**Esclusione sociale e disparità territoriali: mappatura della povertà e della disuguaglianza mediante nuovi metodi di stima per piccole aree**

**PROGETTO DI RICERCA**

**Obiettivo del progetto di ricerca.** Obiettivo del progetto è la valutazione sul territorio italiano delle disparità concernenti povertà (at-risk of poverty rate) e disuguaglianza (indici di Gini, Theil e Atkinson). L'UE ha da decenni l’obiettivo di promuovere la coesione territoriale. Ad acuire la disparità tra ed entro aree geografiche concorre l’intensificarsi del divario centro-periferia tra aree metropolitane ed interne: le prime in grado di attrarre sempre più cittadini e caratterizzate da un’importante offerta di servizi essenziali; le seconde contraddistinte da una intrinseca carenza di servizi e da un progressivo abbandono. Per questa ragione, anche a fronte della pandemia di COVID-19, i produttori di statistiche ufficiali stanno affrontando una crescente richiesta di stime relative ad indicatori di esclusione sociale disaggregate territorialmente. Le indagini ufficiali da cui derivano tali stime sono pianificate per fornire stime affidabili a livello di macroarea (nella migliore delle ipotesi, a livello NUTS-2, le nostre Regioni), e non consentono di raggiungere un livello di dettaglio territoriale più fine.

Ad esempio, l'Indagine sul reddito e sulle condizioni di vita condotta da Eurostat (EU-SILC), che viene correntemente utilizzata per stimare parametri di povertà e disuguaglianza, è in grado di fornire stime affidabili dei parametri in oggetto solo a livello NUTS-2 o per le macroregioni. La disponibilità di stime a livello NUTS-3 (le province), con diversificazione tra aree metropolitane, urbane e rurali, permetterebbe, al contrario, di effettuare una analisi più articolata della disparità territoriale e di valutarne la componente entro le regioni, oltre a quella tra le stesse. In statistica il problema della insufficiente numerosità campionaria in sottopopolazioni viene affrontato con metodi cosiddetti di "small area estimation” (SAE). In breve, gli usuali stimatori basati sull'informazione campionaria forniscono generalmente stime non affidabili in piccole aree mentre gli stimatori “da modello” permettono l'ottenimento di stimatori indiretti significativamente più efficienti. Per una rassegna si vedano Pfeffermann (2013), Rao e Molina (2015) e Tzavidis et al. (2018).

Il principale problema metodologico che si incontra nell'implementazione di modelli SAE di tipo “area level” per la stima indicatori di povertà e disuguaglianza è la loro scarsa flessibilità a fronte di stimatori con peculiari caratteristiche distributive, quali marcata asimmetria e leptocurtosi. Più nel dettaglio l'assunzione di normalità distributiva degli stimatori non può essere giustificata con richiamo al teorema del limite centrale a causa della scarsa numerosità campionaria. In aggiunta, tale ipotesi non risulta adeguata nel caso di indicatori definiti sull'intervallo unitario (indici di disuguaglianza e tassi di povertà).

Obiettivo principale di questa proposta è, dunque, quello di incorporare assunzioni distributive alternative alla comune assunzione di normalità, frequentemente utilizzata nel contesto SAE, che siano maggiormente flessibili e in grado di considerare le peculiarità distributive degli stimatori in oggetto. Un'innovativa proposta metodologica verrà sviluppata tramite un approccio basato su modelli mistura di distribuzioni Beta. Ci si concentrerà, inoltre, sulla specificazione di funzioni di varianza approssimate per particolari stimatori non lineari da incorporare nel modello stesso.

La seconda parte della ricerca riguarderà la costruzione di mappe di esclusione sociale e l’utilizzo delle stime prodotte per la valutazione delle disparità territoriali. In questo ambito si utilizzeranno gli strumenti della statistica spaziale e specifiche metodologie di clustering territoriale (Singleton et al., 2020).

