Fabio Pammolli[1], Francesco Porcelli[2],
Francesco Vidoli[3], Monica Auteri[3], Guido Borà[4]

La spesa sanitaria delle Regioni in Italia

Saniregio2016

first draft

1. *Politecnico di Milano, CERM*
2. *University of Exeter (UK), CERM*
3. *University of Roma Tre, CERM*
4. *University of Siena, CERM*

Cermlab, 2016

INDICE

Elenco delle figure IV

Elenco delle tabelle V

Elenco delle abbreviazioni VI

1 INTRODUZIONE 1

2 OGGETTO DELLA VALUTAZIONE 12
2.1 Sistemi di controllo della spesa e valutazione delle politiche pubbliche 12
2.2 Il modello italiano di assistenza sanitaria: una panoramica 14
2.3 Questioni di metodo riguardanti la definizione e la misurazione dei fabbisogni sanitari 17
2.4 Domanda e offerta: una critica delle proxy esistenti 21
2.5 Localizzazione della domanda e dell'offerta 22
2.6 Non solo domanda ed offerta: efficacia ed efficienza nei sistemi sanitari nazionali 24

3 SANIREGIO2016 27
3.1 SaniRegio1 e 2: cosa abbiamo imparato 27
3.1.1 SaniRegio1 27
3.1.2 SaniRegio2 28
3.1.3 SaniRegio2015 30
3.2 Obiettivi e struttura di SaniRegio2016 31

4 L'IMPIANTO METODOLOGICO 34
4.1 Stima del livello delle prestazioni, del livello degli input e dell'efficienza tecnica 35
4.1.1 Gli indicatori compositi del livello delle prestazioni offerte e degli input impiegati 35
4.1.2 L'efficienza tecnica 37
4.2 La funzione di domanda e la funzione di spesa 37
4.2.1 La funzione di domanda 38
4.2.2 La funzione di spesa 40

5 L'EFFICIENZA TECNICA PRODUTTIVA PER LE AMMINISTRAZIONI REGIONALI 42
5.1 Il livello delle prestazioni sanitarie regionali 42

5.1.1 Costruzione dell'indicatore composito di output 45
5.2 Analisi degli input del servizio 45
5.2.1 Costruzione dell'indicatore composito di input 53
5.3 Stima dell'efficienza tecnica produttiva 60
5.4 Un esercizio di simulazione per gli anni 2013-2015 62

6 VALORIZZAZIONE DELLA SPESA SANITARIA STANDARD E ANALISI DELLE CAUSE DI SCOSTAMENTO DALLA SPESA STORICA 65
6.1 La stima della funzione di domanda e dell'output-gap 65
6.2 Stima della funzione di spesa 69
6.3 Applicazione dei modelli per la valutazione della spesa standard 73
6.4 Riassunto dei risultati e confronti con la spesa storica 79

7 DISAGGREGAZIONE SPAZIALE DEI LIVELLI DI DOMANDA ED OFFERTA SANITARIA A LIVELLO COMUNALE 84
7.1 Domanda di ricerca 84
7.2 Metodi di disaggregazione spaziale di tipo Chow-lin 86
7.3 Una prima applicazione al caso italiano 89
7.3.1 Indicatore composito della domanda potenziale a livello provinciale . . . 89
7.3.2 Identificazione e stima del modello a livello provinciale 91
7.3.3 Disaggregazione spaziale a livello comunale: un approccio di tipo Chow-Lin 94
7.3.4 Domanda ed offerta a livello comunale tra prossimità e specializzazione 95
7.4 Osservazioni conclusive 100

8 CONCLUSIONI 103

9 APPENDICE 106

ELENCO DELLE FIGURE

Figura 1 Spesa sanitaria anno 2013 e differenziale previsto nel 2060 per paese, fonte: Joint Report on Health Care and Long-term Care systems & Fiscal Sustainability, 2016 . . . 2
Figura 2 Spesa sanitaria corrente. Fonte: Istat (Health for All), dato 2015: Legge 190/2014, dato 2016: Legge 208/2015, dati 2017-2019 previsti dalla Legge di bilancio 2017, inflazione 2017-2019 pari a 1%. . . . 4
Figura 3 Analisi dell'inefficienza tecnica. Fonte: analisi proprie su dati Istat (Health for All) . . . 8
Figura 4 Dinamica del fattore di output 1 (dimensionale) per Regione ed anno, anni 1998-2012 . . . 46
Figura 5 Dinamica del fattore di output 2 (qualità) per Regione ed anno, anni 1998-2012 . . . 47
Figura 6 Dinamica del fattore di output 3 (lungodegenza e riabilitazione) per Regione ed anno, anni 1998-2012 . . . 48
Figura 7 Distribuzione dell'indicatore composito robusto di output per Regione - anno 2012 . . . 49
Figura 8 Dinamica dell'indicatore composito robusto di output per Regione e per anno, anni 1998-2012 . . . 50
Figura 9 Distribuzione dell'indicatore composito robusto di output per Regione al variare del metodo (BoD e RBoD) - anno 2012 . . . 51
Figura 10 Dinamica del fattore di input 1 (infermieri e tecnici) per Regione ed anno, anni 1998-2012 . . . 54
Figura 11 Dinamica del fattore di input 2 (beni strumentali) per Regione ed anno, anni 1998-2012 . . . 55
Figura 12 Dinamica del fattore di input 3 (medici) per Regione ed anno, anni 1998-2012 . . . 56
Figura 13 Dinamica del fattore di input 4 (posti letto) per Regione ed anno, anni 1998-2012 . . . 57
Figura 14 Distribuzione degli indicatori compositi di input Lavoro e Capitale per Regione, anno 2012 . . . 58
Figura 15 Dinamica media degli indicatori compositi di input Lavoro e Capitale per anno . . . 59
Figura 16 Dinamica dell'efficienza tecnica produttiva per Regione ed anno . . . 61
Figura 17 Distribuzione dell'inefficienza globale in percentuale della spesa storica. . . . 74

Figura 18 Inefficienza tecnica e di prezzo per Regione (% della spesa storica) . . . 75
Figura 19 Distribuzione dell'output-gap globale in percentuale della spesa storica. 76
Figura 20 Output gap proprio e legato alla mobilità per Regione (% della spesa storica) 77
Figura 21 Output gap proprio e legato alla mobilità per Regione (mln di €) 78
Figura 22 Differenze tra spesa standard e storica per Regione (% della spesa storica) 80
Figura 23 Indicatore composito di domanda a livello provinciale: analisi di tipo ACP 92
Figura 24 Indicatore composito RBoD di domanda a livello provinciale 93
Figura 25 Stima dell'indicatore composito RBoD di domanda a livello comunale: metodo spaziale di tipo Chow-Lin 96
Figura 26 Stima dell'indicatore composito RBoD di domanda a livello comunale: Regioni Nord-Occidentali 97
Figura 27 Offerta e domanda stimata a livello comunale 98
Figura 28 Offerta e domanda stimata a livello comunale: Regioni Nord-Occidentali 99
Figura 29 Domanda stimata e distanza minima dalla struttura sanitaria (in metri) 100
Figura 30 FDH, CRS-DEA e VRS-DEA 114
Figura 31 Free disposal hull per la generica unità A 114
Figura 32 Distribuzione BoD 117
Figura 33 Distribuzione BoD-PCV 118
Figura 34 Distribuzione BoD-PCV in presenza di dati anomali 119
Figura 35 Presenza di outliers in un problema di stima di frontiera 120
Figura 36 Supporto della generica unità C 120
Figura 37 Criteri di selezione del sottoinsieme in un modello di tipo Order-m . . 121
Figura 38 Distribuzione RBoD in presenza di outlier 122
Figura 39 Distribuzione RBoD-PCV in presenza di outlier 122

ELENCO DELLE TABELLE

Tabella 1 Spese correnti per l'assistenza sanitaria per sistemi di finanziamento, Italia 2014, Fonte: Sistema dei conti della salute (SHA - System health accounts) . . . 10
Tabella 2 Spese correnti per l'assistenza sanitaria per sistemi di finanziamento e funzione, Italia 2014, Fonte: Sistema dei conti della salute (SHA - System health accounts). . . . 11
Tabella 3 Analisi in componenti principali - fattori di output (pro capite), anni 1998-2012 . . . 44
Tabella 4 Analisi in componenti principali - fattori di input (pro capite), anni 1998-2012 . . . 52
Tabella 5 Stima dell'andamento dell'efficienza tecnica produttiva per gli anni 2013-2015 . . . 64
Tabella 6 Funzione di domanda per modello di stima, variabile dipendente = indicatore composito di output, anno 2014. . . . 67
Tabella 7 Funzione di spesa per modello di stima, variabile dipendente = spesa sanitaria pubblica corrente per abitante, anno 2014 . . . 71
Tabella 8 Spesa storica e standard per Regione, valori monetari (mld di €), anno 2014 . . . 82
Tabella 9 Spesa storica e standard per Regione, valori percentuali (% della spesa storica), anno 2014. . . . 83
Tabella 10 Indicatore composito di domanda a livello provinciale: analisi di tipo ACP . . . 90
Tabella 11 Indicatore composito di domanda a livello provinciale: metodi a confronto . . . 91
Tabella 12 Indicatore BoD robusto e caratteristiche della domanda a livello provinciale: modelli a confronto . . . 94
Tabella 13 Coefficiente di codispersione tra indicatore di rischio e distanza al variare della classe di distanza . . . 101
Tabella 14 Variabili elementari utilizzate nell'analisi di efficienza . . . 107
Tabella 15 Punteggi fattoriali e indicatore composito RBoD per Provincia (1/2) . . 109
Tabella 16 Punteggi fattoriali e indicatore composito RBoD per Provincia (2/2) . . 110
Tabella 17 Tecniche di stima dell'efficienza per tecnologia di produzione . . . 112

