

Documenti di discussione dell'Ufficio Studi

2007/2

**Stima dei consumi interni nei comuni di
Lombardia, Toscana e Sicilia**

*Stefano Casini Benvenuti
Valentina Tortolini
Alessandro Viviani*

Documenti di discussione dell'Ufficio Studi

2007/2

Stima dei consumi interni nei comuni di Lombardia, Toscana e Sicilia

Stefano Casini Benvenuti()*

*Valentina Tortolini(**)*

*Alessandro Viviani(**)*

() Istituto per la Programmazione Economica della Toscana (IRPET)*

*(**) Università di Firenze*

I documenti di discussione non riflettono necessariamente l'opinione ufficiale dell'Agenzia delle Entrate ed impegnano unicamente gli autori.

Possono essere liberamente utilizzati e riprodotti per finalità di uso personale, studio, ricerca o comunque non commerciali a condizione che sia citata la fonte secondo la seguente dicitura, impressa in caratteri ben visibili:

<http://www.agenziaentrate.it/ufficiostudi>.

Informazioni e chiarimenti: ae.ufficiostudi@agenziaentrate.it

Stima dei consumi interni nei comuni di Lombardia, Toscana e Sicilia

Lo studio propone una metodologia di stima dei consumi interni a livello sub-provinciale. L'oggetto della stima rappresenta un indicatore economico molto potente per quantificare le potenzialità economiche di piccole aree. Oltre ai risultati ottenuti il lavoro fornisce spunti metodologici per colmare le lacune nella disponibilità di informazioni per micro aree geografiche.

Indice

PREFAZIONE DELL'UFFICIO STUDI.....	5
SINTESI DIREZIONALE.....	6
1. PREMESSA	8
2. LA LETTERATURA SULLE STIME PER PICCOLE AREE.....	12
2.3 I METODI BASATI SU MODELLO	14
2.4 IL MODELLO A LIVELLO DI AREA	15
2.5 I PREDITTORI BLUP ED EBLUP.....	17
3 IL MODELLO DI STIMA ADOTTATO.....	18
3.1 UNA STIMA A DUE STADI.....	18
3.2 LA METODOLOGIA ADOTTATA	21
4. LE STIME OTTENUTE.....	22
4.1 I CONSUMI PROVINCIALI.....	22
4.2 LE STIME A LIVELLO LOCALE.....	25
4.3 UN COMMENTO DEI RISULTATI	27
APPENDICE – CONFRONTO TRA I DUE MODELLI.....	30

Prefazione dell'Ufficio Studi

Il lavoro che pubblichiamo è importante per diversi motivi. Innanzitutto si tratta di una proficua collaborazione tra enti centrali (l'Agenzia delle Entrate) e periferici (l'Istituto di ricerca della regione Toscana, IRPET) della pubblica amministrazione, dove i primi hanno messo a disposizione l'informazione, nazionale e regionale, di fonte amministrativa (le dichiarazioni Irap nel caso di specie) ed i secondi hanno integrato tali fonti con la conoscenza diretta del territorio. Il tutto finalizzato ad approfondire aspetti di analisi economica rilevanti per la programmazione territoriale.

In secondo luogo si sono messe a fattor comune conoscenze tecnologiche che hanno permesso all'Irpet di delineare, nelle pagine che seguono, una tecnica per disaggregare i dati a livello sub provinciale. Il percorso è iniziato con l'elaborazione di un prototipo, effettuato nel 2004, riferito alla sola regione Toscana, esteso, in questa sede anche ad altre due regioni: Lombardia e Sicilia.

Infine, il contributo dell'Irpet colma un vuoto conoscitivo rilevante. Il risultato quantitativo raggiunto, infatti, consente di poter disporre di informazioni sui consumi finali della *popolazione presente* su base sub-regionale.

Per le difficoltà connesse alla definizione dell'aggregato, chiaramente esposte nelle pagine seguenti, queste stime non sono fornite dagli enti istituzionalmente preposti alla diffusione dell'informazione statistica. Tale cautela è assolutamente condivisibile, poiché il conseguimento dei risultati è subordinato alla formulazione di ipotesi "forti" alla base della procedura statistica. Consapevoli di questo, i risultati acquisiti devono essere interpretati con cautela, ma, nonostante ciò, consentono interessanti spunti di riflessione sulla realtà economica di piccole aree.

L'interesse dell'amministrazione finanziaria, ad esempio, è quello di conoscere la potenzialità economica di aree relativamente piccole e comunque su base sub-provinciale. E' questo un passaggio cruciale per costruire successivamente indicatori di *compliance* incentrati sugli uffici territoriali e per meglio analizzare le caratteristiche dei bacini di utenza.

I risultati ottenuti non sono, quindi, da considerarsi come un punto di arrivo della ricerca, ma sono offerti all'amministrazione ed alla comunità scientifica per una valutazione e per ricevere ulteriori spunti di riflessione.

Sintesi direzionale¹

Il presente lavoro rappresenta un tentativo di procedere alla stima di una grandezza –i consumi interni a *livello subregionale*- non conosciuta, ma importante per motivi diversi. Innanzitutto perché i consumi interni rappresentano una componente rilevante della domanda finale che si rivolge al sistema produttivo, in particolare per tutte quelle attività cosiddette “locali”, ovvero attività per cui la vicinanza tra produttore e utente è indispensabile: le scarpe che i consumatori acquistano possono provenire anche da molto lontano, ma il commerciante che le vende deve stare a stretto contatto con l’acquirente.

È evidente che alcune attività produttive –quelle di servizio soprattutto- dipendono largamente dal comportamento dei consumi interni e, quindi, il volume d’affari che esse realizzano dipende in larga misura dagli acquisti che i consumatori presenti nell’area fanno. Quindi, se su alcune di queste attività, disponessimo di informazioni corrette e rappresentative del volume d’affari realizzato (o una sua proxy) queste ultime rappresenterebbero una buona base per la stima dei consumi interni.

Naturalmente la premessa fondamentale è che il volume d’affari sia non solo conosciuto ma anche esente dai vizi che potrebbero compromettere la sua capacità di rappresentare l’effettivo ammontare degli acquisti fatto da consumatori presenti.

Sul primo punto –la disponibilità di informazioni sul volume d’affari- i dati a disposizione dell’Agenzia delle Entrate relativi alle dichiarazioni IRAP sono senza dubbio la fonte primaria quando si scende ad un livello territoriale inferiore alla regione (ricordiamo che l’ISTAT fornisce stime dei consumi interni solo a livello regionale); sul secondo punto –la correttezza e rappresentatività del dato- i vizi possono essere molti, ma senza dubbio il principale è legato all’evasione fiscale. Di qui la necessità di utilizzare dati relativi alle dichiarazioni IRAP di quelle attività in cui il fenomeno dell’evasione, per motivi diversi, è praticamente inesistente.

Il ragionamento che si fa è dunque il seguente:

1. alcune dichiarazioni fiscali sono effettivamente rappresentative del volume di affari del settore dichiarante;
2. il volume di affari che se ne deduce è un ottimo indicatore del totale degli acquisti del genere venduto dal settore, effettuato dai consumatori presenti nell’area;
3. esiste una relazione facilmente stimabile tra gli acquisti di alcuni generi e il totale degli acquisti fatti dai consumatori presenti.

Queste tre considerazioni sono tutte fondamentali per la realizzazione della stima che è stata prodotta e riportata nelle pagine che seguono e che punta quindi, in estrema sintesi, alla individuazione della relazione che lega le dichiarazioni IRAP che hanno le

¹ Il lavoro è stato svolto nell’ambito di un incarico affidato dall’Agenzia delle Entrate all’IRPET ed è stato realizzato da Stefano Casini Benvenuti dirigente IRPET e dal Prof. Alessandro Viviani dell’Università di Firenze che in particolare hanno curato l’Introduzione e la premessa e dalla dott. Valentina Tortolini che ha curato il resto del lavoro. Naturalmente l’impostazione e le conclusioni del lavoro sono condivise dai tre autori.

caratteristiche evocate nel primo dei punti precedenti al totale dei consumi interni a livello subregionale.

I due punti successivi ci indicano però che una stima diretta di questo tipo non sarebbe corretta, vi sono infatti motivi diversi che fanno sì che il volume d'affari del settore, seppure esente da fenomeni di evasione, non sia di fatto rappresentativo di tutti gli acquisti di quel genere effettuati dai consumatori presenti (ad esempio, non tutti i soggetti presenti in un comune comprano il giornale in quel comune); non solo, ma la relazione tra i consumi di alcuni generi (poniamo medicinali) e il volume complessivo dei consumi dipende anche dalle caratteristiche socio-economiche dei consumatori (età, livello culturale,...).

Tutte queste circostanze vanno quindi tenute presenti per ottenere una stima non distorta dei consumi interni a partire dalle dichiarazioni IRAP e ci impongono una corretta individuazione delle variabili esplicative dei diversi fenomeni sopra richiamati.

Resta, infine, da risolvere il problema principale di come stimare i parametri che consentono di passare dalle variabili esplicative alla variabile da stimare. La procedura utilizzata segue un approccio particolare riconducibile ai sistemi di stima per piccole aree. Il ragionamento seguito è molto semplice e consiste nello stimare la relazione tra la variabile dipendente (nel nostro caso i consumi interni) e un set di variabili ritenute esplicative del fenomeno a quel livello territoriale in cui tutte le variabili necessarie per effettuare la stima econometrica siano note (quindi il livello regionale). I parametri ottenuti serviranno per estendere la stima anche ai livelli territoriali inferiori, qualora ovviamente esistano informazioni sulle variabili esplicative.

