni mirate nei territori che presentano criticità, al fine di uniformare il bagaglio culturale sull'intero piano nazionale. A tal proposito, ci avvaliamo dei risultati delle indagini INVALSI, che permettono di misurare il raggiungimento di alcune competenze di base nella popolazione studentesca. INVALSI produce stime su diversi gradi scolari, ovverosia sulle classi II e V della scuola primaria, sulla classe III della scuola secondaria di primo grado, e sulle classi II e V della scuola secondaria di secondo grado, rispettivamente corrispondenti ai gradi scolari 2°, 5°, 8°, 10°, 13°. Tale rilevazione viene fatta in diversi ambiti disciplinari, ossia matematica, italiano, e inglese. Nella fattispecie, ci avvaliamo dei soli risultati dei test di matematica, per due diversi gradi di istruzione, ovvero il grado 2°, rappresentato dalla classe seconda della scuola primaria, e il grado 10°, rappresentato dalla classe seconda della scuola secondaria superiore. INVALSI produce due diverse indagini. Dapprima, un'*Indagine Standard* (IS), effettuata sulla quasi totalità delle classi, nei diversi anni scolari. In questo caso le prove vengono somministrate dal corpo docente della scuola. In secondo luogo, viene fatta un'*Indagine Campionaria di Controllo* (ICC). In tal caso, viene estratto un campione casuale con metodo a due stadi: nel primo stadio sono campionate le scuole, e nel secondo, di norma, due classi per ogni scuola selezionata allo stadio precedente. Nelle classi così selezionate lo svolgimento delle prove, che rimangono le stesse, avviene alla presenza di un osservatore esterno che garantisce il pieno rispetto del protocollo di somministrazione. Lo scopo dell'ICC è quello di fare una correzione per il *cheating*, ossia il fenomeno per cui il corpo docente della scuola sarebbe più portato ad aiutare gli allievi nella risoluzione delle prove rispetto a quanto non facciano i valutatori esterni. Nonostante, l'ICC sia soggetta all'errore del disegno campionario, questa risulta meno affetta dall'errore di misurazione che deriva dal *cheating* stesso, che dunque porta a sovrastimare i rendimenti degli allievi.

2. Materiali e metodi

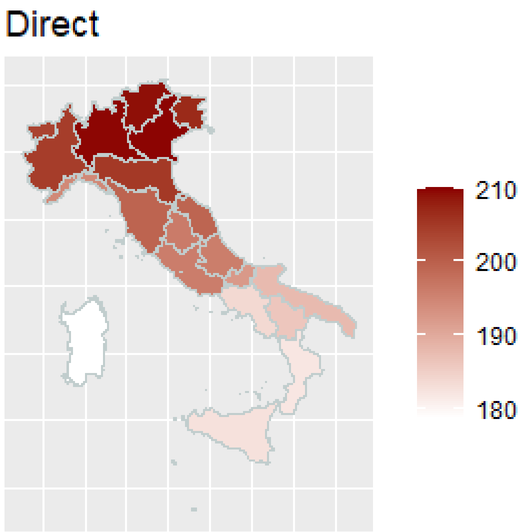
I dati utilizzati provengono dall'*Indagine Campionaria di Controllo* effettuata a livello regionale. Nel caso specifico, ci riferiamo ai dati dell'anno accademico 2021-2022, riferiti alla sola prova di matematica, e ai soli gradi scolari 2° e 10°. Nell'applicazione dei modelli *area-level*, la variabile risposta è costituita dal punteggio medio regionale del test di matematica. Essendo quest'ultimo una stima effettuata su un campione di scuole, esso va a costituire, di fatto, lo stimatore diretto (Horvitz e Thompson, 1952). Ad esso è associato un errore di campionamento, la cui conoscenza è importante ai fini dei modelli per piccole aree, per non incorrere in problemi di identificabilità. Fig. 1 mostra la distribuzione territoriale dello stimatore diretto per il grado scolare 2°.