ELENCO DELLE ABBREVIAZIONI

2SLS	Two-Stage least squares
ACP	Analisi in Componenti Principali
ASL	Aziende Sanitarie Locali
BoD	Benefit of Doubt
BoD-PCV	Benefit of Doubt Penalized by Coefficient of Variation
CI	Composite Indicator
CRS-DEA	Data Envelopment Analysis with Constant Returns to Scale
DEA	Data Envelopment Analysis
DFA	Deterministic Frontier Analysis
DMU	Decision Making Unit
FDH	Free Disposal Hull
GAM	Generalized Additive Model
LEA	Livelli essenziali di assistenza
MPCV	Method of Penalties for Coefficient of Variation
Order-m	Robust Nonparametric Efficiency Analysis - Order-m
RBoD	Robust Benefit of Doubt
RBoD-PCV	Robust Benefit of Doubt Penalized by Coefficient of Variation
SFA	Stochastic Frontier Analysis
SSN	Sistema Sanitario Nazionale
StoNED	Stochastic semi-Nonparametric Envelopment of Data
StoNEZD	Stochastic semi-Nonparametric Envelopment of Z variables Data
VRS-DEA	Data Envelopment Analysis with Variable Returns to Scale

1 INTRODUZIONE

> "*The Council considers that achieving the twin aim of ensuring fiscal sustainability and access to good quality health care services for all, by improving the efficiency and effectiveness of health and long-term care systems, is therefore particularly important*", Council of the European Union[a], 8 Novembre 2016.

a http://data.consilium.europa.eu/doc/document/ST-14182-2016-INIT/en/pdf.

La spesa sanitaria assorbe una significativa quota di risorse nei paesi dell'Unione Europea; le stime più recenti indicano un notevole incremento di tale aggregato (pari in Italia al 6,9% del PIL, anno 2015[1]) nel medio-lungo termine, un incremento (comprensivo anche degli aspetti legati all'assistenza) che è stimato essere tra il 2015 ed il 2060 maggiore di quello pensionistico.
Più specificatamente, il "*Joint Report on Health Care and Long-term Care systems & Fiscal Sustainability*" (JRHC, EU Directorate-General for Economic and Financial Affairs and Economic Policy Committee (Ageing Working Group), 2016) sottolinea le ampie differenze nel livello di base e negli andamenti previsti (AWG Reference Scenario) della spesa sanitaria (si veda figura 1).

La sfida fiscale derivante dall'aumento della spesa sanitaria nell'UE e' significativa e differenziata nei diversi paesi dato che sono differenti i tassi di invecchiamento della popolazione, di introduzione di nuove tecnologie e le condizioni di partenza dovute ad esempio agli alti livelli di debito da riassorbire.
A politiche invariate, la spesa sanitaria è destinata a crescere (in rapporto al PIL) in misura significativa, trainata da fattori demografici e non demografici richiedendo un monitoraggio attento e costante dei singoli sistemi sanitari.
Il rapporto JRHC 2016 individua alcuni criteri di analisi e di valutazione dei diversi sistemi di governance in ambito UE con l'indicazione di possibili aree di intervento/potenziamento sia sul versante sanitario che assistenziale al fine di migliorare il rapporto costo-efficacia del sistema, ridurre gli sprechi ed i trattamenti inappropriati e combattere le frodi.
Sono essenzialmente due le raccomandazioni di policy che vengono suggerite agli Stati membri:

1 Ragioneria Generale dello Stato (2016), Il monitoraggio della spesa sanitaria, Rapporto n.3, http://www.rgs.mef.gov.it/_Documenti/VERSIONE-I/Attivit--i/Spesa-soci/Attivit-monitoraggio-RGS/2016/IMDSS-RS02_11_08_2016.pdf.

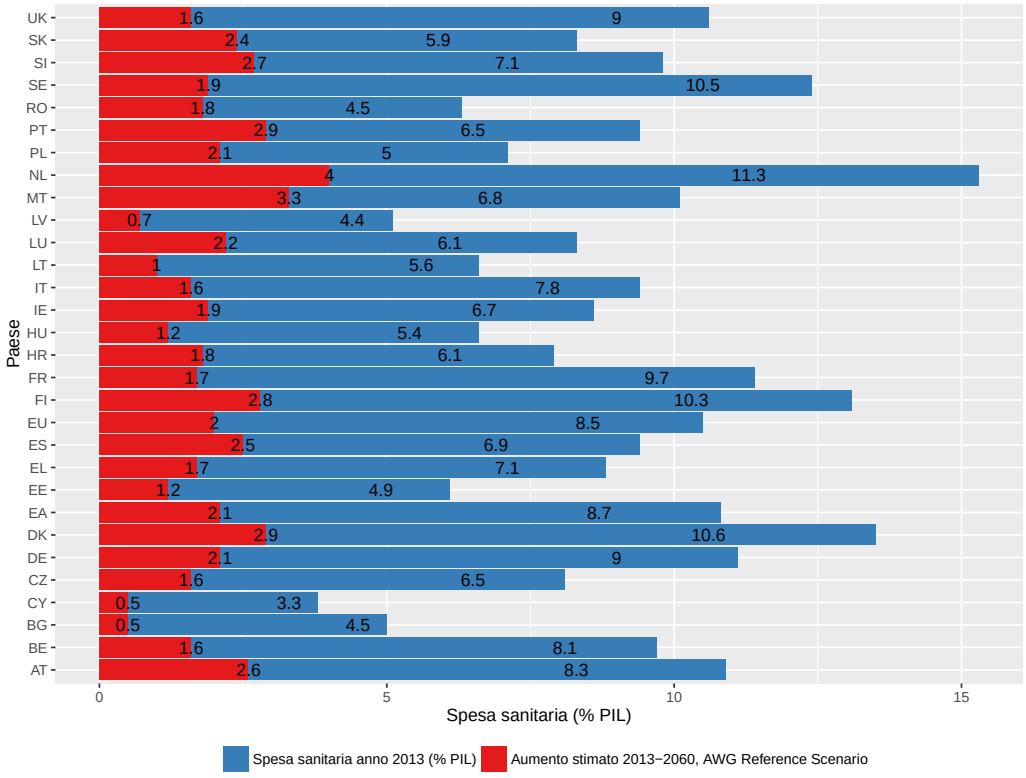


Figura 1: Spesa sanitaria anno 2013 e differenziale previsto nel 2060 per paese, fonte: Joint Report on Health Care and Long-term Care systems & Fiscal Sustainability, 2016

A *Migliorare la governance dei sistemi sanitari*: attraverso il (A1) rafforzamento della cooperazione tra autorità incaricate della politica fiscale e e di quella sanitaria (A2) sviluppando e potenziando i sistemi informativi e di gestione dei dati che favoriscano il monitoraggio e la governance (A3) valutando formalmente e sistematicamente, ex-ante ed ex-post, le riforme sanitarie e (A4) utilizzando strumenti di pianificazione per una gestione efficace e strategica delle risorse umane;

B *Migliorare la sostenibilità finanziaria e il rapporto costo-efficacia della spesa*: (B1) riducendo la prevalenza delle cure ospedaliere e migliorando l'efficienza degli ospedali, (B2) migliorando la performance dei sistemi di cura primaria rafforzandone il ruolo di "gatekeeping" e (B3) promuovendo un uso dei farmaci basato sul rapporto costi-efficacia.

Il rapporto di ricerca SaniRegio2016 vuole rispondere a queste raccomandazioni suggerendo metodi, strumenti e stime puntuali sia riguardo al miglioramento dell'efficienza della spesa sanitaria al fine di un più puntuale contenimento dei costi della spesa sanitaria (capitoli 5 e 6) sia sul versante di un miglior utilizzo a fini programmatori e di valutazione del sistema informativo sanitario a livello nazionale per facilitare un monitoraggio approfondito a livello sub-regionale (capitolo 7).

L'evoluzione della spesa sanitaria pubblica corrente in Italia dalla fine degli anni Novanta è segnata, come mostrato in figura 2, da due tendenze principali.