**Carattere di originalità e innovatività della proposta progettuale.** Nonostante alcune eccezioni, la maggior parte degli studi regionali sull'esclusione sociale tende a focalizzarsi su dati e pattern geografici tra nazioni piuttosto che sull'esplorazione di potenziali pattern tra regioni e di connessioni interregionali. La valutazione della coesione territoriale basata su povertà, deprivazione e diseguaglianza richiede la produzione di stime ad un dettagliato livello territoriale. Essendo disponibili stime a livello nazionale (NUTS-1) e regionale (NUTS-2), l'analisi di tali fenomeni viene generalmente portata avanti a questi livelli. Un contributo che permetta il rilascio di stime a livello di province (NUTS-3) o più fine oppure su specifiche sotto-popolazioni (stratificazione per genere, età od etnia) può portare un'informazione dettagliata da cui trarre un quadro più articolato. Tali stime possono rivelarsi preziose nell'approfondire ulteriormente i trend delle disparità regionali, nonché nell'identificazione delle aree che possono costituirsi come driver del fenomeno a livello nazionale, nella classificazione delle aree per la diversificazione place-based delle politiche, e nello studio degli spillover spaziali (Cavanaugh and Breau, 2018).

Dal punto di vista metodologico, la letteratura dei modelli per piccole aree è composta, per la maggior parte, da modelli con assunzioni di normalità a livello di distribuzione campionaria. Tali assunzioni, come detto, non risultano adeguate per indicatori definiti sull'intervallo unitario e caratterizzati da distribuzioni asimmetriche a code alte. La letteratura, nel caso di indicatori definiti tra 0 ed 1, si divide in due filoni: il primo sfrutta specifiche trasformazioni sulle stime dirette che permettono di pervenire ad un supporto non limitato, su cui è possibile imporre l'assunzione di gaussianità; il secondo si basa sulla regressione Beta. Nessuno dei due, tuttavia, affronta il problema della marcata asimmetria e leptocurtosi. Per una rassegna si vedano Janicki (2020), Esteban et al. (2020) e Rojas-Perilla et al. (2020). La letteratura che affronta l'utilizzo di adeguate distribuzioni flessibili è esigua (Diallo and Rao, 2018; Graf et al., 2019). Con il fine di potenziare l'affidabilità degli stimatori, c'è dunque spazio per innovativi modelli per piccole aree basati su assunzioni distributive particolarmente flessibili.

La proposta sarà sviluppata mediante un approccio di tipo Hierarchical Bayes, che ha diversi vantaggi nel contesto SAE in quanto consente di gestire facilmente ipotesi distributive non gaussiane e di catturare l'incertezza su tutti i parametri target attraverso la distribuzione a posteriori. I modelli utilizzati sono di tipo “area level”, che non richiedono covariate individuali ma sono a livello di area e consentono naturalmente di incorporare le probabilità di inclusione nel processo di stima.

**Impatto della ricerca.** L'impatto previsto si sviluppa sia a livello di definizione di policy place-based, sia a livello scientifico/conoscitivo.

Il primo livello riguarda l'impatto in termini di definizione e di valutazione ex-post delle politiche regionali, con particolare attenzione al livello locale. Il raggiungimento della coesione territoriale, obiettivo chiave delle politiche UE, è un tema da tempo dibattuto da ricercatori e policymaker, che ha trovato nuova linfa a partire dalla crisi finanziaria del 2007-2008, dalle crisi determinate dalla pandemia di COVID-19 e dal conflitto russo-ucraino, eventi in grado di influenzare pesantemente le condizioni socioeconomiche in Europa. I fondi EU di coesione e di sviluppo regionale mirano a stimolare la convergenza geografica di sviluppo degli Stati membri. Nonostante ciò, le differenze entro gli Stati risultano spesso più rilevanti di quelle tra Stati. Data l'ampiezza e l'eterogeneità economica delle regioni europee, la disponibilità di stime ad un dettaglio territoriale fine di indicatori di esclusione sociale consentirebbero una adeguata identificazione del fenomeno di disparità entro le macroregioni, il disegno e la valutazione di politiche place-based ed il riconoscimento di specifiche determinanti locali.