Figura 1. *Distribuzione territoriale dello stimatore diretto del punteggio medio regionale del test di matematica per il grado scolare 2°.*



L’analisi di Fig. 1 non suggerisce grandi differenze territoriali rispetto ai risultati conseguiti dagli scolari della classe seconda della scuola primaria. Del resto, si evince una situazione di svantaggio nell’Italia insulare e nel Piemonte. Gli score più alti si registrano, invece, nell’Italia centro-meridionale, con valore massimo registrato nel Molise. Fig. 2 mostra lo stesso stimatore per il grado scolare 10°.

Figura 2. *Distribuzione territoriale dello stimatore diretto del punteggio medio regionale del test di matematica per il grado scolare 10°.*



Dall’analisi di Fig. 2 emerge chiaramente una polarizzazione dei risultati più alti nell’Italia centro-settentrionale. Questo dato necessita di un maggiore approfondimento di quelle che possano essere le cause di tale sconnessura tra nord e sud.

Lo score nei test di matematica per ogni grado di istruzione di interesse viene analizzato in risposta ad alcune covariate, specifiche per la scuola primaria e secondaria, che vengono selezionate *ad hoc* per i singoli gradi scolari. La scelta delle variabili ausiliarie viene fatta sulla base di un’accurata analisi esplorativa, al fine di far emergere le relazioni più forti. In particolare, per la classe seconda della scuola primaria (grado scolare 2°), vengono selezionate le seguenti variabili: numero di iscritti per classe, numero di insegnanti per scuola, stranieri per ogni cento iscritti. Le prime due possono essere viste come variabili di tipo infrastrutturale, mentre l’ultima ha una connotazione di natura sociologica. Tutte le variabili esplicative si riferiscono all’anno solare 2021. Per quanto concerne, invece, la classe seconda della scuola secondaria superiore (grado scolare 10°), le variabili di interesse sono: numero di stranieri per ogni cento iscritti, femmine per ogni cento iscritti, numero di ripetenti per classe. Stavolta, tutte le variabili possono considerarsi collegate a fenomeni sociali, e sono sempre riferite all'anno solare 2021. Si noti che nell'applicazione dei modelli di riferimento, per entrambi i gradi scolari, tutte le variabili sono state standardizzate. Altresì, tutte le suddette variabili provengono da rilevazioni fatte dall'Istituto Nazionale di Statistica (ISTAT), e pertanto è stato necessario allineare fonti di dati diverse. Tuttavia, qualche dato mancante è stato reperito tramite opportune documentazioni. In generale, ci riferiamo alla provincia autonoma di Trento in luogo del Trentino Alto-Adige, dal momento che INVALSI non rilascia stime relative all'intera regione.

Per mettere in relazione gli stimatori diretti, a livello di regione, con le variabili ausiliarie ci avvaliamo del modello di Fay & Herriot (1979). Esso consta di due livelli: un *sampling model*, ed un *linking* model, ovvero un modello lineare che mette in relazione le medie d'area con alcune covariate. In particolare, il primo livello può essere espresso come

$$\hat{\theta}_i^D = \theta_i + e_i, \quad i = 1, \dots, m$$

dove $\hat{\theta}_i^D$ è lo stimatore diretto non distorto delle medie areali di popolazione θ_i. Il termine e_i rappresenta gli errori di campionamento, che sono considerati indipendenti e con distribuzione assunta normale, $e_i \sim N(0, \varphi_i)$, mentre m è il numero di *small areas*, che nel nostro caso sono le regioni italiane, $m = 20$. Il parametro φ_i è solitamente noto, o al più viene stimato esternamente al modello, pertanto, nei modelli Bayesiani, esso non viene trattato come variabile aleatoria. Il secondo livello viene espresso come

$$\theta_i = x_i^T \beta + v_i$$

dove x_i è un set di covariate specifiche d'area, β è un vettore di coefficienti, e v_i rappresenta gli effetti casuali, che, nel modello standard di Fay & Herriot, sono indipendenti e identicamente distribuiti con distribuzione normale, $v_i \sim N(0, \sigma_v^2)$. Combinando i due livelli si ottiene uno speciale modello lineare ad effetti misti, in cui la struttura di covarianza è di tipo diagonale,