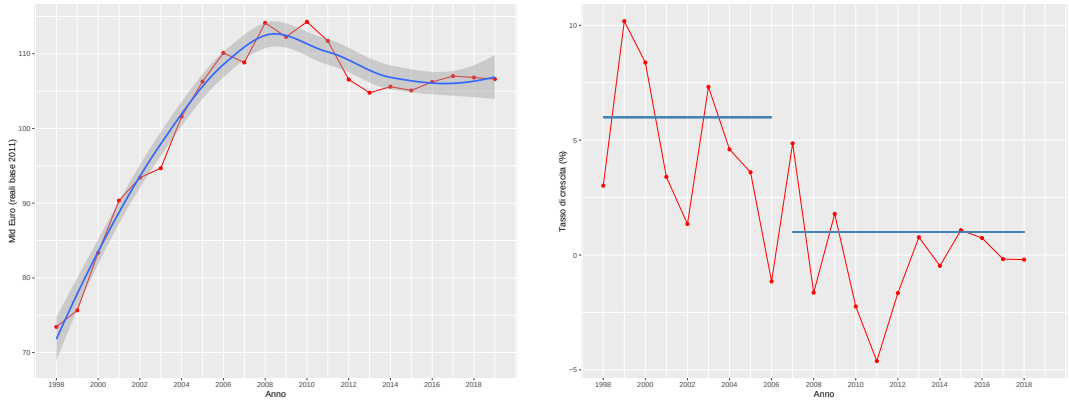
Sino al 2006 si è realizzato un forte trend crescente che ha portato la spesa al 6,6% del PIL con un incremento significativo rispetto al 5,0% del 1998. A partire dal 2006, questo trend sembra essersi arrestato, se è vero che la spesa è rimasta sostanzialmente costante in valori monetari. L'incidenza rispetto al PIL ha continuato su un trend crescente a causa della contrazione dovuta alla crisi finanziaria attestandosi al 7,8% nel 2013, ultimo anno disponibile nel database *Health For All* dell'ISTAT utilizzato per questa analisi. I valori di spesa corrente più recenti riferiti al 2014 e 2015 diffusi dalla Ragioneria Generale dello Stato (RGS 2016)[2], ma non ancora inseriti dall'ISTAT nella banca dati *Health for All*, mostrano come che nel 2014 e nel 2015 la spesa si è mantenuta costante in percentuale del PIL rimanendo ferma al 6,9%, riducendosi però in termini assoluti.

Un quadro, quello richiamato, che trova un riscontro nell'andamento dei tassi di crescita, rimasti positivi sino al 2006 (anche se tendenzialmente decrescenti), e divenuti prevalentemente negativi negli anni successivi.

Nei confronti internazionali l'Italia, con una spesa sanitaria corrente pari al 6,9% del PIL, si colloca di poco sopra la media OCSE pari al 6,4% del PIL, posizionandosi dietro paesi come la Francia (8,7%), la Germania (8,4%) e gli Stati Uniti (8,3%).

Il contenimento della spesa ottenuto dopo il 2006 è il risultato dell'implementazione, con i Patti per la Salute 2007-2009 (Intesa Stato-Regioni del 5 ottobre 2006), 2010-2012 (Intesa

2 Ragioneria Generale dello Stato (2016), Il monitoraggio della spesa sanitaria, Rapporto n.3, http://www.rgs.mef.gov.it/_Documenti/VERSIONE-I/Attivit--i/Spesa-soci/Attivit-monitoraggio-RGS/2016/IMDSS-RS02_11_08_2016.pdf.



(a) (b)

Figura 2: Spesa sanitaria corrente. Fonte: Istat (Health for All), dato 2015: Legge 190/2014, dato 2016: Legge 208/2015, dati 2017-2019 previsti dalla Legge di bilancio 2017, inflazione 2017-2019 pari a 1%.

Stato-Regioni del 3 dicembre 2009) e 2014-2016 (Intesa Stato-Regioni del 10 luglio 2014), di un nuovo assetto di regole incentrato sul passaggio dal sistema fondato sulla "sindrome da vincolo di bilancio soffice" e sull'aspettativa da parte delle Regioni di ripiano dei disavanzi da parte dello Stato, ad un sistema basato sul "principio della forte responsabilizzazione" attraverso l'introduzione dei cosiddetti "Piani di rientro" per le regioni in disavanzo, confermati anche per il triennio 2013-2015.
Tali interventi di efficientamento e razionalizzazione del sistema sono stati particolarmente apprezzabili avendo riguardato una pluralità di aspetti e di soluzioni sia medico che tecnologiche; hanno, in particolare, avuto un forte impatto contenitivo le politiche di contenimento della spesa per il personale (turn-over, e dinamiche retributive del Pubblico Impiego), il miglioramento dell'appropriatezza delle prestazioni sanitarie (attraverso in particolare una riduzione delle prestazioni ospedaliere con ricovero in favore del day hospital o di interventi ambulatoriali), il trasferimento alla sfera socio-assistenziale (ad esempio l'assistenza agli anziani non autosufficienti) di componenti di spesa impropriamente coperte da funzioni acute del sistema sanitario ed il potenziamento dei sistemi centralizzati di acquisto e dei sistemi di controllo (tramite ad esempio il sistema della tessera sanitaria).

Significativi differenziali territoriali dei costi pro capite, però, pur depurati degli effetti indotti dalla diversa struttura demografica, permangono ancora tra le diverse Regioni evidenziando ampi margini di un ulteriore efficientamento del sistema.
L'andamento della spesa lungo il territorio nazionale mostra trend simili nelle diverse macro-

aree geografiche[3]: regioni a statuto speciale del Nord Ovest (NRS)[4], regioni del Nord Est (NR)[5], regioni del Centro (CR)[6], regioni del Sud (SD)[7], regioni a statuto speciale del Sud (SDS)[8].
Se si guarda ai valori monetari, la spesa per abitante si discosta pochissimo lungo il territorio nazionale, passando da un valore medio di circa 1200 euro del 1998 a circa 1700 euro del 2016 (valori reali su base 2011). Fanno eccezione le Regioni a statuto speciale del Nord, che presentano una spesa per abitante notevolmente più alta rispetto alle altre Regioni, passando da circa 1400 euro del 1998 a valori che superano i 1950 euro del 2016 (valori reali su base 2011).
Se si considera la spesa in percentuale del PIL, invece, il Paese risulta diviso in due gradi gruppi, tipici del dualismo economico italiano: le regioni SD e SDS presentano una spesa passata dal 7% del PIL nel 1998 a quasi il 10% del PIL nel 2018, mentre le altre regioni (NRS, NR, CR) passano, in media, dal 5% al 7% del PIL.
Un dato, quello che confronta la spesa per abitante con la spesa in percentuale dei PIL, che conferma il ruolo cruciale della perequazione delle risorse sanitarie pubbliche nell'assicurare un livello di spesa pro capite relativamente omogeneo lungo il territorio nazionale.

L'analisi della composizione della spesa corrente tra diretta e in convenzione mostra tendenze molto diverse, sia in termini di trend, sia come distribuzione lungo il territorio nazionale.

La spesa corrente diretta, che rappresenta circa il 70% del totale (anno 2016), mostra le stesse caratteristiche riscontrate in precedenza per la spesa pubblica corrente totale.

La spesa in convenzione per prestazioni sociali mostra un trend crescente generalizzato sino al 2005; dal 2006 in poi invece si registra un andamento leggermente decrescente su tutto il territorio nazionale. La distribuzione di questa componente di spesa nelle macro-aree geografiche è molto diversa rispetto alla distribuzione della spesa diretta.

Tra il 1998 e il 2016 la spesa in convenzione per assistenza farmaceutica mostra un divario di molto allargato tra le regioni SD e il resto del Paese. Rispetto al 1998 l'analisi della spesa in convenzione per assistenza farmaceutica non mostra un chiaro trend comune alle diverse aree territoriali del Paese: le regioni NRS, NR e CR presentano un trend decrescente, mentre le regioni SD e SDS un andamento prevalentemente crescente.

Da ultimo, l'analisi della spesa sanitaria delle famiglie (corrispondente alla spesa privata nel suo complesso) mostra, a fronte di un andamento piuttosto costante nel tempo, interrotto solo da un leggero incremento negli anni della crisi finanziaria, valori molto diversi lungo il territorio nazionale.

Il Patto per la Salute 2014-2016 ha rilanciato il tema dei costi standard che erano stati introdotti in sanità dal d.lgs. 68/2011 (calcolo del costo standard per il riparto 2013 e, in seguito

3 L'analisi si basa su dati di fonte ISTAT tratti dal database *Health for All*.
4 Valle d'Aosta, Trentino Alto Adige, Friuli Venezia Giulia (regioni NRS).
5 Piemonte, Lombardia, Veneto, Liguria, Emilia Romagna (regioni NR).
6 Toscana, Umbria, Marche, Lazio (regioni CR).
7 Abruzzo, Molise, Campania, Puglia, Basilicata, Calabria (regioni SD).
8 Sicilia, Sardegna (regioni SDS).

anche per gli anni successivi, sulla spesa media delle tre Regioni più virtuose).
L'art. 1 comma 601 della l. 190/2014 (Legge di stabilità 2015) ha modificato l'articolo 27 comma 7 del d.lgs. 68/2011 secondo un orientamento diverso da quello attuale [9] disponendo che, per l'anno in corso, i pesi per il riparto siano rivisti con Decreto Ministeriale previa intesa in sede di Conferenza Stato – Regioni sulla base dei criteri indicati dall'art. 1 comma 34 della l. 662/1996, tenendo conto, nella ripartizione del costo e del fabbisogno sanitario standard regionale, del percorso di miglioramento per il raggiungimento degli standard di qualità.[10] La Corte dei conti,[11] a tale riguardo, ha più volte osservato che l'attuale criterio di quantificazione dei fabbisogni sanitari regionali è tuttora prevalentemente basato sulla popolazione pesata di ciascuna regione e propone come obiettivo di "*pervenire a un set d'indicatori tali da valutare i livelli di efficienza e di appropriatezza raggiunti in ciascuna regione, con riferimento ad un aggregato di prestazioni rese all'interno di ciascuno dei tre macrolivelli di assistenza*".[12]

Il nuovo rapporto SaniRegio fornisce, in linea con le edizioni precedenti, il calcolo della spesa sanitaria standard delle singole Regioni attraverso la stima di una funzione di spesa in cui la distribuzione della popolazione per fasce di età rappresenta il driver principale dei fabbisogni standard determinati prendendo a riferimento la spesa standard di una o più regioni benchmark.