L'approccio, metodologicamente consolidato, ha fornito risultati apprezzabili relativamente ai comuni (o loro aggregazioni) di tre regioni (Lombardia, Sicilia e Toscana) e può essere agevolmente esteso ad ogni altra regione italiana dal momento che si basa su dati disponibili presso l'Agenzia delle Entrate, integrati da altri dati presenti nelle statistiche ufficiali.

Naturalmente, non tutti i dubbi sono dissolti, da un lato, perché trattandosi di stimare una variabile latente non esistono riscontri empirici sulla bontà della stima, dall'altro perché restano ancora aperti alcuni dubbi oltre a quelli, sempre presenti, sulla qualità dei dati disponibili. Tra questi dubbi il più rilevante riguarda la scala territoriale minima di analisi; fino a che livello territoriale si può scendere, cioè, senza perdere di significatività nella stima. Per motivi che saranno richiamati nel testo la scala adottata è stata quella dei distretti fiscali.

1. Premessa

La presente ricerca costituisce un approfondimento ed una estensione di un precedente lavoro finalizzato alla stima dei consumi interni nei comuni della Toscana²; l'approfondimento nasce dalla esigenza di un migliore utilizzo delle informazioni disponibili; l'estensione si riferisce invece al fatto che la metodologia di stima è stata applicata anche ai comuni di Lombardia e Sicilia.

Come già sottolineato nella precedente ricerca, la necessità di ricorrere a modelli di stima per determinare i consumi interni a livello locale nasce dal fatto che le informazioni fornite dalle fonti ufficiali su tale grandezza sono disponibili solo a livello regionale.

Con l'espressione "consumi interni", si intende l'insieme dei consumi che diversi soggetti presenti in un determinato territorio –quindi sia residenti che non residenti– realizzano al suo interno in un determinato anno. Come dicevamo, tale grandezza è rilevata dalle fonti statistiche ufficiali solo a livello regionale, oltre che nazionale³, e viene distinta, a tale livello, anche per 12 funzioni di spesa (tavola 1).

Tavola 1. Le funzioni di spesa rilevate dall'ISTAT a livello regionale

1)	Generi alimentari e bevande non alcoliche
2)	Bevande alcoliche, tabacco, narcotici
3)	Vestiario e calzature
4)	Spese per l'abitazione, elettricità, gas ed altri combustibili
5)	Mobili, elettrodomestici, articoli vari e servizi per la casa
6)	Spese sanitarie
7)	Trasporti
8)	Comunicazioni
9)	Ricreazione e cultura
10)	Istruzione
11)	Alberghi e ristoranti
12)	Beni e servizi vari
	Totale

Se i dati sui consumi si fermano a livello regionale, a livello subregionale sono, invece, presenti informazioni su variabili che indirettamente possono essere ricondotte ai consumi dei soggetti presenti nell'area, nel senso che possono essere interpretate come causa o come effetto della capacità di spesa esercitata nell'area: l'ISTAT, ad esempio, a livello provinciale produce una contabilità ridotta (limitata a valore aggiunto ed unità di lavoro), mentre a livello comunale esistono, molte informazioni di tipo

² Casini Benvenuti, Viviani (2004).

³ Per le stime più aggiornate si veda ISTAT (2007).

amministrativo tra le quali, appunto, i dati relativi alle dichiarazioni IRAP, disponibili presso l'Agenzia delle Entrate ed utilizzati nella precedente stima.

L'ipotesi di fondo adottata in questo, come nel precedente lavoro, si basa sull'idea che esista una relazione tra i consumi interni (noti solo a livello regionale) e un insieme di variabili esplicative (note, invece, anche a livello provinciale e comunale) e che, una volta stimata a scala regionale questa relazione, essa possa essere estesa a tutti i livelli territoriali.

Rispetto al precedente lavoro è stato introdotto un elemento di importante novità con l'obiettivo di un utilizzo più efficiente dei dati disponibili. In particolare la stima precedente si limitava ad estendere immediatamente a livello comunale la relazione tra i consumi interni e le dichiarazioni IRAP stimata a livello regionale. In tal modo, la procedura non teneva in alcun conto le informazioni esistenti a livello provinciale che, essendo superiori (per quantità e qualità) a quelle note a livello comunale consentono, invece, di procedere ad una prima stima dei consumi interni provinciali. Questa stima, essendo basata su di una maggiore massa di informazioni, consente di ottenere una misurazione abbastanza attendibile dei consumi interni provinciali, consumi che nella seconda fase della procedura rappresentano la variabile dipendente del modello che consente di estendere la stima a livello comunale.

In sintesi, l'analisi è stata organizzata nei seguenti due passi successivi: inizialmente sono stati stimati i consumi interni a livello provinciale utilizzando come variabili esplicative alcuni dati relativi alle province italiane; successivamente, i consumi provinciali così stimati rappresentano la variabile dipendente per la stima a livello comunale, nella quale le variabili esplicative sono rappresentate dalle dichiarazioni IRAP su tabacchi, farmaci e giornali, opportunamente intergrate con altri indicatori in grado di influenzare la relazione tra consumi e le variabili suddette.

Vale la pena di richiamare in sintesi quelli che sono i presupposti di base di questa seconda fase del lavoro:

- a) dalle dichiarazioni IRAP sono stati estratti i dati delle vendite di tabacchi, giornali e farmaci assunte come rappresentative del totale dei consumi, avvenuti nell'area, dei generi cui l'IRAP si riferisce;
- b) esiste una relazione valida ovunque tra il totale dei consumi interni ed i consumi dei generi presi in esame
- c) quest'ultima relazione può essere influenzata da altre variabili, che vanno dunque identificate per ottenere una stima corretta dei consumi interni a partire dai consumi dei generi suddetti.

Il primo dei tre punti non è qui messo in discussione nel senso che le tre tipologie di beni sono scarsamente affetti da forme di evasione riconducibili alla sottofatturazione, per cui si può ragionevolmente assumere che le dichiarazioni IRAP corrispondano alle vendite effettive. Inoltre, le tipologie di produzione individuate rappresentano attività di servizio che si suppone siano fruite dalla popolazione presente e, quindi, approssimano la definizione di consumo che si intende stimare. Qualche dubbio si pone qui sul livello

di disaggregazione territoriale, dal momento che quando si scelgono livelli territoriali molto minuti è possibile che il genere preso in esame venga acquistato nel comune vicino. Non solo affinché l'affermazione a) sia valida occorre anche che nessun altro soggetto diverso da tabaccherie, giornali e farmacie venda i generi da loro venduti: come noto, invece, giornali e prodotti farmaceutici possono essere venduti anche in altri esercizi commerciali.

Per quel che riguarda, invece, il punto b) una prima idea del legame che può esservi tra consumi interni totali e consumi dei tre generi considerati può essere verificata a livello nazionale stimando appunto la relazione sulla serie storica dei consumi delle famiglie ed utilizzando le voci ISTAT più vicine a quelle di cui alle dichiarazioni IRAP utilizzate; in particolare:

- tabacchi
- prodotti medicinali, articoli sanitari e materiale terapeutico
- giornali, libri ed articoli di cancelleria

Le stime effettuate (tabella 1) mettono tutte in evidenza una positiva e in genere molto significativa relazione tra le tre voci ed i consumi interni totali. Delle tre grandezze tuttavia, sono i tabacchi quella meno significativa; inoltre l'elasticità dei consumi totali ai consumi dei tre generi è molto elevata nel caso dei tabacchi e dei giornali, mentre è bassa nel caso dei farmaci, lasciando intendere che i primi due sono generi più voluttuari (tab. 2).

Il punto c) richiama invece il fatto che le tre suddette relazioni sono in realtà influenzate anche da altre variabili. In effetti il rapporto esistente tra il consumo di ciascuno dei tre generi considerati ed i consumi totali dipende anche da molti altri fattori riconducibili sostanzialmente a due aspetti.

1. per ogni soggetto il rapporto tra i suoi consumi di tabacco, di giornali e di farmaci ed i suoi consumi totali dipende dalle sue caratteristiche; ricondotte a livello di area si può ipotizzare che cultura ed età siano altrettanti aspetti che incidono su tale rapporto;
2. il rapporto non è poi lo stesso a seconda del momento in cui il consumo del genere avviene; quando si compra un giornale in vacanza il rapporto tra la spesa complessivamente fatta in quel luogo e la spesa sostenuta presso il giornalaio è diverso da quello che avviene quotidianamente nei luoghi della propria residenza.

Poiché tutti questi fenomeni modificano il rapporto tra il consumo dei generi considerati ed i consumi totali occorre tenerne conto per arrivare ad una stima corretta di questi ultimi a partire da una conoscenza dei primi.