$$\hat{\theta}_i^D = x_i^T \beta + v_i + e_i.$$

Lo stimatore di Fay-Herriot, anche detto *empirical best linear unbiased predictor* (EBLUP) di θ_i, è ottenuto sostituendo alla varianza dell'effetto casuale, σ_v^2, e ai coefficienti di regressione, β, delle quantità stimate, rispettivamente $\hat{\sigma}_v^2$ e $\hat{\beta}$. Occorre, altresì, evidenziare che, nel paradigma Bayesiano, questi parametri vengono considerati aleatori, a meno che non siano noti. Ne segue che lo stimatore può essere scritto come

$$\hat{\theta}_i^{FH} = \hat{\gamma}_i\hat{\theta}_i^D + (1 - \hat{\gamma}_i)x_i^T\hat{\beta}$$

ove $\hat{\gamma}_i = \frac{\hat{\sigma}_v^2}{\hat{\sigma}_v^2+\varphi_i}$ viene detto fattore di *shrinkage*. Lo stimatore di Fay-Herriot è pertanto una media ponderata dello stimatore diretto e dello stimatore sintetico che deriva dalla regressione lineare. Quanto più la varianza del disegno campionario è piccola, e quanto più la varianza dell'effetto casuale è grande, tanto più si tende a privilegiare lo stimatore diretto; viceversa, nel caso opposto.

L'analisi dei dati si avvale del pacchetto *emdi* del software R, che incorpora sia modelli *unit-level* sia modelli *area-level* con relative estensioni. Tra quest'ultime, al fine di cogliere anche gli aspetti di carattere territoriale, abbiamo stimato i divari di apprendimento fra regioni utilizzando il modello di Fay-Herriot con connotazione spaziale. Quest'ultimo prende ad input la matrice di prossimità fra le regioni.

3. Risultati

Il modello standard di Fay-Herriot migliora nettamente le stime rispetto allo stimatore diretto in termini di errore quadratico medio, e questo avviene sia per quanto concerne il grado scolare 2° sia per il grado scolare 10°, evidenziando un concreto potere esplicativo delle variabili ausiliarie. Tuttavia, un ulteriore beneficio in termini di stime lo si ha considerando l'estensione spaziale del modello di Fay-Herriot stesso. Anche in questo caso osserviamo una significativa riduzione degli errori quadratici medi per entrambi i gradi scolari, sebbene questa sia più marcata per il grado scolare 10°. Una possibile spiegazione di questo comportamento potrebbe risiedere nel fatto che le variabili ausiliarie utilizzate per il grado scolare 2° riescano di per sé a catturare bene la correlazione spaziale esistente fra i risultati regionali. Invece, le variabili utilizzate per il grado 10°, seppure abbiano un elevato potere predittivo, sembrano spiegare meno la correlazione spaziale esistente. In aggiunta, i test di Moran e di Geary, che sono implementati nel pacchetto *emdi*, rivelano una non significativa correlazione spaziale per il grado 2°, mentre questa è fortemente significativa per il grado scolare 10°, come si evince anche da Fig. 2. Di seguito, riportiamo alcune misure sintetiche per entrambi i gradi scolari.