Diversamente dalla prime due edizioni e similmente a SaniRegio2015, il nuovo SaniRegio isola due componenti importanti della spesa storica: la quota di spesa relativa all'inefficienza tecnica, ovvero la percentuale di spesa riconducibile agli input in eccesso rispetto a quelli compatibili con una produzione efficiente degli attuali livelli di servizio; la quota di spesa attribuibile ai livelli - quantitativi e qualitativi - di output realizzati in misura superiore o inferiore rispetto ad un valore di output standard definito attraverso la stima di una funzione di domanda. Queste due componenti di spesa, se correttamente individuate, consentono di affiancare al calcolo dei coefficienti di riparto delle risorse del comparto sanitario, una misura del livello ottimale dei fondi pubblici oggetto del riparto.

Con questo obiettivo, il nuovo SaniRegio impiega una varietà di tecniche di stima e aggregazione non parametriche, per ottenere tre misure sintetiche che approssimano la funzione di produzione dei servizi sanitari: a) un indicatore composito di output, costituito da una componente dimensionale misurata in relazione alle giornate di degenza e da una componente qualitativa legata principalmente ai flussi della mobilità sanitaria; b) un indicatore composito dell'input lavoro, che comprende sia la componente legata agli infermieri, ai tecnici e

9 La ripartizione delle risorse del SSN è sempre avvenuta su proposta delle Regioni. La Conferenza delle Regioni e delle Province autonome, in sede di osservazioni alla legge di stabilità, ha chiesto lo stralcio della proposta preferendo l'auto-revisione.

10 Ossia popolazione residente, frequenza dei consumi sanitari per età e per sesso, tassi di mortalità della popolazione, indicatori relativi a particolari situazioni territoriali ritenuti utili al fine di definire i bisogni sanitari delle regioni e indicatori epidemiologici territoriali.

11 Si veda da ultimo: Corte dei Conti, Rapporto 2015 sul coordinamento della finanza pubblica, Roma p. 197.

12 Assistenza ospedaliera, assistenza distrettuale e assistenza collettiva in ambiente di vita e di lavoro che pesano sul finanziamento della spesa sanitaria rispettivamente il 44%, il 51% e il 5%.

al personale riabilitativo in generale, sia la componente più specialistica legata al personale medico; c) un indicatore composito dell'input capitale che misura il livello delle dotazioni strumentali considerando la distribuzione dei beni strumentali (quali TAC, tavoli operatori, tavoli radio ecc.) e la distribuzione dei posti letto. Successivamente, utilizzando queste misure di output e input, il grado di inefficienza tecnica di ogni regione per ogni anno tra il 1998 e il 2012 è stato calcolato in modo non parametrico utilizzato la Data Envelopment Analysis. In prima battuta, sulla base dei dati disponibili nel dataset *Health for All* dell'ISTAT l'analisi si è fermata al 2012, nella versione finale del lavoro s'intende estendere le stime al 2014 e dove è possibile anche al 2015 reperendo i dati mancanti da altre fonti.

Come evidenziato in figura 3, la metodologia, applicata in prima istanza per il periodo 1998-2012, evidenzia una tendenza molto chiara. Tra il 1998 e il 2012, l'intero sistema presenta livelli d'inefficienza mediamente vicini a 0,5 su di una scala compresa tra zero (massima efficienza) e uno (massima inefficienza). Sino al 2002 si osserva, in media, un incremento dell'inefficienza tecnica che spiega parte del forte incremento della spesa registrato in quegli anni.

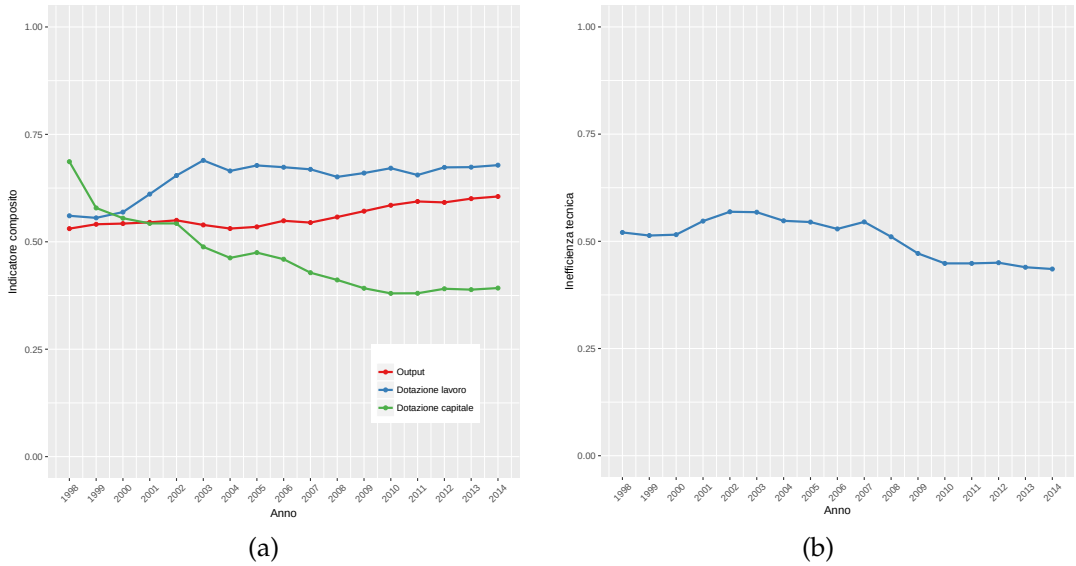
Dal 2002 in poi, invece, si registra una costante riduzione dell'inefficienza, legata a tre fattori: una trend costante nel livello dell'input lavoro per abitante; un trend decrescente nel livello delle dotazioni strumentali per abitante; un incremento dell'output per abitante, soprattutto a partire dal 2008, anno dal quale si consolidano sia i maggiori recuperi di efficienza sia le prime riduzioni nel livello della spesa corrente in termini reali. Conseguentemente, gli incrementi di spesa osservati tra il 2002 e il 2006 possono essere attribuiti in modo prevalente al livello dei prezzi (salari, prezzi d'acquisto per beni e servizi).

Un'ipotesi, questa, confermata dalla Ragioneria Generale dello Stato che, fin nel primo rapporto sul monitoraggio della spesa sanitaria (RGS 2014), indica tra i principali driver della spesa una dinamica dei rinnovi contrattuali superiore all'inflazione programmata, la mancanza di controllo sui fondi contrattuali che hanno determinato una amplificazione dell'impatto dei rinnovi, la mancanza di controllo sugli ordinatori di spesa e l'insufficiente sviluppo di acquisti di beni e servizi secondo procedure centralizzate e/o attraverso commercio elettronico.

Se si sposta l'analisi dall'intero sistema alle singole macro-aree territoriali, si nota, in primo luogo, che l'output per abitante si distribuisce in tutti gli anni in modo decrescente dal Nord al Sud del Paese, con le regioni NR che si distinguono per i livelli di produzione per abitante più alti pari, nel 2012, a quasi il doppio rispetto a quelli osservati nelle altre regioni.

Dal punto di vista dinamico l'andamento del livello delle prestazioni segue un trend crescente nelle regioni NR, NRS e CR soprattutto a partire dal 2008 in linea con l'andamento nazionale; invece nelle regioni SD e, in particolare nelle regioni SDS, si osserva una leggero trend decrescente nei livelli di output per abitante lungo tutto il periodo dell'analisi.

Gli input seguono in tutte le macro-aree un trend molto simile a quello nazionale descritto in precedenza con solamente alcune sporadiche eccezioni. In particolare le regioni SD e SDS si distinguono per essere da un lato quelle con più bassa dotazione dell'input lavoro per



(a) (b)

Figura 3: Analisi dell'inefficienza tecnica. Fonte: analisi proprie su dati Istat (Health for All)

abitante, dall'altro lato, invece, risultano quelle con la dotazione strumentale per abitante più alta.