Tabella 1 Modello di regressione tra consumi interni e le tre voci di consumi considerate – Stima effettuata su variabili trasformate logaritmicamente

TABACCHI			
	<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>
Intercetta	2.660	2.032	1.309
Variabile X	1.128	0.218	5.182
<i>Statistica della regressione</i>			
R ²		0.441	
R ² corretto		0.425	
Errore standard		0.197	
n. osservazioni		36	
PRODOTTI MEDICINALI, ARTICOLI SANITARI E MATERIALE TERAPEUTICO			
	<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>
Intercetta	11.057	0.037	297.684
Variabile X	0.257	0.004	57.853
<i>Statistica della regressione</i>			
R ²		0.990	
R ² corretto		0.990	
Errore standard		0.026	
n. osservazioni		36	
GIORNALI, LIBRI ED ARTICOLI DI CANCELLERIA			
	<i>Coefficienti</i>	<i>Errore standard</i>	<i>Stat t</i>
Intercetta	0.746	0.660	1.131
Giornali	1.343 0.071	18.872	
<i>Statistica della regressione</i>			
R ²		0.913	
R ² corretto		0.910	
Errore standard		0.078	
n. osservazioni		36	

2. LA LETTERATURA SULLE STIME PER PICCOLE AREE

2.1 La natura del problema

Al tema dell'analisi statistica per piccole aree viene dedicata attualmente grande attenzione, a livello teorico ed empirico, per la crescente domanda di informazioni statistiche relative a territori e domini ridotti. Le stime di grandezze economico-sociali a livello di piccola aree sono sempre state di interesse generale; purtroppo tali stime risultano spesso disponibili (ed affidabili) solo in occasione dei censimenti, o saltuariamente, attraverso indagini ad hoc. Nel periodo intercensuario, quando le rilevazioni sono campionarie, l'accuratezza delle stime a livello di piccole aree è molto limitata e risultano affidabili soltanto quelle a livello superiore: esempio emblematico delle due situazioni sono i dati a livello comunale e regionale rispettivamente.

Se le indagini si rivolgono a macro aree, quali il territorio nazionale, un insieme di regioni (circoscrizione territoriale) o regione, i dati sono ottenuti sulla base di disegni di campionamento, più o meno complessi, che coinvolgono un campione numeroso in modo tale da garantire una conoscenza affidabile delle grandezze oggetto d'indagine e di studio. Se, invece, si è interessati a conoscere il valore di talune grandezze con riferimento ad aree di ridotte dimensioni (comuni o frazioni) è necessario che queste siano considerate come domini al momento del disegno dell'indagine; in caso contrario, non si potranno utilizzare gli stimatori basati sul disegno relativo all'indagine perché produrrebbero stime non affidabili.

Una soluzione banale (ma costosa), a questo riguardo, potrebbe essere costituita da un'estensione del campione, ma si comprende che il problema dell'accuratezza delle stime si può porre comunque, in quanto dipendente dalla dimensione dell'area cui si è interessati. In generale, quindi, la necessità di disporre di informazioni a livello di piccole aree stimola la ricerca di metodologie, che utilizzando tutte le informazioni disponibili, siano in grado di produrre stime sufficientemente accurate a tale livello.

Come fa notare Chiandotto (1996), la scelta di una metodologia di stima dipende essenzialmente dal tipo di dati disponibili. Tuttavia, una caratteristica comune a tutti i problemi di stima per piccole aree è rappresentata dalla esiguità di dati ed informazioni a livello disaggregato, mentre i dati sono disponibili –ed affidabili– a livello aggregato. Un esempio, relativo al tema qui affrontato, può chiarire questo aspetto: si conoscono i livelli di reddito e di consumo a livello regionale, unitamente ad una serie di variabili che ne consentono l'analisi, e si vuole disporre di quei dati a livello comunale.

Una soluzione a questo problema è l'utilizzo di stimatori che consentono di migliorare l'efficienza delle stime rispetto a quello che si otterrebbe sulla base del disegno di campionamento. In questo contesto si collocano i metodi di analisi statistica relativi alla stima per piccole aree (Small Area Estimation – SAE) che fanno uso di informazioni ausiliarie (Estevao e Särndal, 1999). Le informazioni ausiliarie possono

essere note a livello d'unità campionaria oppure possono essere riferite alle piccole aree d'interesse, ma in ogni caso, i modelli di stima per piccole aree ipotizzano effetti casuali non correlati fra zone. È, inoltre, possibile ipotizzare la presenza di una relazione spaziale tra variabili rilevate su aree territoriali definite vicine, secondo un criterio di distanza e, quindi, considerare la posizione geografica di ciascun'area, assumendo che gli effetti casuali fra aree vicine siano correlati e che l'intensità di tale correlazione diminuisca all'aumentare della distanza. Quindi, con i metodi di stima per piccole aree si supera il problema dovuto alla ripartizione geografica o ai domini più ristretti rispetto a quelli previsti dal disegno campionario originario.

Un vantaggio di tali modelli deriva dalla possibilità di utilizzare fonti statistiche già esistenti come i dati censuari, gli archivi amministrativi, i dati di diverse indagini campionarie, etc., combinandole insieme al fine di ottenere una stima affidabile per il carattere desiderato nell'ambito dell'analisi svolta. Non risulta necessario, quindi, progettare indagini "ad hoc" per ottenere le informazioni desiderate, con il risultato di una notevole diminuzione di costi e di tempi.

Quella che seguirà è una trattazione introduttiva all'argomento, dove saranno esaminati aspetti generali teorici che riguardano la stima per piccole aree.

2.2 I metodi di stima

Per prima cosa si deve definire cosa si intende per "piccola area".

Brackstone (1987) definisce piccola area una qualunque area per cui non si possono ottenere stime accurate (accettabili) con i metodi di stima classici dato un certo campione, ma è necessario utilizzare nuovi metodi di stima.

La piccola area (Rao, 2003) è un'area geografica di piccole dimensioni, come può essere una provincia, un comune od una sezione di censimento, dove la numerosità campionaria non consente di ottenere stime accurate. Il termine spesso è usato come sinonimo di piccolo dominio, cioè di sub-popolazioni segmentate in base a caratteristiche quali sesso, razza, età, che appartengono ad un'unità geografica più ampia, che solitamente è la nazione. Questo indica che la piccola area non è un concetto di significato assoluto, legato a parametri fisici, ma relativo poiché per definirlo bisogna fare riferimento alla dimensione campionaria e al processo inferenziale. Nel caso qui trattato, la dimensione di riferimento è quella comunale, in quanto sono disponibili informazioni "attendibili" a livello regionale su consumi e reddito, grandezze per le quali c'è l'obiettivo di ottenerne una stima ai livelli istituzionali inferiori.

I metodi di stima per piccole aree permettono di ottenere, con riferimento a porzioni di territorio ristrette, stime più efficienti rispetto a quelle che si otterrebbero basandoci sul disegno di campionamento.

Questi possono essere classificati in tre categorie secondo il tipo di inferenza:

- metodi basati su disegno;
- metodi assistiti da modello;
- metodi basati su modello.

a) I metodi basati su disegno (o campionari) forniscono direttamente l'espressione di stimatori per il parametro caratteristico di piccola area, senza assumere esplicitamente un modello di distribuzione del fenomeno tra aree. Il parametro d'interesse della piccola area è stimato utilizzando i metodi campionari classici fondati sulla distribuzione di probabilità indotta dal disegno di campionamento: si riportano le osservazioni del campione alle piccole aree e si stima direttamente il valore del parametro in ogni area.

A seconda dei dati disponibili per le aree di riferimento, si possono costruire stimatori per piccole aree diretti o indiretti: il parametro da stimare è pensato come una costante. Lo stimatore diretto della variabile Y_i relativo all'area i -esima, è ottenuto a partire dai valori della variabile obiettivo osservati con riferimento alle sole unità del campione appartenenti alla piccola area i . Questo stimatore è corretto sulla base del disegno, ma può essere caratterizzato da elevata variabilità. Il problema principale è che la variabilità degli stimatori cresce al diminuire della numerosità del campione; inoltre, può essere che nessuna unità campionaria sia presente nella piccola area, impedendo di ottenere una stima della grandezza d'interesse.

b) Metodi assistiti da modello: sono quei metodi nei quali l'inferenza è basata sia sul disegno che sul modello. A partire dal disegno di campionamento, si ottengono stimatori corretti utilizzando il disegno stesso ed una serie di ipotesi fatte dal ricercatore sul legame fra variabili ausiliarie e vettore dei parametri incogniti riferiti alla piccola area, al fine di ridurre la variabilità. Nella sostanza, in tutti questi metodi, diretti o indiretti che siano, la piccola area assume un peso che è proporzionale alle unità del campione in essa contenute.

c) Metodi basati su modello: il parametro oggetto di studio non è visto come una costante, ma come una variabile casuale. L'approccio prevede l'introduzione esplicita di un modello probabilistico di superpopolazione relativo alla distribuzione del fenomeno tra le aree, da cui derivare il predittore ottimo corretto a livello di piccola area. Vi appartengono i modelli di piccola area che prevedono la presenza di effetti casuali di area (Rao, 2003).

2.3 I metodi basati su modello

I metodi di stima per piccole aree basati su modello prevedono l'introduzione di modelli probabilistici che assumono effetti casuali specifici di area per esprimere la variabilità dei valori della variabile d'interesse tra aree, oltre quella spiegata dalle variabili ausiliarie incluse nel modello (Pfeffermann, 2002).

Nello scegliere il modello di superpopolazione adeguato alle finalità dell'analisi, è importante trovare un compromesso fra un numero di parametri elevato, che meglio permetta di rappresentare la popolazione oggetto e la numerosità esigua del campione (Cicchitelli G., Herzel e A.Montanari,G.E.; 1992),. Considerando che tali metodi prevedono l'indipendenza della stima dal disegno campionario e l'utilizzo di un modello probabilistico, ponendo specifiche condizioni, è possibile stimare un parametro anche per un dominio non campionato.

Tali modelli sono classificati in due categorie:

- modelli a livello di area (Area Level Random Effects Model), che legano la stima diretta di piccola area con i valori di covariate specifiche dell'area stessa (esiste un unico dato riassuntivo della piccola area per ogni covariata). L'impiego di questi modelli è necessario quando non sono disponibili informazioni ausiliarie a livello di unità campionarie;
- modelli a livello di unità, che mettono in relazione i valori delle variabili di studio osservati sul campione con quelli delle covariate disponibili a livello di unità campionarie e quindi si stima il parametro utilizzando le covariate osservate su ogni unità della piccola area.