3.1 Grado scolare 2°

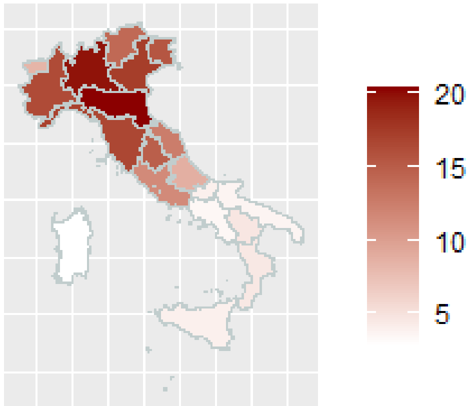
In questa sezione mostriamo i risultati dell'analisi, relativamente ai test di matematica per la classe seconda della scuola primaria. Nella fattispecie, mostriamo i risultati delle stime del modello lineare ad effetti misti, sia nella sua versione standard, sia nella sua estensione spaziale. In particolare, mettiamo in luce come l'adozione dei modelli a*rea-level* migliori le stime in termini di errore quadratico medio sia quando consideriamo il modello standard sia quando introduciamo una componente spaziale. Si tenga presente che il beneficio in termini di stime è sempre attribuibile alle variabili antecedenti, e dunque, a differenza di quanto vedremo per il grado scolare 10°, per il grado scolare 2° non riusciamo a dimezzare gli errori quadratici medi degli stimatori diretti. Ciò deriva anche dal fatto che, per il grado scolare 2°, gli stimatori diretti, di partenza, hanno errori quadratici medi più contenuti. Tab. 1 riporta l'output del modello standard di Fay-Herriot per il grado scolare 2°.

Tabella 1. *Stime da modello standard di Fay-Herriot degli effetti lineari delle variabili ausiliarie sui punteggi dei test di matematica per il grado scolare 2°.*

Variabili ausiliarie	Coefficienti	P-value
Intercetta	192,805	< 0,001
# stranieri per 100 iscritti	4,372	0,012
# iscritti per classe	-9,662	< 0,001
# insegnanti per scuola	6,610	0,003

Da Tab. 1 si evince come il numero di studenti stranieri, e il numero di insegnanti per scuola vadano ad influire positivamente sul risultato medio regionale delle prove INVALSI. Viceversa, si nota come classi più affollate abbiano un effetto negativo sugli score regionali. Anche se apparentemente potrebbe sembrare controintuitivo, il numero di studenti stranieri è correlato positivamente col risultato medio regionale. Il motivo di tale fenomeno è sussunto in Fig. 3. Difatti, il numero di studenti stranieri può essere considerato come una proxy della ricchezza delle regioni, dacché essi tendono a concentrarsi nelle regioni, per l'appunto, più ricche. Fig. 3 mostra come la grande maggioranza di bambini stranieri sia concentrata nelle regioni dell'Italia centro-settentrionale. Viceversa, al sud la loro presenza è senza dubbio scarsa, con una piccola rappresentanza nelle regioni di primo approdo, quali la Calabria e la Sicilia. Pertanto, possiamo concludere che la positività del coefficiente della variabile *studenti stranieri per 100 iscritti* debba essere attribuita ad un fenomeno di tipo economico che mette in relazione la ricchezza regionale con la presenza di stranieri.

Figura 3. *Distribuzione regionale del numero di studenti stranieri ogni 100 iscritti per il grado scolare 2°.*



Per verificare l'impatto dell'inserimento della componente spaziale sulle stime dei coefficienti, ci avvaliamo, in prima battuta, dei test di Moran e di Geary. L'indice *I* di Moran ha un campo di variazione compreso tra -1 e 1, ove valori uguali a -1 e 1 sanciscono perfetta autocorrelazione spaziale negativa e positiva, rispettivamente. L'indice *C* di Geary varia, invece, tra 0 e un valore positivo maggiore di 1, ove 1 rappresenta assenza di autocorrelazione spaziale, valori via via inferiori a 1 indicano una crescente autocorrelazione spaziale positiva, viceversa valori via via maggiori di 1 denotano una crescente autocorrelazione spaziale negativa. Nel caso in questione la statistica *I* di Moran risulta uguale a 0,064, con un p-value associato pari a 0,241, mentre la statistica *C* di Geary ha un valore uguale a 0,716, con un p-value pari a 0,045. Entrambi i test rivelano una correlazione spaziale non significativa. Tab. 2 riporta le stime derivanti dal modello di Fay-Herriot con componente spaziale.