Dalla distribuzione territoriale dell'inefficienza tecnica emerge, negli anni dell'analisi, un forte divaricamento in termini di efficienza tra le regioni del Centro-Nord (NR, NRS e CR) e quelle del Sud (SD e SDS). Mentre nel 1998 le regioni risultavano tutte molto vicine, nel corso degli anni si assiste ad una forte recupero di efficienza nelle regioni del Centro e del Nord a fronte di un sostanziale immobilismo delle regioni meridionali che offrono progressivamente meno servizi con un livello di inefficienza più alto rispetto al resto del Paese. Nel 2012 le regioni del Sud, incluse le Isole, presentano valori di inefficienza quasi il doppio più alti rispetto alle regioni del Centro e del Nord (incluse le regioni a statuto speciale).

Anche se lo studio delle cause dell'inefficienza tecnica richiede indagini molto dettagliate ancora in corso di definizione, l'evidenza empirica appena descritta mostra che le cause del mancato recupero di efficienza osservato nelle regioni meridionali vanno ricercate non solo in relazione ai più bassi livelli di output per abitante offerti rispetto a quanto registrato nel Centro e nel Nord della penisola ma, probabilmente, anche in relazione alla non corretta proporzione tra gli input che risulta, come discusso sopra, sbilanciata nei confronti del capitale rispetto a quanto osservato nelle regioni più efficienti.

Da ultimo, tra le cause del ritardo in termini di efficienza delle regioni del Sud è importante approfondire l'impatto generato dall'adozione dei piani di rientro che hanno interessato

principalmente le regioni meridionali[13].

Secondo quanto si legge nel rapporto RGS (2014), attualmente "*i piani di rientro costituiscono uno strumento fondamentale del sistema di governance del SSN diretto alla risoluzione delle problematiche inerenti l'efficienza e l'efficacia nell'utilizzo delle risorse messe a disposizione in relazione alla manifestazione, in talune regioni, di elevati ed insostenibili disavanzi strutturali e la presenza di gravi carenze nell'erogazione appropriata dei Livelli essenziali di assistenza (LEA)*". I singoli piani di rientro nascono con l'obiettivo di individuare ed affrontare selettivamente le cause che hanno determinato strutturalmente l'emersione di significativi disavanzi, configurandosi come veri e propri programmi di ristrutturazione industriale che incidono sui fattori di spesa sfuggiti al controllo delle regioni ripristinando una più appropriata ed efficiente erogazione dei LEA[14].

Dalla nostra analisi emerge che, probabilmente, non tutti questi obiettivi sono stati centrati soprattutto sul versante degli stimoli ad una più efficiente erogazione delle prestazioni in senso tecnico.

L'acuirsi, in termini di efficienza nella fornitura dei servizi, del divario tra il Sud e il Centro-Nord del Paese in concomitanza dell'introduzione dei piani di rientro è una evidenza empirica che, sia pur da verificare attraverso test econometrici, genera il sospetto che i piani di rientro abbiano migliorato i saldi di finanza pubblica agendo principalmente sul versante dei prezzi degli input senza apportare, però, miglioramenti nell'efficienza tecnica e soprattutto senza stimolare un livello delle prestazioni più alto in termini di quantità e qualità.

Da qui la necessità di riformare parzialmente questo strumento di governance sviluppando sistemi di monitoraggio che pongano maggiore enfasi sul livello dei servizi erogati a parità di input al fine di garantire che, a fronte di un miglioramento dei saldi finanziari, non si registri un deterioramento delle prestazioni.

Con riferimento al 2014, l'ultimo anno in cui è possibile al momento estendere la nostra analisi, dalle prime stime risulta che il recupero totale di inefficienza tecnica sia quantificabile in circa 3 mld di euro concentrati per oltre il 40% nel meridione (includendo le Isole) e per il 15% nel Lazio[15]. Le regioni che mostrano i livelli di inefficienza, in percentuale della spesa storica, più elevati, rispetto alla media nazionale del 3,4%, sono: la Calabria (5,7%), la Basilicata (5,3%), la Valle d'Aosta (5,8%). Nel gruppo delle regioni più virtuose ritroviamo, invece, il Veneto e la Lombardia con livelli di spesa inefficiente prossimi allo zero, tutte le altre regioni registrano valori di inefficienza tecnica tra il 2% e il 4% della loro spesa storica.

13 Lazio, Abruzzo, Molise, Campania, Calabria e Sicilia hanno avviato il piano di rientro nel periodo 2006-2010 e sono ancora in vigenza di piano; Piemonte e Puglia sono, invece, soggette a un piano di rientro "leggero" a partire dall'anno 2010.

14 La normativa vigente dall'anno 2010 prevede l'obbligatorietà del Piano di rientro per le regioni che presentano un disavanzo pari o superiore alla soglia del 5% del livello del finanziamento del SSR ovvero per le regioni per le quali, pur avendo un livello di disavanzo inferiore alla soglia del 5%, gli automatismi fiscali o altre risorse di bilancio della regione non garantiscano con la quota libera la copertura integrale del disavanzo.

15 Questo recupero di efficienza è da considerarsi parziale in quanto non tiene conto della componente di inefficienza legata ai prezzi degli input.

E' importante precisare che in questo rapporto, come anche nei precedenti, si considererà solamente, a fini perequativi, la spesa sanitaria pubblica finanziata dal governo centrale attraverso i Patti della Salute. Non sono considerate altre tipologie di finanziamento, che pur, come evidenziato in tabella 1 grazie ai dati del sistema SHA per la prima volta disponibili per l'anno 2014, sono consistenti e rimangono in carico essenzialmente alle famiglie; tali spese, addebitabili principalmente a motivi riabilitativi e di acquisto di beni sanitari (si veda la tabella 2), impattano potenzialmente molto sull'equità del sistema perchè sostenute indipendentemente dal reddito familiare e distintamente per regione di residenza.
Come anche, infatti, afferma Agenas nel suo rapporto "La compartecipazione alla spesa nelle Regioni, anno 2015"[16] *"è evidente che la variabilità dei sistemi di compartecipazione al costo può determinare condizioni di non equità tra i cittadini in relazione alla residenza, in cui i pazienti sono essere chiamati a corrispondere importi differenti a fronte della stessa prestazione. A rendere ancora meno equo il sistema è l'attuale sistema di esenzioni, talvolta esteso a livello regionale, che riduce il numero di cittadini tenuti a corrispondere il ticket".*
Un focus particolare dovrà sicuramente essere portato avanti nei prossimi anni sulle dinamiche distributive di tale sforzo fiscale al fine di meglio comprendere le differenze territoriali del trade-off tra spesa pubblica e privata.

Sistemi di finanziamento	Anno 2014	
	Milioni di €	% PIL
HF.1 Governo e sistemi di finanziamento sanitari obbligatori	110.642	6,9
- *HF.1.1 Governo*	110.198	6,8
- *HF.1.2 Sistemi di finanziamento sanitari obbligatori*	444	0,0
HF.2 Sistemi volontari di pagamento per l'assistenza sanitaria	3.153	0,2
- *HF.2.1 Sistemi di assicurazione sanitaria volontaria*	2.157	0,1
- *HF.2.2 Sistemi di finanziamento NPISH*[17]	505	0,0
- *HF.2.3 Sistemi di finanziamento aziendale*	491	0,0
HF.3 Pagamenti delle famiglie out-of-pocket	32.143	2,0
HF.4 Altre tipologie di finanziamento		
Tutti gli HF	145.938	9,1

Tabella 1: Spese correnti per l'assistenza sanitaria per sistemi di finanziamento, Italia 2014, Fonte: Sistema dei conti della salute (SHA - System health accounts)

Infine, il rapporto è organizzato nel modo seguente: nella sezione 2 si richiama brevemente l'importanza dell'analisi dell'efficienza nel settore pubblico; nella sezione 3 si ripercorrono i punti salienti delle precedenti edizioni di SaniRegio per descrivere, poi, in modo più dettagliato le finalità di SaniRegio2016; nella sezione 4 si descrivono gli step metodologici che

16 http://www.agenas.it/images/agenas/monitoraggio/spesa_sanitaria/normativa_ticket_regioni_2015.pdf
17 NPISH = Non-profit institutions serving households.

Funzioni di assistenza sanitaria	Tutti i sistemi di finanziamento		HF. 1 Governo	
	Milioni €	Perc. (%) PIL	Milioni €	Perc. (%) PIL
HC.1.1+HC.2.1 *Inpatient* cure e riabilitazione	41.867	2,6	40.315	2,5
HC.1.2+HC.2.2 *Day-hospital* cure e riabilitazione	6.403	0,4	5.884	0,4
HC.1.2+HC.2.3 *Outpatient* cure e riabilitazione	33.214	2,1	21.173	1,3
HC.1.4+HC.2.4 *Home-based* cure e riabilitazione	658	0,0	495	0,0
HC.3 Long-term care	13.813	0,9	10.295	0,6
HC.4 Servizi ausiliari	12.498	0,8	9.195	0,6
HC.5 Articoli medicali	28.847	1,8	16.064	1,0
HC.6 Cure di prevenzione	5.982	0,4	5.402	0,3
HC.7 Governance e amministrazione sistema sanitario	2.656	0,2	1.819	0,1
Spese correnti sanitarie	145.938	9,1	110.642	6,9

Tabella 2: Spese correnti per l'assistenza sanitaria per sistemi di finanziamento e funzione, Italia 2014, Fonte: Sistema dei conti della salute (SHA - System health accounts).

contraddistinguono l'analisi condotta in SaniRegio2016; i risultati sono riportati e discussi nelle sezioni 5 e 6 dedicate, rispettivamente, alla descrizione dei livelli di efficienza e di spesa. Nel capitolo 7, infine, si propone una prima disaggregazione spaziale dei livelli di domanda ed offerta sanitaria a livello comunale. Le conclusioni sono riportate nella sezione 8. Da ultimo i dettagli metodologici di carattere più tecnico sono riportati in Appendice.