La scelta del modello è importantissima, poiché l'inferenza si basa sulla distribuzione probabilistica imposta dal modello assunto: se il modello scelto non si adatta ai dati, gli stimatori basati su tale modello saranno distorti e le stime sbagliate.

I vantaggi di tali metodi sono vari: si possono ottenere stimatori ottimi di piccola area, si può associare una misura di variabilità ad ogni piccola area, si può scegliere il modello più adatto rispetto alla natura della variabile oggetto di studio, si può stimare il valore del parametro d'interesse anche in quelle aree che non sono rappresentate nel campione.

Il modello applicato è quello a livello di area, con variabili ausiliarie i cui valori sono riferiti ai comuni.

2.4 Il modello a livello di area

Il presupposto consiste nel disporre di una serie di osservazioni estratte da una popolazione (indagine ISTAT) e riferite ad un'area geografica. Questa viene partita in piccole aree, come possono essere le aree censuarie e, nel caso, i comuni. Si ipotizza di avere un certo numero di osservazioni, non necessariamente per ogni piccola area. Allo stesso tempo si suppone di avere, per ogni piccola area, delle covariate. Le informazioni ausiliarie devono essere disponibili sia per le piccole aree campionate, che per quelle dove nessuna unità è stata estratta. Tramite un modello statistico che utilizza sia le osservazioni campionate, sia le variabili ausiliarie, si stima la media del carattere cui siamo interessati nella popolazione e vi si associa il suo EQM (Errore Quadratico

Medio), con l'obiettivo di ottenere una stima corretta ottima: devono quindi essere rispettati i principi di correttezza, efficienza e consistenza della stima (Petrucci A., Pratesi M. e Salvati N.; 2003),.

Si specifica che:

- le piccole aree oggetto dell'analisi sono m ;
- Θ è il parametro della popolazione che si vuole stimare;
- ci sono p variabili ausiliarie (covariate), indicate con $x = (x_{i1}, x_{i2}, \dots, x_{ip})^T$

Si assume che:

$$\theta_i = x_i^T \beta + z_i u_i \quad (1.1)$$

cioè il parametro d'interesse è legato attraverso una certa relazione alle variabili ausiliarie, dove z_i sono costanti positive note, β è il vettore di regressione e u_i sono variabili casuali indipendenti ed identicamente distribuite con

media $E(u_i) = 0$ e varianza $V(u_i) = \sigma_u^2$

Per stimare θ_i ci si avvale del modello

$$\hat{\theta}_i = \theta_i + e \quad (i = 1 \dots m) \quad (1.2)$$

Dove e_i è l'errore di campionamento nell'area i con $E(e_i | \theta_i) = 0$ e $V(e_i | \theta_i) = \psi_i$. Questo significa che lo stimatore è corretto rispetto al disegno.

Unendo la (1.1) con (1.2) si ottiene il modello lineare ad effetti misti di Fay e Herriot (1979):

$$\hat{\theta}_i = x_i^T \beta + z_i u_i + e_i \quad (i = 1 \dots m) \quad (1.3)$$

La stima (1.3) è formata da una componente sistematica $x_i^T \beta$, e da due componenti casuali: u_i che rappresenta l'effetto dovuto all'area i (la piccola area trattata) ed e_i che rappresenta l'errore dovuto al campionamento. Si assume che gli effetti casuali di area u_i e gli errori di campionamento e_i siano indipendenti, cioè non correlati: $E(e_i | u_i) = 0$. Da tutte queste ipotesi risulta che la stima per piccola area è corretta.

2.5 I predittori BLUP ed EBLUP

Assumendo il modello 1.3 si può ottenere il predittore BLUP⁴ per la stima di parametri della popolazione. Questo è un predittore⁵ ottimo lineare e corretto, vale a dire che è lo stimatore che minimizza l'errore quadratico medio nella classe degli stimatori lineari corretti. Quindi è: lineare, in quanto funzione lineare dei dati; corretto, in quanto il valore stimato corrisponde al valore vero; ottimo, poiché ha la varianza più piccola fra tutti gli stimatori lineari corretti.

Il predittore BLUP, a partire dal modello lineare ad effetti misti, permette di stimare valori della popolazione solo quando il vettore dei parametri è noto. Molto sinteticamente, rimandando per maggiori approfondimenti e spiegazioni a Rao (2003), illustriamo la stima BLUP, funzione delle osservazioni y e del vettore delle variabili ausiliarie δ :

$$\hat{T}(\delta, y) = \hat{u} = l^T \hat{\beta} + m^T \hat{u} \quad (1.4)$$

dove:

- $\hat{\beta}$ è la stima del vettore dei coefficienti di regressione;
- \hat{u} è la stima del vettore che rappresenta l'effetto casuale di area;
- l è il vettore delle covariate dell'area di interesse ed m è un vettore che può assumere valore 1 o 0, a seconda che si utilizzi un modello a livello di area o un modello a livello di unità.

Nella realtà pratica, però, molto spesso i valori delle componenti di varianza, cioè il valore delle variabili ausiliarie, quindi il vettore dei parametri inseriti nel modello (δ), non sono noti. È possibile sostituire questi valori con le loro stime, tramite l'utilizzo del predittore empirico ottimo lineare e corretto (EBLUP)⁶: $\hat{T} = \left(\hat{\delta}, y \right)$.

⁴ Best Linear Unbiased Predictor

⁵ La distinzione tra predittore e stimatore si riferisce al fatto che il primo è uno stimatore degli effetti casuali, mentre il secondo degli effetti fissi: lo stimatore fa riferimento ad un modello lineare ad effetti fissi $y = X\beta + e$, dove la variabilità della variabile casuale di studio è interamente spiegata dalle covariate della X ; il predittore invece si basa sul modello a effetti misti $y = X\beta + Zu + e$, nel quale Zu è la variabilità, dovuta alla relazione fra aree (che non è spiegata nel modello a effetti fissi)

⁶ Empirical Best Linear Unbiased Predictor

3 IL MODELLO DI STIMA ADOTTATO

3.1 Una stima a due stadi

Come abbiamo già ricordato, la logica di fondo del modello utilizzato ripercorre quella già impiegata nel precedente lavoro, ma con l'introduzione di una fase intermedia. L'idea di base è sempre quella di risalire alla stima dei consumi interni comunali a partire dalle alcune "tracce" lasciate sul territorio dai consumatori: nel nostro caso le tracce sono rappresentate dai consumi di alcuni generi così come vengono rilevati dalle dichiarazioni IRAP. Dalle tracce, tuttavia, è possibile risalire ai consumi solo se si conosce la relazione che lega le tracce (dichiarazioni IRAP) al fenomeno sotto osservazione (i consumi interni), una relazione, questa, stimabile solo al livello territoriale per cui si dispone simultaneamente delle informazioni sulle dichiarazioni IRAP ed i consumi interni, quindi il livello regionale o, come vedremo, quello provinciale.

Secondo questa procedura, le dichiarazioni IRAP, di fatto, servono per disaggregare a livello subregionale i dati sui consumi interni rilevati a livello regionale. La procedura proposta si colloca quindi nell'ambito dei "metodi basati su modello" richiamati nel capitolo precedente

L'elemento di maggiore novità rispetto alla precedente procedura di stima consiste nell'aver introdotto una fase intermedia finalizzata ad una stima dei consumi interni a livello di provincia. Questa novità nasce dalla esigenza di sfruttare al meglio le numerose informazioni esistenti a tale livello e che consentono di procedere ad una stima attendibile dei consumi interni provinciali. I dati IRAP vengono pertanto utilizzati in una seconda fase e servono per disaggregare a livello subprovinciale le stime dei consumi interni ottenute a livello provinciale.

In sintesi, indicando con i pedici c , p , r i livelli territoriali, rispettivamente, comunali, provinciali e regionali, la stima passa attraverso due fasi:

1. livello regionale: $C_r = f(\sum_{p \in r} V_p)$ da cui si ottiene un stima di C_p (consumi provinciali);
2. livello provinciale: $C_p = f(\sum_{c \in p} X_c)$ da cui si ottiene un stima di C_c (consumi comunali).

Nella prima fase, come dicevamo, la massa di informazioni disponibili è certamente significativa e consente una buona stima dei consumi provinciali. In particolare, a tale livello, disponiamo di stime ISTAT sul valore aggiunto provinciale pro capite, inteso qui come una buona approssimazione del livello di reddito prodotto dell'area.

Sappiamo, inoltre, che sui consumi interni incidono anche i consumi dei non residenti, ovvero dei lavoratori pendolari e, soprattutto, dei turisti. Quindi per procedere ad una stima corretta dei consumi interni utilizzando come principale variabile esplicativa il valore aggiunto prodotto nell'area occorre introdurre almeno due correttivi volti a cogliere la diversa presenza di lavoratori pendolari e turisti.

L'incidenza dei due fenomeni la si può cogliere dagli indicatori sulle presenze turistiche nelle strutture ricettive ufficiali aggiornate annualmente (anche se, salvo il caso della Toscana, non smarriva mai a livello di comune) e dagli indicatori sui movimenti pendolari per motivi di studio e lavoro rilevanti con i censimenti della popolazione.