Tabella 2. *Stime da modello spaziale di Fay-Herriot degli effetti lineari delle variabili ausiliarie sui punteggi dei test di matematica per il grado scolare 2°.*

Variabili ausiliarie	Coefficienti	P-value
Intercetta	190,190	< 0,001
# stranieri per 100 iscritti	4,589	0,028
# iscritti per classe	-8,349	< 0,001
# insegnanti per scuola	4,337	0,031

Dai risultati contenuti in Tab. 2 si nota come l'inclusione nel modello della correlazione spaziale non modifichi più di tanto l'entità e la natura dei coefficienti, sebbene ne risenta leggermente la significatività di qualcuno di essi. Le variazioni degli errori standard, che per brevità non mostriamo, unitamente a quelle dei coefficienti di regressione, danno luogo, in alcuni casi, a valori della statistica test meno estremi, e dunque a coefficienti meno significativi.

Il modello spaziale ha un impatto anche sugli errori quadratici medi dei singoli domini, come suggerisce Tab. 3.

Tabella 3. *Errori quadratici medi per lo stimatore diretto, lo stimatore di Fay-Herriot, e lo stimatore di Fay-Herriot nel modello spaziale, per il grado scolare 2°.*

Domini	MSE Diretto	MSE FH	MSE FH spaziale
Abruzzo	3,120	2,824	2,766
Basilicata	4,118	3,630	3,230
Calabria	4,018	3,589	3,441
Campania	4,214	3,755	3,584
Emilia-Romagna	2,238	2,131	2,038
Friuli-Ven. Giulia	2,709	2,501	2,481
Lazio	2,729	2,700	2,694
Liguria	2,861	2,631	2,556
Lombardia	2,354	2,233	2,156
Marche	2,878	2,650	2,532
Molise	3,092	2,919	2,857
Piemonte	3,129	2,835	2,618
Prov. Aut. Trento	2,865	2,658	2,618
Puglia	2,328	2,272	2,228
Sardegna	2,949	2,731	2,770
Sicilia	2,841	2,677	2,680
Toscana	3,249	3,006	2,876
Umbria	2,963	2,698	2,669
Valle d'Aosta	2,817	2,660	2,617
Veneto	2,733	2,551	2,429

Per brevità, riportiamo solo gli errori quadratici medi (Tab. 3), e non le stime delle medie d'area, che di fatto tendono a fare *shrinkage* attorno allo stimatore diretto. Tab. 3 pone in evidenza come le variabili antecedenti utilizzate per predire lo score regionale per il grado scolare 2° riescano a corroborare lo stimatore diretto. Difatti, lo stimatore di Fay-Herriot ha un errore quadratico medio inferiore a quello dei dati osservati per tutti i domini territoriali. Altresì, un ulteriore miglioramento deriva dall'inserimento dell'autocorrelazione spaziale. L'inserimento nel modello dell'effetto spaziale mediante la considerazione della matrice di prossimità comporta una diminuzione del MSE per tutte le regioni, eccezion fatta per le due isole, che di fatto non traggono beneficio dall'adiacenza di alcun'altra regione.

3.2 Grado scolare 10°

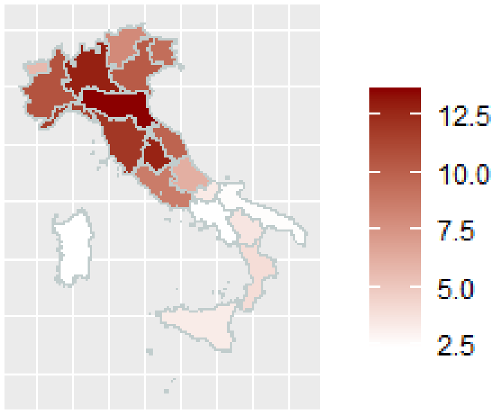
Come abbiamo già visto in precedenza per la classe seconda della scuola primaria, ora riportiamo alcuni risultati relativi ai test di matematica per la classe seconda della scuola secondaria superiore. Si voglia sottolineare che la peculiare spaccatura dell'Italia evintasi da Fig. 2 richiede una trattazione specifica. Pertanto, abbiamo considerato variabili ausiliarie differenti, al fine di far emergere le relazioni più importanti. In questo caso, le variabili considerate hanno un notevole potere predittivo, e si riferiscono maggiormente e fattori socioeconomici, piuttosto che a questioni di natura infrastrutturale. Come abbiamo già sottolineato, la scelta di buone covariate ha un impatto considerevole sulla precisione degli stimatori di Fay-Herriot. In Tab. 4 sono riassunti i risultati derivanti dall'applicazione del modello standard di Fay-Herriot per il grado scolare 10°.