2 OGGETTO DELLA VALUTAZIONE

Indice

2.1 Sistemi di controllo della spesa e valutazione delle politiche pubbliche 12
2.2 Il modello italiano di assistenza sanitaria: una panoramica 14
2.3 Questioni di metodo riguardanti la definizione e la misurazione dei fabbisogni sanitari 17
2.4 Domanda e offerta: una critica delle proxy esistenti 21
2.5 Localizzazione della domanda e dell'offerta 22
2.6 Non solo domanda ed offerta: efficacia ed efficienza nei sistemi sanitari nazionali 24

2.1 SISTEMI DI CONTROLLO DELLA SPESA E VALUTAZIONE DELLE POLITICHE PUBBLICHE

"*In the coming decades, the size and age-structure of Europe's population will undergo dramatic demographic changes. Ageing populations will pose major economic, budgetary and social challenges. It is expected to have a significant impact on growth and lead to significant pressures to increase public spending. This will make it difficult for Member States to maintain sound and sustainable public finances in the long-term. This is of particular importance in Economic and Monetary Union, as high deficits and rising debt in some countries leading to unsustainable public finances might have an adverse impact on macro-economic conditions for other EMU countries*".
Communication from the Commission to the Council and the European Parliament, European Economy, October 2006.

I cambiamenti demografici, le diverse dinamiche territoriali in materia di innovazione, nonché le delocalizzazioni dei centri di produzione all'estero stanno avendo un profondo impatto sulle dinamiche macroeconomiche e sulla sostenibilità a lungo periodo dei conti pubblici nazionali.
La crisi dello stato sociale indotta dal rallentamento economico, dai cambiamenti demografici, insieme con la costruzione non ancora pienamente compiuta dell'Unione Europea hanno spinto a prescrivere discipline di bilancio sempre più severe per gli Stati membri, spostando l'attenzione dei policy maker sul contenimento della spesa pubblica sia a livello centrale sia,

in particolare, degli enti locali.
Per i paesi UE l'esigenza di soddisfare i vincoli di bilancio previsti dal Patto di stabilità e crescita ha imposto un controllo più rigoroso delle finanze pubbliche e in ultima analisi "*has affected local and regional authorities and their capacity to provide public services*" (Bloomfield, 2006).
Il focus dei sistemi pubblici di contabilità nazionale si sta, dunque, spostando verso la necessità, sia nel breve termine, ma soprattutto nel medio-lungo termine, di costruire sistemi di controllo della spesa stabili, flessibili e proattivi, sia a livello centrale e, soprattutto, a livello locale.
Dati questi vincoli, la vera sfida per i governi è quella di costruire sistemi di controllo e di perequazione che permettano di mantenere i livelli di servizio pubblico, per quanto possibile, inalterati, minimizzando le inefficienze dovute ad allocazioni errate dei fattori di produzione o a disallineamenti nel dimensionamento ottimo degli Enti Locali.
Negli ultimi decenni l'interesse scientifico verso metodi di valutazione delle politiche pubbliche è aumentata proporzionalmente con la manifestazione di alcuni fattori critici che caratterizzano l'azione di governo; si pensi ad esempio alla crescente complessità dei processi decisionali che devono tener conto della specializzazione dei diversi livelli di governo (si considerino, ad esempio, i diversi livelli di governo - Commissione europea, Stato, Regioni, Province e Comuni - relativi alle competenze concorrenti nella funzione sanità), all'interdipendenza tra i programmi di Enti concorrenti tra loro, alla maggiore versatilità e gamma degli obiettivi nei settori tradizionali della politica pubblica ed infine, come si sottolineava in precedenza, alla simultanea contrazione delle risorse disponibili.
L'insieme di tali fattori ha reso i percorsi decisionali delle politiche pubbliche meno lineari e difficilmente ascrivibili entro i confini istituzionali tradizionali, assumendo, quindi, morfologie insolite e dinamiche poco chiare.
I metodi di valutazione delle politiche pubbliche hanno lo scopo di analizzare retrospettivamente i processi ed i risultati, ampliando la capacità cognitiva dei decisori e delle istituzioni politiche. I destinatari dei risultati sono quindi i responsabili politici, coloro che devono sorvegliare la produzione dei servizi pubblici ed aumentare il loro controllo/qualità.
La vera sfida, quindi, è quella di includere nella valutazione delle politiche pubbliche i vari aspetti che le qualificano anche ai fini di una successiva programmazione: la qualità e l'efficienza del servizio percepita dagli utenti, le sue prestazioni rispetto agli obiettivi che motivano la produzione, il ruolo sociale del servizio pubblico all'interno del territorio, l'accessibilità, etc; tali aspetti, spesso di difficile stima econometrica, dovrebbero, a parere degli scriventi, affiancare misure di efficienza e di efficacia che negli ultimi anni sono state maggiormente sviluppate sia da un punto di vista metodologico che applicativo.

2.2 IL MODELLO ITALIANO DI ASSISTENZA SANITARIA: UNA PANORAMICA

Negli ultimi decenni, l'Italia ha intrapreso un processo di decentramento politico e amministrativo il più delle volte senza tenere in debito conto delle differenze geografiche e socio-economiche, sia dal lato della domanda che dal lato dell'offerta. È questo il caso del comparto sanitario. Il Sistema Sanitario Nazionale (SSN) italiano, istituito nel 1978 per fornire copertura assicurativa sanitaria completa e standard a tutti i cittadini e residenti legali, è stato progettato come un sistema basato regionalmente ed organizzato a livello nazionale, regionale e locale. Dal punto di vista organizzativo, il SSN è costituito da tre diversi livelli: il governo centrale in alto, i 20 governi regionali nel mezzo[1], e le Aziende Sanitarie Locali (ASL) e gli ospedali indipendenti (IHS) in basso.
In particolare, uno dei principi fondanti del SSN è che l'assistenza sanitaria deve essere gratuita in ogni punto di erogazione sul territorio facendo sì che medesimi cittadini con stesse esigenze potessero usufruire di parità di accesso, a prescindere quindi dalle condizioni reddituali personali o del luogo nel quale risiedono. Il concetto di equità è, quindi, un requisito fondamentale per il SSN. Posta tale premessa e data la notevole variabilità delle esigenze di assistenza sanitaria della popolazione, l'assegnazione del macro budget annuale alle singole autorità sanitarie locali risulta essere un compito fondamentale, proprio in termini di equità. L'"*equitable fund allocation*" è quindi logicamente collegata con la possibilità per ogni autorità sanitaria di acquistare un pacchetto "*standard*" di assistenza sanitaria per i suoi residenti tenendo conto della domanda locale, dei costi locali di assistenza sanitaria e in ultima analisi del budget assegnato dal governo nazionale.
Per meglio modulare la pianificazione dell'offerta sulla base della domanda locale reale, a partire dal 1992, una prima serie di riforme è stata specificamente progettata al fine di aumentare l'autonomia delle Regioni in termini di finanziamento e di offerta di assistenza sanitaria. L'aumento del decentramento amministrativo e una maggiore dipendenza da fonti di finanziamento regionali, hanno necessariamente spinto ad una maggiore autonomia dal lato del finanziamento e della spesa per l'assistenza sanitaria al fine di compensare le divergenze interregionali in particolare per le Regioni italiane caratterizzate da notevoli necessità economiche, demografiche e di capacità fiscale. Gli ampi squilibri interregionali esistenti hanno quindi favorito la domanda politica volta all'introduzione di una più forte forma di federalismo fiscale[2] e, nel 2001, di una riforma costituzionale[3].
L'introduzione di un maggior decentramento fiscale ha contribuito ad abolire formalmente i trasferimenti centrali alle Regioni sostituiti da quote di imposte riscosse dagli Enti Centrali, imposte essenzialmente basate sul valore aggiunto (IVA). Contemporaneamente, un Fondo di solidarietà nazionale finanziato dall'IVA è stato istituito per ridurre le differenze tra i

1 Più precisamente 19 Regioni e due Province autonome
2 Decreto legislativo 56/2000.
3 Legge costituzionale 3/2001

redditi regionali pro capite medi e la media nazionale. Oltre a ciò, le Regioni sono state autorizzate ad introdurre proprie entrate supplementari, portando quindi ad ulteriori differenze negli schemi di finanziamento tra Regioni. Ogni anno la Conferenza permanente per i rapporti tra lo Stato, le Regioni e le Province autonome di Trento e Bolzano (con i presidenti delle Regioni ed i rappresentanti del governo centrale come suoi membri) stabilisce i criteri (di solito basati sulle consistenze della popolazione e sui profili di età) con i quali destinare le quote del fondo tra le diverse Regioni. L'assistenza sanitaria è diventata, quindi, oggetto di legislazione concorrente tra Stato e Regioni: questo significa che le Regioni hanno un'ampia autonomia nell'organizzazione e nella gestione dei servizi di assistenza sanitaria all'interno del proprio territorio, mentre il governo nazionale definisce il macro budget e le regole di allocazione finanziaria.
Il processo di regionalizzazione sanitaria ha consentito ad ogni singola Regione di organizzare e regolamentare il proprio Servizio sanitario, non solo, come in passato, in termini di quantità, qualità ed efficienza dei servizi, ma modificando il quadro normativo che regola i servizi regionali e la sua interfaccia con le altre istituzioni (per una descrizione più dettagliata, si veda il box seguente). Pertanto oggi, le Regioni godono di una notevole autonomia nella determinazione della macro struttura dei loro sistemi sanitari e delle unità sanitarie locali intra-regionali.