Un secondo impatto può definirsi di tipo scientifico/conoscitivo. Come sottolineano Ramos and Royuela (2014), la diseguaglianza intra-regionale negli utlimi decenni è aumentata, con differenze tra regioni maggiori di quelle tra stati. Con riferimento all'Italia, una disparità in termini di PIL è stata rilevata su Roma e Milano a svantaggio della restante parte, dove spesso i dettagli entro le aree urbane di piccole dimensioni campionarie ma con elevata popolazione sono celate da un'analisi portata avanti a livello NUTS-2 (Ballas et al., 2017). La coesione socioeconomica territoriale viene generalmente studiata sulla base del PIL pro capite, aggregato disponibile a dettagliati livelli spaziali. Tuttavia, le policy con obiettivi di coesione, espresse anche dal PNRR, non si focalizzano solo sulla crescita del PIL, soggetto a numerose critiche rispetto al suo utilizzo come base di comprensione della coesione territoriale: il PIL non dà contezza di parametri cruciali quali la qualità delle condizioni di vita, l'inclusione sociale, lo sviluppo sostenibile della crescita economica. L'utilizzo, dunque, in questa sede, di misure che tengano conto della distribuzione interpersonale (individuale) del reddito e della ricchezza, permette di considerare la dimensione assistenziale (dei trasferimenti),famigliare e patrimoniale, utili alla valutazione delle condizioni di vita e di inclusione sociale di una popolazione.

Gli obiettivi di impatto misurabili riguardano il rilascio di codici e/o librerie R che permettano un’agevole implementazione delle metodologie proposte, la simulazione di una banca dati georeferenziata che includa le stime dei diversi indicatori e la pubblicazione di un contributo conferibile in VQR.

**PIANO DI ATTIVITA’**

**Implementazione della ricerca.** I dati utilizzati dal progetto derivano dall'indagine EUROSTAT sul Reddito e le Condizioni di Vita (EU-SILC). Le informazioni necessarie per lo sviluppo del progetto saranno a breve in possesso di ricercatori afferenti al gruppo di ricerca del PI, in quanto intestatari di un protocollo di intesa per la ricerca suggellato con ISTAT ed in fase avanzata di approvazione, e con ottenuto consenso per l'ammissione al Laboratorio ADELE di ISTAT. Il Laboratorio per l’Analisi dei Dati ELEmentari (ADELE) è l’ambiente dell’Istat attraverso il quale viene offerto l’accesso da remoto ai dati elementari d’indagine, da utilizzare a scopi di ricerca, cui non sono stati applicati metodi di controllo per la tutela della riservatezza.

Il programma delle attività associate al progetto avrà durata temporale di un anno e sarà suddiviso in 4 fasi:

*Fase 1*

I primi 2 mesi saranno dedicati alla rassegna ed alla discussione della recente letteratura sui temi per aggiornare il framework concettuale del progetto. Questa attività sarà svolta a partire dall’analisi dei lavori riportati nella successiva sezione di Riferimenti Bibliografici. In particolare, si effettuerà una rassegna della letteratura dei modelli per piccole aree, sviluppati sia a livello di area (area-level) che a livello individuale (unit-level). La rassegna sarà svolta considerando diverse direzioni. Il focus metodologico specifico verterà sui modelli per dati non-Gaussiani e per indicatori definiti sull'intervallo unitario, mentre un secondo focus verterà sulla costruzione di una rassegna di lavori relativi ad indicatori di esclusione sociale che prendano in considerazione sia il reddito sia la ricchezza. Un secondo filone di attività riguarderà l’analisi della letteratura relativa alla disparità territoriale tra ed entro le regioni, riferita ai fenomeni oggetto di studio, alla individuazione di partizioni territoriali che consentano una adeguata comprensione della diffusione spaziale dei fenomeni considerati, a metodi di utilizzo a fini conoscitivi delle stime che si otterranno con le tecniche di stima per piccole aree.