Tabella 4. *Stime da modello standard di Fay-Herriot degli effetti lineari delle variabili ausiliarie sui punteggi dei test di matematica per il grado scolare 10°.*

Variabili ausiliarie	Coefficienti	P-value
Intercetta	196,276	<0,001
# stranieri per 100 iscritti	4,681	< 0,001
# femmine per 100 iscritti	4,616	< 0,001
# ripetenti per classe	-3,910	< 0,001

Tab. 4 rivela come il numero di studenti stranieri, e la quantità relativa di studentesse siano variabili positivamente correlate col risultato al test di matematica. In maniera non sorprendente, invece, si ha che classi con più ripetenti arrecano un effetto negativo ai risultati regionali. Tutte le variabili ausiliarie e l'intercetta mostrano coefficienti di regressione significativi ad un livello di significatività prossimo allo zero. Di nuovo, Fig. 4 riassume come il numero di studenti stranieri sia indicativo della ricchezza della regione; ciò spiega anche il motivo per cui la variabile *numero di studenti stranieri per 100 iscritti* presenti un coefficiente di regressione positivo, sebbene questo potrebbe sembrare, inizialmente, controintuitivo.

Figura 4. *Distribuzione regionale del numero di studenti stranieri ogni 100 iscritti per il grado scolare 10°.*



Ancora, volendo verificare la presenza di correlazione spaziale, ci avvaliamo dei suddetti test d'ipotesi. In questo caso, la statistica I di Moran presenta un valore uguale a 0,647, con un p-value associato di gran lunga inferiore a 0,001, mentre, la statistica C di Geary risulta pari a 0,095, con un p-value anch'esso molto vicino allo zero. Da entrambi i test si evince una correlazione spaziale positiva e fortemente significativa. Tuttavia, la linea di demarcazione che divide il nord dal sud (Fig. 2) fa sì che questa correlazione spaziale non riesca bene ad esplicarsi a livello regionale; e forse sarebbe meglio colta ad un livello di disaggregazione territoriale maggiore. Tab. 5 riporta le stime derivanti dal modello di Fay-Herriot con componente spaziale.

Tabella 5. *Stime da modello spaziale di Fay-Herriot degli effetti lineari delle variabili ausiliarie sui punteggi dei test di matematica per il grado scolare 10°.*

Variabili ausiliarie	Coefficienti	P-value
Intercetta	193,145	<0,001
# stranieri per 100 iscritti	4,085	< 0,001
# femmine per 100 iscritti	3,583	< 0,001
# ripetenti per classe	-3,098	< 0,001

In Tab. 5 non si notano grosse differenze rispetto a Tab. 4 in termini di magnitudo dei coefficienti. Vale la pena rilevare, però, che, sebbene non mostrati, gli errori standard delle stime dei coefficienti sono sempre superiori a quelli derivanti dal modello standard di Fay-Herriot. Tuttavia, tutti i coefficienti rimangono significativi ad un livello di significatività α dell'1 per mille. L'analisi congiunta di Fig. 2 e dei risultati dei test di Moran e Geary suggerisce come l'aggiunta nel modello dell'autocorrelazione spaziale riduca sensibilmente gli errori quadratici medi, salvo in alcune regioni di frontiera, quali la Valle d'Aosta e il Friuli-Venezia Giulia, e, ancora, le due isole, che non beneficiano della prossimità di altre regioni. Gli errori quadratici medi risultanti dall'applicazione delle versioni standard e spaziale del modello di Fay-Herriot, nonché quelli dello stimatore diretto, sono riassunti in Tab. 6. D'altronde, occorre sottolineare come, di partenza, l'errore del disegno campionario condotto da INVALSI, per il grado scolare 10°, risulti ben più elevato rispetto a quello del grado scolare 2°. Probabilmente questo è ascrivibile al fatto che le scuole secondarie superiori sono suddivise in diverse tipologie, cosa che non avviene per le scuole primarie. Si può, altresì, notare come le regioni più piccole, quali, ad esempio, la Valle d'Aosta e il Molise, presentino un errore quadratico medio, associato allo stimatore diretto, piuttosto elevato. Questo avvalora ulteriormente l'adozione dei modelli in questione.