Per quanto riguarda le informazioni sulle presenze turistiche occorre fare una duplice ipotesi. La prima, la più plausibile, è il che il numero di turisti sia un buon indicatore della funzione di spesa di questa categoria di non residenti. In secondo luogo, dobbiamo supporre che il numero di turisti in alloggi privati (non registrati dalle fonti ufficiali) sia ovunque esattamente proporzionale alle presenze turistiche registrate in strutture ricettive ufficiali. Vi sono in questa seconda ipotesi alcune evidenti semplificazioni, dal momento che l'uso di seconde case (in affitto o in proprietà) è diffuso ovviamente in località turistiche, ma non necessariamente vi è, tra questa tipologia di turismo e quella che fa riferimento alle strutture alberghiere o extra-alberghiere, una relazione strettamente proporzionale; non mancano infatti i casi in cui le due formule ricettive siano tra loro alternative.

Purtroppo non sono disponibili informazioni che siano in grado di cogliere densità e effettivo utilizzo di seconde case, per cui l'ipotesi di inserire le presenze ufficiali rilevate presso le strutture ricettive alberghiere ed extra come *proxy* del totale delle presenze turistiche era l'unica accettabile.

Non solo, ma è solo per la Toscana che si dispone al momento di informazioni ufficiali sulle presenze turistiche a livello comunale, riconducibili quindi ai raggruppamenti dell'Agenzia delle Entrate della Toscana, mentre per Lombardia e Sicilia le informazioni sono più aggregate e si riferiscono alle circoscrizioni turistiche di scala sovracomunale. In questo caso siamo stati costretti a riportare le presenze a livello comunale sulla base della dotazione di strutture ricettive di ciascun comune, introducendo implicitamente l'ipotesi di un uguale tasso di utilizzo delle strutture all'interno della stessa circoscrizione.

Se per i turisti sono presenti dati ufficiali, anche se incompleti, le informazioni relative ai pendolari sono ancora più scarse. Da qui la necessità di elaborare un indicatore che rappresenti il livello di mobilità nel territorio di riferimento. L'indicatore scelto è il rapporto tra occupati nelle unità locali (esclusa l'agricoltura) ubicate nel territorio di riferimento e gli occupati (sempre esclusa l'agricoltura) residenti nello stesso territorio; i dati utilizzati per il calcolo dell'indicatore sono di fonte ISTAT (censimento dell'industria e censimento della popolazione) Quando questo indicatore assume valori minori di uno significa che il flusso di pendolari è verso l'esterno del territorio, mentre i valori maggiori di uno indicano un flusso in entrata nel territorio.

Analizzando i dati relativi ai due suddetti fenomeni limitatamente alla Toscana (tab. 2), si osservano differenze rilevanti in entrambi i casi, ma mentre per i flussi pendolari le differenze variano tra il minimo di Aulla (0,75) ed il massimo di Firenze 3 (1,368), nel caso del turismo le differenze sono assai più accentuate passando dal minimo di 1344 per San Miniato al massimo di 109.531 per Portoferraio. In questo secondo caso i valori

massimi stanno ben al di fuori del campo di oscillazione presente a livello provinciale, che come già ricordato è la base sulla quale viene stimata la funzione del consumo.

Tabella 2 Pendolarità e presenze turistiche per raggruppamenti comunali delle agenzie delle entrate in toscana. 2001

Agenzia	Pendolarità	Presenze turistiche per 1000 ab.
Arezzo	1,031	2.944
Aulla	0,750	2.030
Borgo San Lorenzo	0,794	6.298
Carrara	0,927	1.472
Castelnuovo Garfagnana	0,804	3.485
Empoli	0,954	3.976
Firenze 1	1,196	16.238
Firenze 2	0,872	4.474
Firenze 3	1,368	6.661
Grosseto	0,910	20.965
Livorno	0,914	13.246
Lucca	1,073	3.483
Massa	0,983	19.795
Massa Marittima	0,876	16.573
Montepulciano	0,944	25.577
Montevarchi	0,862	2.347
Orbetello	0,907	28.291
Pescia	0,919	21.729
Pietrasanta	0,994	21.182
Piombino	1,074	21.861
Pisa	0,939	9.636
Pistoia	0,932	2.765
Poggibonsi	0,934	8.974
Pontedera	0,975	5.594
Portoferraio	0,972	109.531
Prato	1,030	1.854
San Miniato	1,105	1.344
San Sepolcro	0,943	2.742
Siena	1,080	14.998
Viareggio	0,893	14.232

Fonte: elaborazioni IRPET su dati ISTAT

La base informativa principale della seconda fase della procedura di stima adottata è rappresentata dalle dichiarazioni IRAP relative a tabaccherie, giornali e farmacie; sono queste le informazioni su cui è necessario fare le ipotesi più forti. In primo luogo dobbiamo ipotizzare che i dati IRAP non siano soggetti ad evasione fiscale o distorsione di altro tipo.

La natura dei beni presi in esame ci fa, tuttavia, supporre che tali dati non dovrebbero contenere elementi di evasione e quindi dovrebbero essere una rappresentazione fedele dei consumi di tabacchi, giornali e farmaci nelle aree considerate.

Ciò nonostante, rispetto a questa affermazione occorre introdurre alcuni elementi dubitativi non tanto sulla possibilità che vi sia una qualche forma di evasione, quanto sul fatto che le vendite dei generi suddetti, registrata nelle dichiarazioni IRAP delle tre tipologie di esercizio, sia effettivamente rappresentativa del totale dei consumi dei tre generi avvenuti all'interno dell'area: vi possono infatti essere altri esercizi che vendono gli stessi generi (si pensi ai giornali nei supermercati), oppure alcuni residenti dell'area che possono comprare i generi suddetti nelle aree vicine. Questo secondo problema può essere, in realtà, risolto adottando un criterio di territorializzazione che superi i suddetti problemi: il criterio adottato è, come nel precedente lavoro, le zone di competenza dell'agenzia delle entrate⁷.

Una prima analisi esplorativa dei dati può consentire una valutazione indiretta della affidabilità dei dati disponibili attraverso un confronto temporale (i due anni disponibili) e spaziale (il dettaglio provinciale) degli stessi. In realtà le differenze tra i due anni sono spesso significative e tali da impedirci di utilizzare il *pool* complessivo delle informazioni. In particolare per i tabacchi possiamo osservare che le differenze tra i dati 2000 e i dati 2001 variano da un +15% ad un -25%, per i farmaci da un -4% ad un +61% e per le edicole da un -36% ad un +9%. Per questo motivo si è preferito utilizzare per la stima dei consumi comunali i dati IRAP, di un solo anno (2001).

3.2 La metodologia adottata

Il nostro obiettivo è dunque quello di stimare i consumi che tutti i soggetti (residenti e non) presenti su un dato territorio (il comune) effettuano nell'arco di un anno a partire dai dati delle dichiarazioni IRAP, che rappresentano delle "tracce" che i soggetti lasciano durante la loro permanenza nel luogo, e dai consumi interni a livello regionale registrati nello stesso arco di tempo.

Invece di stimare direttamente i consumi comunali partendo da quelli regionali si è preferito effettuare un step intermedio passando per i consumi provinciali e poi estendendo l'analisi all'interno dei comuni della provincia.

In altre parole, relativamente alle 20 regioni, si procede alla stima della relazione tra i consumi interni regionali C_r e le tracce $V_{r,j} = \sum_{p \in r} V_{p,j}$ dove $V_{p,j}$ rappresentano i dati regionali ottenuti come somma dei dati provinciali della traccia j -esima. La relazione

⁷ le zone adottate per la Toscana sono, ad esempio, le seguenti:

1 Arezzo	7 Firenze 1	13 Massa	19 Pietrasanta	25 Portoferraio
2 Aulla	8 Firenze 2	14 Massa Marittima	20 Piombino	26 Prato
3 Borgo S. Lorenzo	9 Firenze 3	15 Montepulciano	21 Pisa	27 San Miniato
4 Carrara	10 Grosseto	16 Montevarchi	22 Pistoia	28 San Sepolcro
5 Castelnuovo G.	11 Livorno	17 Orbetello	23 Poggibonsi	29 Siena
6 Empoli	12 Lucca	18 Pescia	24 Pontedera	30 Viareggio

stimata a livello regionale verrà applicata a V_p per stimare i consumi provinciali C_p , ponendoci in vincolo che $C_r = \sum_{p \in r} C_p$.

Una volta ottenuti i consumi provinciali C_p si procede alla stima dei consumi comunali in modo analogo, ovvero si stima la relazione tra C_p e le tracce $X_{c,i} = \sum_{c \in p} X_{c,i}$.

4. LE STIME OTTENUTE

4.1 I consumi provinciali

I consumi provinciali sono stati stimati sfruttando la relazione esistente tra consumi interni regionali e valore aggiunto (sempre regionale) e inserendo nella relazione l'informazione relativa ai turisti al fine di valutare la popolazione insistente sul territorio e non solo quella residente. Non sono stati considerati i pendolari perché si suppone che il saldo dei pendolari a livello regionale non contenga differenze tali da rendere significativa la variabile.

I dati a nostra disposizione sono dunque quelli riportati nella tabella seguente:

Variabile	Regioni	Province
Valore Aggiunto	Noto	Noto
Turisti	Noto	Noto
Consumi Interni	Noto	Ignoto - Stimato

La stima della relazione tra consumi regionali valore aggiunto e turisti è stata effettuata con il seguente modello lineare nei logaritmi:

$$\ln(C_r) = \ln \alpha_0 + \alpha_1 \ln(V_{r,1}) + \alpha_2 \ln(V_{r,2}) + \epsilon_r$$

Dove:

C_r = consumi interni regionali con $r = 1, \dots, 20$;

$V_{r,1}$ = valore aggiunto della regione r ;

$V_{r,2}$ = turisti registrati nella regione r .