*Fase 2*

Nei successivi 2 mesi ci si dedicherà all’ottenimento delle stime basate sul disegno dei diversi indicatori sopra elencati, al livello territoriale NUTS 3,ed alla stima dei rispettivi coefficienti di variazione, mediante tecniche di tipo bootstrap o mediante linearizzazione. I risultati di questa operazione consentiranno di valutare se sarà verosimilmente possibile ottenere stime da modello affidabili a tale livello di dettaglio territoriale ose, almeno per alcuni parametri, sarà possibile arrivare ad un livello di dettaglio territoriale più fine. Si passerà, quindi, alla individuazione del vettore di covariate più adeguato per ogni indicatore ed al reperimento dei relativi dati al livello territoriale designato (verosimilmente dati provenienti da dichiarazioni fiscali, disponibili con aggregazione comunale, dati demografici, ecc).

*Fase 3*

La terza parte del progetto, della durata di 3/4 mesi, sarà dedicata allo sviluppo della proposta metodologica e alla definizione di varie fasi di test da attuare sulla stessa, sia in ambito di applicazione che di simulazione. Il primo filone di metodi verterà sulla proposta di modelli di stima per piccole aree per stimatori con distribuzione asimmetrica a code alte, anche definiti sull'intervallo unitario. Un secondo filone di metodi, al contrario, si concentrerà sul trattamento delle stime da modello a fini di studio fenomenico, attraverso l'implementazione di algoritmi di classificazione e/o modelli di dipendenza spaziale.

*Fase 4*

Infine, la parte conclusiva sarà dedicata alla finalizzazione dell'analisi, alla stesura di 1 manoscritto, alla condivisione e documentazione dei metodi sviluppati attraverso codici e/o librerie R ed alla mappatura dei risultati attraverso la creazione di una banca dati georeferenziata.

Per ogni fase del progetto saranno previsti incontri settimanali tra assegnista, PI e CO-PI con lo scopo di verificarne lo stato di avanzamento e di garantire un alto livello di collaborazione tra l'assegnista ed i responsabili. Le attività ed i risultati saranno sistematicamente condivisi in modo da assicurare la qualità dell'attività di ricerca svolta e della sua finalizzazione in termini di pubblicazioni.

**Riferimenti bibliografici**

Ballas, D., Dorling, D., & Hennig, B. (2017). Analysing the regional geography of poverty, austerity and inequality in Europe: A human cartographic perspective. Reg Studies, 51(1), 174-185.

Cavanaugh, A., & Breau, S. (2018). Locating geographies of inequality: publication trends across OECD countries. Reg Studies, 52(9), 1225-1236.

Diallo, M.S., & Rao, J.N.K. (2018). Small area estimation of complex parameters under unit‐level models with skew‐normal errors. Scand J of Statistics, 45(4), 1092-1116.

Esteban, M.D., Lombardía, M.J., López-Vizcaíno, E., Morales, D., & Pérez, A. (2020). Small area estimation of proportions under area-level compositional mixed models. TEST, 29(3), 793-818.

Graf, M., Marín, J.M., & Molina, I. (2019). A generalized mixed model for skewed distributions applied to small area estimation. TEST, 28(2),565-597.

Janicki, R. (2020). Properties of the beta regression model for small area estimation of proportions and application to estimation of poverty rates. Communications in Statistics-Th & Meth, 49(9), 2264-2284.

Pfeffermann, D. (2013). New important developments in small area estimation. Stat Science, 28(1), 40-68.

Rao, J.N., & Molina, I. (2015). Small Area Estimation. John Wiley & Sons.

Rojas‐Perilla, N., Pannier, S., Schmid, T., & Tzavidis, N. (2020). Data‐driven transformations in small area estimation. Journal of the Royal Statistical Society: Series A (Stat in Soc), 183(1), 121-148.

Singleton, A., Alexiou, A., & Savani, R.(2020). Mapping the geodemographics of digital inequality in Great Britain: An integration of machine learning into small area estimation. Computers, Environment and Urban Systems, 82, 101486.

Tzavidis, N., Zhang, L. C., Luna, A., Schmid, T., & Rojas‐Perilla, N. (2018). From start to finish: a framework for the production of small area official statistics. Journal of the Royal Statistical Society: Series A (Stat in Soc), 181(4), 927-979