Tabella 6. *Errori quadratici medi per lo stimatore diretto, lo stimatore di Fay-Herriot, e lo stimatore di Fay-Herriot nel modello spaziale, per il grado scolare 10°.*

Domini	MSE Diretto	MSE FH	MSE FH spaziale
Abruzzo	5,291	3,161	2,795
Basilicata	4,294	3,364	2,703
Calabria	4,090	3,260	3,248
Campania	3,935	3,090	2,610
Emilia-Romagna	4,938	3,325	2,347
Friuli-Ven. Giulia	5,279	3,053	3,345
Lazio	4,558	2,812	2,240
Liguria	4,850	3,232	2,766
Lombardia	4,351	3,137	2,436
Marche	5,401	2,961	2,372
Molise	6,413	4,854	4,050
Piemonte	4,489	2,920	2,327
Prov. Aut. Trento	5,875	4,800	4,323
Puglia	4,351	3,072	2,511
Sardegna	4,024	3,035	3,210
Sicilia	3,616	2,986	3,107
Toscana	4,486	3,447	2,654
Umbria	5,653	3,623	3,287
Valle d'Aosta	7,711	5,123	5,849
Veneto	4,954	3,066	2,202

Rispetto al caso del grado scolare 2°, in cui l'adozione del modello di Fay-Herriot, sia nella sua versione standard sia in quella spaziale, riduceva sì gli errori quadratici medi rispetto a quelli dello stimatore diretto, ma non di molto, in Tab. 6, abbiamo che gli errori quadratici medi dello stimatore diretto vengono in alcuni casi addirittura dimezzati. Questo è attribuibile al forte potere esplicativo insito nelle variabili ausiliarie utilizzate, sebbene le varianze degli stimatori diretti fossero originariamente abbastanza elevate. Altresì, rileviamo come il modello spaziale va ulteriormente a ridurre la varianza delle stime, eccezion fatta per la Valle d'Aosta, il Friuli-Venezia Giulia, la Sardegna e la Sicilia, traendo, forse, beneficio da variabili ausiliarie che da sole non colgono in maniera decisiva quella che è l'autocorrelazione spaziale esistente.

4. Conclusioni

In questo lavoro, abbiamo adottato i modelli per piccole aree, nella fattispecie i modelli di tipo *area-level*, per migliorare le stime provenienti dall'indagine campionaria condotta da INVALSI. In particolare, gli stimatori diretti traggono forza sia dalle variabili ausiliarie sia dalla componente di natura spaziale. Gli stimatori diretti per il grado scolare 2° hanno una distribuzione abbastanza omogenea, o tutt'al più non così diversa, sul territorio nazionale; d'altronde per il grado scolare 10°, si osserva una spaccatura tra nord e sud dell'Italia. A tal proposito, cerchiamo di capire quali siano le determinanti di tale divario nel profitto scolastico. Queste possono essere attribuite sia a fattori socioeconomici sia a deficit di carattere infrastrutturale, come, ad esempio, delle classi troppo numerose, o un sottodimensionamento dell'organico scolastico. Ancora, gli stimatori diretti prodotti da INVALSI vengono distorti dal fenomeno del *cheating*. Benché INVALSI produca questi dati campionari sotto il controllo di osservatori esterni al corpo docente, ciò non preclude la possibilità che i risultati dei test siano frutto di contaminazioni fra studenti. Pertanto, il nostro obiettivo è proprio quello di tenere conto di queste possibili contaminazioni, e dunque fornire delle stime più attendibili.