Per le considerazioni effettuate precedentemente, in particolare per il fatto che le osservazioni a nostra disposizione erano solo 20 e poiché i dati relativi alla variazione di valore aggiunto regionale mostrano un andamento simile per tutte le regioni per i due anni di osservazione (2000 e 2001) lasciando ipotizzare l'assenza di fenomeni particolari che potevano risultare distorcitori dal punto di vista della stima, si è optato per un modello *pooled* che considerasse simultaneamente le informazioni del biennio 2000-2001.

Nel modello utilizzato, la variabile dipendente “consumi” è stata stimata attraverso due variabili indipendenti: il valore aggiunto e i turisti. L’adattamento del modello è risultato molto buono –l’R² corretto è infatti uguale a 0,947- inoltre come era logico aspettarsi, entrambi i coefficienti dei predittori sono positivi ed altamente significativi.

Tabella 3. Il modello utilizzato

Model	R	R ²	R ² aggiustato	Std. Error della stima
1	0,974(a)	0,949	0,947	0,04963
	coefficiente	Std. Error	t di Student	Significatività
Intercetta	2,724	0,368	7,403	0,000
Valore aggiunto	0,675	0,039	17,344	0,000
Turisti	0,059	0,012	4,871	0,000

Il significato dei coefficienti di regressione è, in questo caso, abbastanza intuitivo, e possiamo leggerli in termini di elasticità. Il valore 0,675 associato al valore aggiunto (pro capite) indica che all’aumento dell’1% di questa variabile i consumi interni registrano un aumento dello 0.675%, rimanendo costanti i turisti. Viceversa un aumento dei turisti, rispetto alla popolazione residente, dell’1% porta un incremento del valore aggiunto dello 0,059%. Attraverso i coefficienti standardizzati possiamo valutare l’importanza relativa delle singole variabili indipendenti nell’influenzare il livello di consumi, e come era logico aspettarsi il valore aggiunto ha un effetto più intenso di quello dei turisti.

I risultati ottenuti, a livello regionale, sono riportati in tabella 4 e mostrano un buon adattamento rispetto ai valori osservati. Come si può rilevare il risultato peggiore delle stime è stato ottenuto per la Valle d’Aosta, dove la differenza tra consumi osservati e stimati è del 13,12%; un risultato, questo, non particolarmente sorprendente, soprattutto considerando le dimensioni esigue, in termini di abitanti, e l’elevato flusso turistico che rendono questa regione un “caso unico” nel panorama nazionale.

I parametri così ottenuti sono stati applicati ai dati provinciali per stimare i consumi per ognuna delle 103 province italiane. Nella tabella seguente si riportano i dati relativi alla province di Lombardia, Toscana e Sicilia.

Tabella 4 Valori osservati e stimati. Valori in migliaia di euro

Regione	Consumi stimati	Consumi osservati	Differenza
Piemonte	57.471.572	59.767.330	3,99
Valle d'Aosta	2.000.664	2.263.217	13,12
Lombardia	134.715.381	134.832.300	0,09
Trentino Alto Adige	16.535.625	15.913.990	-3,76
Veneto	67.973.874	65.412.260	-3,77
Friuli Venezia Giulia	17.185.915	16.690.380	-2,88
Liguria	22.748.376	23.961.790	5,33
Emilia Romagna	62.518.285	62.653.910	0,22
Toscana	51.013.600	50.670.820	-0,67
Umbria	10.772.201	10.471.110	-2,80
Marche	19.785.200	19.302.060	-2,44
Lazio	72.902.526	72.971.690	0,09
Abruzzo	14.776.278	14.149.360	-4,24
Molise	3.366.465	3.309.094	-1,70
Campania	54.891.005	53.536.590	-2,47
Puglia	37.794.791	39.749.930	5,17
Basilicata	5.932.352	5.473.771	-7,73
Calabria	18.467.311	19.498.570	5,58
Sicilia	47.429.195	49.426.780	4,21
Sardegna	17.785.600	17.553.280	-1,31

Tabella 5 Consumi stimati per provincia in: Sicilia, Toscana e Lombardia. Valori in migliaia di euro

Regione	Provincia	Consumi stimati
Sicilia	Agrigento	4.029.072
Sicilia	Caltanissetta	2.351.143
Sicilia	Catania	10.415.839.
Sicilia	Enna	1.479.758
Sicilia	Messina	7.261.960
Sicilia	Palermo	12.385.304.
Sicilia	Ragusa	3.097.305
Sicilia	Siracusa	4.334.272
Sicilia	Trapani	4.072.124
Sicilia	Arezzo	4.167.764
Toscana	Firenze	15.272.997.
Toscana	Grosseto	2.779.929
Toscana	Livorno	4.879.615
Toscana	Lucca	5.272.864
Toscana	Massa-Carrara	2.344.213
Toscana	Pisa	5.344.634
Toscana	Pistoia	3.675.502
Toscana	Prato	3.200.562
Toscana	Siena	3.732.735
Lombardia	Bergamo	14.201.965.
Lombardia	Brescia	14.982.548.
Lombardia	Como	8.704.614
Lombardia	Cremona	4.959.619
Lombardia	Lecco	4.253.393
Lombardia	Lodi	2.842.303
Lombardia	Mantova	5.416.618
Lombardia	Milano	57.432.138.
Lombardia	Pavia	7.586.806
Lombardia	Sondrio	2.477.462
Lombardia	Varese	11.974.828.

4.2 Le stime a livello locale

I consumi provinciali così stimati sono stati utilizzati come variabile dipendente della seconda fase della stima, quella cioè rivolta alla stima dei consumi a livello comunale. In altre parole, l'ipotesi adottata è che i consumi suddetti siano di fatto quelli "veri" a livello provinciale e che, a partire da essi, sia possibile stimare una loro relazione con le variabili esplicative note a livello comunale e rappresentate dalle dichiarazioni IRAP.

Poiché, però, la relazione tra consumi dei tre generi di prodotti presi in esame e consumi complessivi è verosimilmente influenzata anche da altre grandezze la stima corretta della relazione richiede che anche tali altre variabili vengano prese in esame.

In particolare si tratta di inserire, come altre esplicative della relazione di base tra dichiarazioni IRAP e consumi interni, l'informazione relativa ai turisti e ai pendolari. Il senso di questo inserimento va chiarito con precisione. Di fatto si parte dalla considerazione che vi sia una relazione valida ovunque tra consumi interni di un'area e i consumi del genere di cui si dispone delle dichiarazioni IRAP (supponiamo i giornali); tale relazione dovrebbe valere per ciascun soggetto che acquista un giornale ad indicare che per ogni acquisto di giornale vi è alle spalle una certa spesa per consumi all'interno di quell'area da parte del soggetto acquirente. Si comprende, invece, bene che il passaggio da acquisto di giornale e acquisti complessivi effettuati nell'area dipende anche da altre circostanze: probabilmente l'età, il titolo di studio, ma sicuramente anche il fatto di essere un turista od un lavoratore che non risiede all'interno dell'area. Nei vari tentativi di stima effettuati età e titolo di studio non sono mai risultati significativi, per cui assieme alle dichiarazioni IRAP abbiamo considerato le presenze turistiche e il pendolarismo:

Variabili esplicative	Province	Comuni
Dichiarazione IRAP Tabacchi	Noto	Noto
Dichiarazione IRAP Farmaci	Noto	Noto
Dichiarazione IRAP Giornali	Noto	Noto
Turisti	Noto	Noto
Pendolari	Noto	Noto
Consumi Interni	Noto	Ignoto - Stimato

La stima della relazione tra consumi provinciali dichiarazioni IRAP turisti e pendolari è stata effettuata con il seguente modello lineare nei logaritmi:

$$\ln(C_p) = \ln \alpha_0 + \alpha_1 \ln(P_{p,1}) + \alpha_2 \ln(T_{p,2}) + \alpha_j \ln(D_{p,j}) + \epsilon$$

Dove:

C_p = consumi interni provinciali con $r = 1 \dots 103$;

$P_{p,1}$ = pendolari della provincia p ;

$T_{p,2}$ = turisti registrati nella provincia p ;

$D_{p,j}$ = IRAP dichiarata per il bene j con $j = 1..3$ (tabacchi, farmaci, giornali).

Per le osservazioni riportate precedentemente, ovvero poiché la variazione dei dati IRAP mostrano un andamento molto diverso per le diverse province per due anni di osservazione (2000 e 2001), si è optato di considerare nel modello solo le informazioni del 2001, inserendo nella stima tutte le province italiane (103) in modo da avere un cospicuo numero di osservazioni omogenee a cui applicare il modello.

Sono stati elaborati diversi modelli che prendessero in considerazione le tre variabili IRAP simultaneamente o singolarmente. Dall'analisi effettuata è risultato che i dati sui Tabacchi sono quelli meno informativi: questa variabile è sempre risultata non significativa.

Inoltre i modelli con le tre variabili IRAP considerate simultaneamente o a coppie, sono risultati peggiori dei modelli in cui si utilizza una sola delle variabili IRAP: come era logico aspettarsi le variabili tabacchi farmaci e giornali sono altamente correlate tra loro questo comporta che utilizzandole insieme nello stesso modello una o più variabili non risultano significative.