Modelli regionali di Health management in Italia

Anche se ogni Regione organizza e regola il proprio servizio di assistenza sanitaria in autonomia, è possibile mettere in evidenza tre macro-tipologie di servizio:

- Modello competitivo: caratterizzato da una forte concorrenza tra le sub- organizzazioni sanitarie regionali. Questo modello è stato adottato soprattutto in Lombardia, dove i cittadini-pazienti possono scegliere liberamente tra una vasta gamma di fornitori pubblici e accreditati di servizi privati in competizione tra loro. Le unità sanitarie locali mantengono il loro ruolo di "progettista" e di controllo delle spese tramite tetti di spesa di volta in volta concordati con ogni fornitore.

- Modello integrato. Si distingue per la cooperazione e l'integrazione della organizzazione sanitaria, indipendentemente dalla proprietà e tipologia ed è caratterizzato da un elevato grado di negoziazione e pianificazione. Questo modello è stato adottato, con differenti caratterizzazioni, in Toscana, Emilia-Romagna e Regioni del Nord-est come Veneto o Friuli Venezia Giulia. In queste reti di fornitura di servizi, ogni struttura è essenziale per l'intero sistema e complementare alle altre. Le Regioni hanno optato per accordi contrattuali negoziati tra fornitori così come produttori. Molti servizi sono concentrati su pochi strutture di grandi dimensioni e il numero di ospedali è limitato secondo un principio di razionalizzazione, che mira a ridurre i costi e a migliorare la qualità dei servizi forniti. L'accreditamento di nuovi soggetti privati è rigorosamente limitato ex-ante e controllato ex-post.

- Modello burocratico: modello nel quale i burocrati gestiscono in modo pervasivo il sistema sanitario con una pianificazione di lungo periodo, un controllo gestionale e accordi contrattuali limitati. Gli ospedali ricevono finanziamenti diretti dalla Regione tramite accordi sui massimali di bilancio. L'integrazione o la cooperazione di enti privati accreditati e pubblico è quindi molto indebolita e di solito non efficiente. Questo modello è stato adottato soprattutto nelle regioni meridionali, come ad esempio Campania, Calabria e Basilicata.

Allo scopo di superare gradualmente le criticità legate al criterio di allocazione finanziaria basato sulla spesa storica, la legge finanziaria del 2008 ha avviato il passaggio alla determinazione del fabbisogno sulla base dei livelli di assistenza, calcolando il finanziamento regionale a partire dai costi effettivi sostenuti delle Regioni considerate più efficienti. Tuttavia è solo a partire dal 2013, che i trasferimenti regionali vengono calcolati in base ai "costi standard". Nella configurazione attuale la metodologia dei costi standard prevede l'identificazione di alcune Regioni di riferimento, scelte tra quelle più efficienti non solo in termini finanziari, ma anche di adeguatezza nella fornitura di servizi sanitari (anche se tale aspetto non viene tenuto in debita considerazione); la presente ricerca, quindi, vuole proporre un metodo più completo ed equo di calcolo dei fabbisogni standard che tenga in considerazione gli aspetti di efficienza tecnica, di efficienza di prezzo e soprattutto del livello di servizio effettivamente erogato.
Attualmente, il macro budget del settore sanità è suddiviso tra le diverse Regioni utilizzando una quota pro capite, parzialmente corretta per la diversa distribuzione dell'età della popolazione senza tenere adeguatamente in considerazione l'intero set di differenze nella domanda di assistenza sanitaria nonché le differenze territoriali riguardanti i costi/prezzi locali. Il governo nazionale da un lato ha cercato - pur con margini di miglioramento da un punto di vista statistico-econometrico - di standardizzare il reddito regionale in base a criteri oggettivi e più strettamente legati alle esigenze di salute. Tuttavia il lato dell'offerta è regolato solamente dai "livelli essenziali di assistenza", un generico pacchetto nazionale di servizi che devono essere garantiti a tutti i residenti in ogni Regione.
Più di recente, e in particolare nel 2014-2016, i Patti per la Salute tra Stato e Regioni hanno cercato di superare il sistema verticale previsto dalla riforma del 1992, sostituendo la moltitudine di unità regionali ed aziende sanitarie ospedaliere, autonome e concorrenti, con l'adozione di una "*wide area holding*" che permetta di ottenere maggiori economie di scala. Tuttavia se la riduzione del numero di delle ASL (fino addirittura a prevedere una singola ASL per l'intera regione), permette di realizzare economie di scala e da maggiore indipendenza alle Regioni, accelera però la frammentazione del sistema in termini di organizzazione dei servizi regionali (France & Taroni, 2005).
La storia recente del SSN dimostra che il decentramento ha amplificato le differenze interregionali (Ferré *et al.*, 2014) nonostante fosse stato previsto un meccanismo di allocazione delle risorse che mirava a garantire equità e a mitigare le disuguaglianze interregionali e

intra-regionali nella fornitura di assistenza sanitaria pubblica.
Con una sempre maggiore tendenza verso un sistema sanitario centrato sul paziente, la richiesta di una migliore corrispondenza tra bisogno e servizio si è intensificata.

Dovendo realizzare un sostanziale decentramento in un paese con forti divari socio-economici, è necessario stimare ex-ante con precisione le differenti esigenze delle comunità locali. In altri termini, le difficoltà nello stabilire standard equi e nel raggiungimento di una migliore corrispondenza tra fabbisogno stimato e servizio effettivo richiedono l'utilizzo di strumenti statistici e metodologici robusti che consentano di sintetizzare l'offerta e la domanda di beni e servizi pubblici locali ad una scala di dettaglio sempre minore. Questa esigenza è più cogente per le Amministrazioni Pubbliche locali per comprendere e - se necessario - modificare la propria offerta sul territorio: sistemi di misurazione robusti e territorialmente micro-fondati delle prestazioni, dovrebbe essere costruiti e monitorati al fine di garantire l'allineamento con le politiche chiave e di identificare le possibili aree di miglioramento (Goddard & Jacobs, 2009).
Date queste premesse, accanto alla tendenza ad una nuova centralizzazione della pianificazione e del controllo, che si estende a tutte le Regioni, si nota una concomitante mancanza di adeguati sistemi di monitoraggio dei bisogni di salute dei gruppi più vulnerabili della popolazione (anziani, disoccupati e residenti nelle aree suburbane). Pertanto nelle pagine seguenti si propone, accanto ad un macro modello di stima dei fabbisogni standard regionali, un possibile strumento di analisi per valutare l'incontro tra domanda ed offerta ad un inedito livello di dettaglio territoriale.
Lo scopo ultimo è quello di stimolare l'attenzione dei policy maker verso gli effetti di una non corretta allocazione dei servizi sul territorio che potrebbe determinare maggiori costi privati soprattutto per le popolazioni lontane dai grandi centri urbani.

2.3 QUESTIONI DI METODO RIGUARDANTI LA DEFINIZIONE E LA MISURAZIONE DEI FABBISOGNI SANITARI

Come già sottolineato nel paragrafo precedente, l'Italia è caratterizzata da ampie differenze spaziali nonché da una distribuzione ineguale delle risorse a livello intra e inter regionale, in particolare in termini di numero di posti letto, di disponibilità di attrezzature mediche avanzate, di servizi privati nonché di servizi di assistenza della comunità. Le disparità di reddito, legate allo stato di salute, sono anch'esse significative, ma simili a quelle rilevate in altri paesi europei (van Doorslaer & Koolman, 2004) come Olanda o Germania. Una delle conseguenze di questa distribuzione ineguale è la forte mobilità interregionale da Sud verso Regioni del Centro o del Nord Italia (France, 1997), mobilità che presenta una crescita significativa proprio negli anni '90. Una sfida importante rimane il crescente divario tra Nord e Sud nella fornitura e qualità dei servizi (Toth, 2014).