L'adozione del modello standard di Fay-Herriot riduce la varianza delle stime per entrambi i gradi scolari considerati, e lo stesso avviene, in generale, quando consideriamo l'estensione spaziale del modello di Fay-Herriot. In questo senso, le variabili ausiliarie giocano un ruolo chiave. Abbiamo considerato variabili ausiliarie diverse per i due gradi scolari. Questo al fine di cogliere le relazioni più forti fra la variabile risposta e le covariate. Quest'ultime sembrano cogliere bene anche le possibili correlazioni fra i risultati regionali; difatti, per il grado scolare 2°, il modello con la componente spaziale riduce di poco gli errori quadratici medi, mettendo dunque in discussione l'ulteriore parametrizzazione derivante dalla componente spaziale, in virtù del principio del Rasoio di Occam. Ciò potrebbe essere ascritto a due ragioni di fondo. La prima, per l'appunto, costituita dall'assenza di una correlazione spaziale significativa, mentre, la seconda insita nella capacità delle variabili ausiliarie utilizzate per il grado scolare 2° di catturare la correlazione spaziale.

Eventuali estensioni di questo lavoro possono tenere conto dell'errore di misurazione nelle variabili ausiliarie, nonché di una ricerca più accurata di quest'ultime, prendendo in considerazione, ad esempio, i tempi di percorrenza casa-scuola, o altre variabili sensibili rispetto alle performance scolastiche.

Riconoscimenti

Lo studio pubblicato è stato finanziato dall'Unione Europea – *NextGenerationEU*, nell'ambito del progetto GRINS – *Growing Resilient Inclusive and Sustainable* (GRINS PE00000018 – CUP H93C22000650001). I punti di vista e le opinioni espresse sono esclusivamente quelle degli autori e non riflettono necessariamente quelle dell'Unione Europea, né può l'Unione Europea essere ritenuta responsabile per esse.

Riferimenti bibliografici

Arima, S.; Datta, G. S.; Liseo, B. (2015). Bayesian estimators for small area models when auxiliary information is measured with error. *Scadinavian Journal of Statistics*, 42: 518-529.

Datta, G. S.; Hall, P.; Mandal, A. (2011). Model selection by testing for the presence of small-area effects, and application to area-level data. *Journal of the American Statistical Association*, 106: 362-374.

Datta, G. S.; Mandal, A. (2015). Small area estimation with uncertain random effects. *Journal of the American Statistical Association*, 110: 1735-1744.

Fay, R. E.; Herriot, R. A. (1979). Estimates of income for small places: An application of James-Stein procedure to census data. *Journal of the American Statistical Association*, 74: 269-277.

Horvitz, D. G.; Thompson, D. J. (1952). A generalization of sampling without replacement from a finite universe. *Journal of the American Statistical Association*, 47: 663-685.

Neves, A.; Silva, D.; Correa, S. (2013). Small domain estimation for the Brazilian service sector survey. *Estadística*, 65: 13-37.

Petrucci, A.; Salvati, N. (2006). Small area estimation for spatial correlation in watershed erosion assessment. *Journal of Agricultural, Biological and Environmental Statistics*, 11: 169-182.

Rao, J. N. K. (2003). *Small area estimation*. Wiley, Hoboken, New Jersey.

Schmid, T.; Bruckschen, F.; Salvati, N.; Zbiranski, T. (2017). Constructing sociodemographic indicators for national statistical institutes using mobile phone data: Estimating literacy rates in Senegal. *Journal of the Royal Statistical Society Series A: Statistics in Society*, 180: 1163-1190.

Ybarra, L. M. R.; Lohr, S. L. (2008). Small area estimation when auxiliary information is measured with error. *Biometrika*, 95: 919-931.