Indici di correlazione tra le dichiarazioni IRAP

Tabacchi-farmaci	tabacchi-giornali	farmaci-giornali
0,945	0,632	0,746

Per queste considerazioni si sono utilizzati due modelli: entrambi tra le variabili esplicative comprendono i turisti e i pendolari ma nel primo modello come variabile IRAP sono stati presi i dati relativi ai giornali e nel secondo quelli relativi ai farmaci.

In primo modello stima i consumi interni attraverso tre variabili esplicative: pendolari, turisti e giornali. L'adattamento del modello è buono (R^2 corretto = 0,779) i predittori risultano tutti altamente significativi e con coefficienti positivi.

Tabella 6. Il modello 1

Modello	R	R^2	R^2 aggiustato	Std. Error della stima
1	0,886(a)	0,786	0,779	0,09661
	coefficiente	Std. Error	t di Student	Significatività
Intercetta	9,198	0,0280,028	329,175	0,000
Giornali	0,209	0,019	11,099	0,000
Turisti	0,029	0,008	3,487	0,001
Pendolari	0,814	0,141	5,779	0,000

La lettura dei parametri questa volta risulta leggermente più complessa. Il parametro associato ai pendolari va letto alla luce di come è stato costruito l'indice di pendolarità.

In verità i consumi sono stimati per aggregazioni di comuni (tali aggregazioni sono effettuate in base alla zona di competenza delle varie agenzie delle entrate ubicate sul territorio) e non per singoli comuni. Questo perchè in alcuni comuni, di dimensioni molto ridotte è possibile che non esistano farmacie o edicole e dunque mancano le informazioni necessarie per la stima stessa.

Il secondo modello stima i consumi interni attraverso tre variabili esplicative: pendolari, turisti e farmaci. L'adattamento del modello è più basso di quello precedente (R^2 adattato = 0,554) ma anche in questo caso i predittori risultano tutti altamente significativi e con coefficienti positivi

Tabella 7. Il modello 2

Modello	R	R^2	R^2 aggiustato	Std. Error della stima
2	0,753(a)	0,567	0,554	0,13732
	coefficiente	Std. Error	t di Student	Significatività
Intercetta	8,367	0,327	25,557	0,000
Farmaci	0,054	0,011	4,894	0,000
Turisti	1,561	0,176	8,878	0,000
Pendolari	0,272	0,082	3,313	0,001

4.3 Un commento dei risultati

I risultati ottenuti (riportati in appendice) espressi in termini di consumi pro capite presentano differenze piuttosto elevate a seconda del modello utilizzato. Nel caso della Toscana le differenze sono meno spiccate, mentre in Lombardia lo sono in modo tale da ritenere il secondo modello inaccettabile.

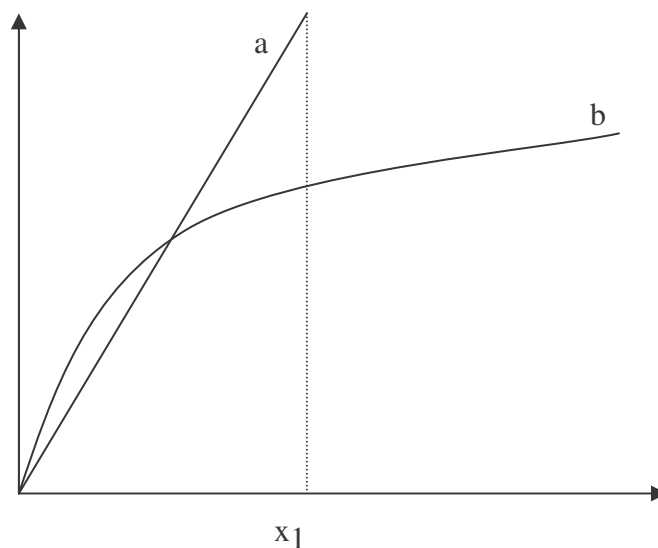
Le differenze possono dipendere da molte circostanze, alcune delle quali perfettamente giustificabili. Tra queste ultime vale la pena di ricordare che il fatto di rapportare i consumi alla popolazione residente rende tale rapporto in generale poco significativo: il numeratore si riferisce infatti ad un comportamento di residenti e non residenti, mentre il denominatore contiene solo i primi. In sintesi i consumi pro capite c (nelle notazioni che seguono i pedici r e t indicano rispettivamente residenti e turisti, il maiuscolo indica che i dati sono espressi in valori assoluti, mentre il minuscolo in valori pro capite) possono esprimersi nel modo seguente

$$c=(C_r +C_t)/R, \text{ ovvero } c=c_r + c_t T/R$$

È evidente che quando il peso dei turisti (T/R) è particolarmente elevato il valore di c è elevato anche se i consumi dei residenti sono bassi: non è un caso che nelle tabelle che seguono in tutte le località turistiche i valori dei consumi pro capite siano molto elevati. Naturalmente il valore dipende anche dalla dimensione del parametro associato al turismo che, in effetti, è particolarmente alto proprio nel modello 2 (l'elasticità è di 1,561 nel modello 2 contro 0,029 del modello 1).

Queste due considerazioni tuttavia mantengono aperta una rilevante incertezza circa il significato di fondo del modello utilizzato: quello cioè di ottenere parametri ad un livello territoriale superiore per poi utilizzarli per la stime al livello territoriale inferiore. Il problema non sta tanto nell'ipotesi che esista un modello comportamento unico, una volta che siano state ben individuate le variabili, quanto nel fatto che il campo di oscillazione di queste ultime consenta di cogliere correttamente il modello di comportamento.

La situazione in cui possiamo spesso trovarci quando passiamo dal livello regionale o provinciale a quello comunale può essere rappresentata efficacemente nel grafico che segue (ma è un problema noto nelle stime econometriche): la stima dei parametri del modello viene effettuata su variabili che si muovono all'interno del campo di oscillazione osservato a livello provinciale; le variabili osservate a livello comunale possono, però, talvolta uscire in modo significativo da tale campo di osservazione accentuando gli errori di stima. Supponiamo ad esempio che la funzione lineare (la retta a) sia una buona approssimazione di una funzione di comportamento diversa (crescente, ma con derivata seconda negativa come la curva b del grafico): nell'ambito del campione osservato l'approssimazione è accettabile, ma se ci si allontana da tale campo ponendoci, ad esempio, nel punto x_1 la differenza comincia ad essere eccessiva e non più accettabile.



Il caso del turismo appartiene a questa tipologia di comportamenti: a livello provinciale il rapporto tra turisti e popolazione varia da un minimo di 616 ad un massimo di 23.317 turisti ogni 1.000 abitanti; quando si passa a livello comunale il valore passa da un minimo di 21 ad un massimo di 109.531 turisti ogni 1.000 abitanti.

A questo problema se ne aggiungono altri legati alla incertezza dei dati utilizzati a livello comunale. Vi sono almeno tre motivi di incertezza:

1. i dati IRAP talvolta riportano valori negativi, che dipendono ovviamente da motivi tecnici, ma che mettono in dubbio la prima ipotesi del modello utilizzato (i dati IRAP sono una indicazione molto precisa delle vendite del prodotto considerato nel luogo considerato);
2. vi possono essere vendite del prodotto preso in esame (giornali, tabacchi e medicinali), anche fuori dai punti di vendita di cui disponiamo di dichiarazioni IRAP;
3. i dati sulle presenze turistiche si riferiscono alla sola tipologia di turismo ufficialmente registrata (mancano in particolare le seconde case) e, a livello comunale, sono stati stimati con una procedura particolare che rende ancora più incerti i valori utilizzati.

I tre problemi sopra evidenziati, assieme a quello più rilevante richiamato prima (il fatto cioè di uscire dal campo di oscillazione delle variabili esplicative) hanno dunque natura molto diversa e non sempre appaiono facilmente risolvibili.

In alcuni casi si tratta di migliorare la qualità dell'informazione (i dati sul turismo appartengono a questa categoria) o di scoprirne gli aspetti più oscuri (perché l'IRAP negativa in alcuni comuni), in altri il problema è piuttosto quella della ricerca di una forma funzionale che limiti i problemi derivanti dal fatto di essere molto esterni al campo di oscillazione delle variabili esplicative. Tutte queste considerazioni ci indicano tuttavia a trovare nel turismo la causa principale della instabilità delle stime, sia perché i dati sono spesso non disponibili, sia perché è proprio su questo fenomeno che le differenze si accentuano enormemente man mano che dalla scala regionale si passa a quella provinciale e poi a quella comunale. Una stima del turismo in seconde case condotta anche a livello comunale che integri quella sulle presenze nelle strutture turistiche ufficiali appare quindi fondamentale

Appendice – Confronto tra i due modelli

I comuni sono stati aggregati secondo la competenza territoriale degli uffici dell’Agenzia delle Entrate eccetto che per i comuni di Milano e Palermo. In particolare si ha la seguente aggregazione territoriale:

UFFICIO DI COMPETENZA	COMUNE
UFFICIO DI MILANO	MILANO
	BOLLATE
	BRESSO
	CESATE
UFFICIO DI MILANO 2	CORMANO
	GARBAGNATE MILANESE
	NOVATE MILANESE
	SENAGO
	COLTURANO
	DRESANO
	MEDIGLIA
	PANTIGLIATE
	PAULLO
	PESCHIERA BORROMEO
UFFICIO DI MILANO 3	SEGRATE
	TRIBIANO
	CARPIANO
	CERRO AL LAMBRO
	MELEGNANO
	SAN COLOMBANO AL LAMBRO
	SAN DONATO MILANESE
UFFICIO DI MILANO 4	SAN GIULIANO MILANESE
	SAN ZENONE AL LAMBRO
	VIZZOLO PREDABISSI
	ASSAGO

UFFICIO DI COMPETENZA

COMUNE

BASIGLIO

BINASCO

BUBBIANO

BUCCINASCO

CALVIGNASCO

CASARILE

CORSICO

LACCHIARELLA

LOCATE DI TRIULZI

NOVIGLIO

OPERA

PIEVE EMANUELE

ROZZANO

TREZZANO SUL NAVIGLIO

VERNATE

ZIBIDO SAN GIACOMO

CESANO BOSCONI

CUSAGO

SETTIMO MILANESE

UFFICIO DI MILANO 5

UFFICIO DI MILANO 6

UFFICIO DI COMPETENZA

COMUNE

UFFICIO DI PALERMO

PALERMO

UFFICIO DI PALERMO 1

BALESTRATE

BORGETTO

CAPACI

CARINI

CINISI

GIARDINELLO

ISOLA DELLE FEMMINE

UFFICIO DI COMPETENZA

COMUNE

UFFICIO DI PALERMO 2

MONTELEPRE
PARTINICO
TERRASINI
TORRETTA
TRAPPETO
ALIA
ALTOFONTE
BISACQUINO
CAMPOFIORITO
CAMPOREALE
CASTRONUOVO DI SICILIA
CHIUSSA SCLAFANI
CONTESSA ENTELLINA
CORLEONE
GIULIANA
LERCARA FRIDDI
PALAZZO ADRIANO
PRIZZI
ROCCAMENA
ROCCAPALUMBA
SAN CIPIRELLO
SAN GIUSEPPE JATO
VICARI
BAUCINA
BELMONTE MEZZAGNO
BOLOGNETTA
CAMPOFELICE DI FITALIA
CEFALA' DIANA
CIMINNA

UFFICIO DI PALERMO 2

UFFICIO DI PALERMO 3

UFFICIO DI COMPETENZA

COMUNE

GODRANO

MARINEO

MEZZOJUSO

MISILMERI

MONREALE

PIANA DEGLI ALBANESI

SANTA CRISTINA GELA

USTICA

VENTIMIGLIA DI SICILIA

VILLABATE

VILLAFRATI

Bacini di utenza degli uffici	Modello 1	Modello 2
Lombardia		
Abbiategrosso	10.952	818
Bergamo 1	16.452	3.828
Bergamo 2	13.111	860
Breno	14.035	37.639
Brescia	21.000	9.613
Brescia 1	13.395	13.699
Brescia 2	11.650	3.143
Busto Arsizio	14.467	162
Cantu'	12.463	135
Casalmaggiore	12.820	3.436
Castiglione delle Stiviere	14.282	1.958
Chiari	12.456	33
Clusone	13.503	11.421
Codogno	12.645	446
Como	15.581	3.737
Corteolona	7.778	201
Crema	10.464	479
Cremona	15.380	2.002
Desio 1	10.623	428
Desio 2	10.083	646
Erba	13.543	1.252
Gallarate	16.391	1.553
Gardone Val Trompia	12.173	302.964
Gavirate	13.742	8.373
Gorgonzola	12.624	808
Lecco	15.827	6.626
Legnano	12.289	1.031
Lodi	11.706	2.764
Lonato	16.373	178.190
Luino	12.590	10.022
Magenta	11.600	840
Mantova	14.940	2.657
Menaggio	9.845	156.642
Merate	12.142	73
Milano	23.022	19.377
Milano 2	9.026	126

Bacini di utenza degli uffici	Modello 1	Modello 2
Lombardia (continua)		
Milano 3	15.579	7.448
Milano 4	13.927	5.421
Milano 5	12.975	14.969
Milano 6	13.535	2.508
Montichiari	14.379	604
Monza 1	11.547	1.618
Monza 2	13.970	808
Morbegno	12.570	18.042
Mortara	13.073	502
Pavia	14.700	1.471
Ponte San Pietro	11.985	240
Rho	14.873	919
Romano di Lombardia	13.450	697
Salo'	13.528	53.598
Saronno	14.806	938
Sondrio	16.326	12.241
Soresina	13.184	121
Stradella	11.854	313
Suzzara	14.018	937
Tirano	15.604	365.005
Treviglio	13.555	266
Varese	14.095	2.249
Verolanuova	13.544	190.626
Vigevano	14.861	342
Vimercate	16.301	2.740
Voghera	14.856	5.656
Zogno	12.962	29.555
Toscana		
Arezzo	13.790	14.491
Aulla	10.035	8.864
Borgo San Lorenzo	12.307	9.547
Carrara	13.330	12.357
Castelnuovo Garfagnana	10.409	9.669
Empoli	12.254	15.046
Firenze 1	18.962	20.688
Firenze 2	13.038	10.084
Firenze 3	18.094	27.178
Grosseto	14.401	13.643
Bacini di utenza degli uffici	Modello 1	Modello 2

Toscana (continua)		
Livorno	13.229	12.315
Lucca	15.093	15.677
Massa	14.818	14.153
Massa Marittima	13.468	12.8
Montepulciano	14.054	14.073
Montevarchi	12.112	10.534
Orbetello	13.685	13.611
Pescia	13.578	1.338
Pietrasanta	17.587	16.699
Piombino	16.306	16.854
Pisa	14.425	12.773
Pistoia	12.254	1.225
Poggibonsi	13.949	12.632
Pontedera	12.793	13.371
Portoferraio	17.470	16.888
Prato	13.614	12.106
San Miniato	12.842	13.786
San Sepolcro	10.459	12.098
Siena	16.578	17.292
Viareggio	14.972	13.541
Toscana	14.486	14.486
Sicilia		
Acireale	7.072	7.215
Agrigento	9.825	4.660
Augusta	11.640	10.918
Bagheria	7.795	2.459
Barcellona Pozzo di Gotto	10.611	3.239
Caltagirone	8.439	149
Caltanissetta	9.354	471
Canicattì	7.225	3
Castelvetrano	9.123	1.599
Catania	11.945	1.238
Cefalù	9.888	151.755
Enna	10.650	739
Gela	6.436	195
Giarre	9.154	5.895
Lentini	7.797	30
Licata	8.591	9
Marsala	9.275	695
Bacini di utenza degli uffici	Modello 1	Modello 2

Sicilia (continua)

Messina	11.390	1.442
Milazzo	11.974	16.164
Ristretta	7.719	2.368
Modica	11.040	1.577
Nicosia	9.596	160
Noto	7.067	6.059
Palermo	11.736	3.027
Palermo1	7.944	11.553
Palermo2	4.449	58
Palermo3	4.307	964
Patti	12.845	39.970
Ragusa	13.443	20.319
Sant'agata di Militello		1.165
Sciacca	9.568	17.329
Siracusa	12.593	4.761
Taormina	13.702	579.008
Termini Imprese	6.828	5.724
Trapani	10.677	8.242
Vittoria	12.235	137

Riferimenti bibliografici

- Brackstone G.J. (1987), *Small Area Data: Policy Issues and Technical Challenges*, in Plateck R., Rao J.N.K., Särndal C.E. and Singh M.P. (Eds), *Small Area Statistics*, Wiley, New York
- Casini Benvenuti S., Viviani A. (2004) “La stima dei consumi interni dei comuni della Toscana”, documento interno dell’Ufficio Studi dell’Agenzia delle Entrate.
- Chiandotto B. (1996), *L’informazione statistica a livello territoriale: significatività, problemi e limiti*. Terza Conferenza Nazionale di Statistica, 24-26 Novembre 1996, Roma
- Cicchitelli G., Herzel A.Montanari G.E. (1992), *Il campionamento statistico*, Il Mulino, Bologna
- Ghosh M., Natarajan K., Stroud T.W.F., Carlin Bradley P. (1998) “Generalized *Linear Models for Small-Area Estimation*” *Journal of the American Statistical Association*, Vol. 93
- Ghosh M., Rao J. N. K., (1994) “*Small Area Estimation: An Appraisal*” *Statistical Science*, Vol. 9, No. 1, pp. 55-76
- Giusti A., Pratesi M. (1996), “*Nuove tecnologie nel processo di produzione dei dati statistici*”, Atti della XXXVIII Riunione Scientifica della Società Italiana di Statistica, 1996, Rimini
- Heeringa, S.G. (1982) “*Statistical Models for Small Area Estimation*,” in *Readings in Population Research Methodology*, 5, 126-132.
- Isaki Cary T., (1990) “*Small-Area Estimation of Economic Statistics*” *Journal of Business & Economic Statistics*, Vol. 8, No. 4, pp. 435-441
- ISTAT (2007) “I consumi delle famiglie – anno 2005”, Collana *Famiglie e Società*, Annuario.
- Petrucci A., Pratesi M., Salvati N. (2003), *Stimatore combinato e correlazione spaziale nella stima per piccole aree*, Dipartimento di Statistica e Matematica Applicata all’Economia, report n. 240, Pisa

Pfeffermann D. (2002), *Small Area Estimation-New Developments and Directions*, International Statistical Review, pg. 125-143

Pratesi, M., Salvati, N. (2005) “*Small Area Estimation: the EBLUP estimator with autoregressive random area effects*”, Working Paper 261, Dipartimento di Statistica e Matematica Applicata all’Economia, Università di Pisa.

Rao J.N.K. (2003), *Small Area Estimation*, Wiley, London

Salvati, N. (2004) *La correlazione spaziale nella stima per piccole aree: metodi proposti e casi di studio*, Tesi di Dottorato in Statistica Applicata, Dipartimento di Statistica “G. Parenti”, Università degli Studi di Firenze.