Ma come valutare correttamente il servizio sanitario dal lato della domanda?
Una maggiore o minore domanda di assistenza sanitaria è solitamente legata alle differenti condizioni di salute di un individuo, al livello di spesa che l'individuo deve affrontare nonché al proprio livello di reddito (si veda il box seguente). Tuttavia, nei paesi, come l'Italia, dove l'assicurazione pubblica in larga parte , o privata, copre solitamente tali spese mediche, il paziente non sopporta tali costi che sono a carico della fiscalità generale. Se questo è il caso, la domanda di assistenza sanitaria non è sensibile solo al prezzo e non è influenzata dalla disponibilità a pagare per tali servizi (Pauly, 1986).
Di conseguenza, nei paesi con un sistema sanitario pubblico, sono i vincoli all'offerta che determinano la dimensione della spesa sanitaria (Tulchinsky & Varavikova, 2009, p. 411): Si assiste, pertanto, ad una costante incapacità dell'offerta ad allinearsi con i livelli di domanda. In questo contesto è possibile determinare i fabbisogni standard relativi, in un'ottica di allocazione di macro budget, ma è molto difficile individuare le proxy utili a determinare la domanda.

Determinanti della spesa in campo sanitario

Un'attenta analisi delle letteratura mostra come le seguenti variabili aiutino a spiegare i differenziali territoriali e possano essere considerate come le principali determinanti della spesa sanitaria:

- Reddito. Un lavoro molto noto in letteratura di Newhouse (1977) identifica la ricchezza di un paese come la principale determinante della spesa sanitaria. Altri studi successivi hanno confermato questo risultato sempre in un setting di tipo cross-section. Tuttavia, negli anni '90, l'uso di serie storiche e la disponibilità sempre maggiore di dati panel permettono di stimare con più precisione l'elasticità della domanda al reddito vicino all'unità, il che implica considerare l'assistenza sanitaria come un bene normale (Culyer, 1990; Hitiris & Posnett, 1992). Questo risultato è confermato nello studio di Hitiris (1997), che stima un'elasticità al reddito tra 1,0 e 1,2. Ulteriori studi, sempre applicando l'analisi dei dati panel, ma aumentando la dimensione del campione e introducendo effetti fissi regionali, sono stati in grado di monitorare una vasta gamma di caratteristiche che variano nel corso del tempo, la cui assenza può distorcere il rapporto tra spesa sanitaria e PIL. Nel complesso, diversi studi nel tempo con diversi approcci econometrici (Auteri* & Costantini, 2004) hanno ampliato di molto il campo di variazione della stima dell'elasticità della domanda al reddito, da valori leggermente negativi (-0,082 in Crivelli *et al.*, 2006) a valori nettamente superiori di 1 come in uno dei modelli di Roberts (2000) con coefficiente pari a 1,25.
  Per quanto riguarda l'Italia, Giannoni & Hitiris (2002) propongono di spiegare le determinanti della spesa sanitaria e le sue differenze regionali nel periodo 1980-1995, stimando che il livello del reddito (con una elasticità pari a 0,33) e l'invec-

chiamento della popolazione siano le principali variabili esplicative. Altre variabili considerate, come ad esempio il numero dei posti letto negli ospedali sembrano indicare l'esistenza di economie di scala a livello regionale, così come il contributo positivo alle spese del personale sanitario e non sanitario per ospedale.
Altri paper identificano il reddito come un chiaro e significativo fattore esplicativo dell'aumento della spesa sanitaria: Herwartz & Theilen (2003) stimano un'elasticità al reddito della domanda di servizi sanitari pari a 0,74; Koenig *et al.* (2003) 0,760 in un unico modello e 0,312 in un altro, mentre Mosca (2007) 0,682.

- Invecchiamento della popolazione. Negli anni '90, numerosi studi hanno analizzato l'impatto dell'invecchiamento della popolazione sulla spesa sanitaria. I risultati di Blomqvist & Carter (1997), in una analisi panel di 24 paesi OCSE per il periodo 1960-1991, indicano come il numero di persone di età superiore ai 65 anni sia un fattore predominante nell'aumento della spesa sanitaria. Altri autori giungono a risultati similari; si vedano ad esempio: Barros (1998); Zweifel *et al.* (1999); Felder *et al.* (2000); Herwartz & Theilen (2003); Seshamani & Gray (2004); Dormont *et al.* (2006); Werblow *et al.* (2007).
  Di Matteo & Di Matteo (1998) identificano l'invecchiamento della popolazione come principale fattore determinante della spesa sanitaria delle province canadesi e risultati simili sono forniti anche da Karatzas (2000) negli Stati Uniti. Nella stessa linea di ricerca, gli studi di Roberts (2000) e Giannoni & Hitiris (2002) indicano l'invecchiamento della popolazione come il principale responsabile per l'aumento della spesa sanitaria, in unione al livello del reddito. Di Matteo (2005), esaminando i gruppi di età della popolazione tra i 18 ed i 44, 45-64 e 64-74 e più di 74 anni, trova una relazione positiva tra la spesa sanitaria e la popolazione over 74 responsabile di essere la principale covariata per l'aumento della spesa sanitaria. Per la Svizzera, Crivelli *et al.* (2006) giunge alla conlusione che sia la percentuale di persone anziane che maggiormente determini l'aumento della spesa sanitaria. Tuttavia, non ci sono prove chiare ed unanimi a considerare l'invecchiamento della popolazione solamente come la principale variabile relativa all'aumento della spesa sanitaria. Getzen (1992), ad esempio, unendo le serie temporali con studi di tipo cross-section in 20 paesi dell'OCSE per il periodo 1960-1988, suggerisce che l'invecchiamento della popolazione non sia necessariamente associabile ad una maggiore spesa sanitaria. Gerdtham (1993) analizzando l'impatto dell'invecchiamento della popolazione della Svezia - dati 1970-1985, trova che i cambiamenti demografici spiegano solamente il 13% del totale dell'aumento della spesa sanitaria nel periodo considerato. Studi recenti sostengono che sia piuttosto la vicinanza alla morte, più che l'invecchiamento della popolazione, la variabile che spiega maggiormente l'aumento della spesa sanitaria.

- Vicinanza alla morte. Zweifel *et al.* (1999) e Felder *et al.* (2000) sono gli autori più importanti per i quali la vicinanza alla morte è il principale fattore determinante della spesa sanitaria. In particolare, lo studio di Zweifel *et al.* (1999) è il precursore di una serie di ricerche che includono la vicinanza alla morte come variabile esplicativa della spesa sanitaria. Secondo i loro risultati, l'invecchiamento della popolazione non spiega del tutto le differenze di spesa sanitaria. Solo gli ultimi due anni di vita (otto trimestri) degli individui, infatti, sono determinanti indipendentemente dall'età del soggetto. I risultati di Felder *et al.* (2000) sono simili a quelli di Zweifel *et al.* (1999); secondo il loro studio, la spesa sanitaria diminuisce con l'aumentare dell'età della persona, anche se aumenta con l'avvicinarsi della morte. Seshamani & Gray (2004) risultati confermano quelli di Zweifel *et al.* (1999); pertanto, anche se l'età e la vicinanza della spesa sanitaria morte influenza, l'influenza di quest'ultimo è molto maggiore, poiché non vi è un aumento della probabilità di essere in ospedale (incorrere in una spesa sanitaria maggiore di zero) con la vicinanza alla morte. Tra gli altri, Breyer & Felder (2006), Werblow *et al.* (2007) e Evans *et al.* (2001) confermano che la vicinanza alla morte, piuttosto che l'invecchiamento della popolazione, è la variabile che meglio spiega l'aumento della spesa sanitaria. Evans *et al.* (2001), in particolare, ipotizzano che l'invecchiamento della popolazione non è sinonimo della spesa sanitaria superiore, ma che quest'ultimo si concentra nei 4 mesi, o l'anno, prima della morte, indipendentemente dall'età del soggetto. Ma non tutti i ricercatori hanno parere unanime: Breyer & Felder (2006) ha trovato che i costi degli ultimi anni di vita provocano l'impatto demografico della spesa pro capite per diminuire. Si stima l'impatto del cambiamento demografico sulla spesa sanitaria in Germania. Inoltre, i risultati di Werblow *et al.* (2007) indicano che l'età è di poco significato se la vicinanza alla morte è controllato per.

- Altre determinanti della spesa sanitaria.

  L'impatto della tecnologia viene in risalto come variabile esplicativa della crescita della spesa sanitaria, con lo studio di Newhouse (1992). Anni dopo, in un articolo di primo piano da Okunade & Murthy (2002) progresso tecnologico si conferma come la principale determinante della spesa sanitaria negli Stati Uniti nel periodo considerato. Negli studi di Herwartz & Theilen (2003) e di Koenig *et al.* (2003), spesa sanitaria si spiega con variabili dipendenti raggruppati nelle seguenti categorie: le condizioni economiche e demografiche generali; stato di salute; pagamento ai fornitori; assicurazione sanitaria; fornitura di medici e specialisti; struttura di mercato dei fornitori; i costi di gestione; regolamentazione delle linee guida e delle tecnologie sanitarie e di trattamento. I risultati principali sono che la struttura demografica della popolazione è l'elemento guida della spesa sanitaria. Altri risultati indicano che un aumento del 10% della popolazione immigrata è associato a un

aumento della spesa per i servizi di assistenza sanitaria di 91 %, Stearns & Norton (2004) confrontare i costi sanitari futuri della popolazione tra i 66 ei 99 anni di età in Inghilterra (1992-1998): i risultati più importanti sono che il genere e la posizione geografica non possono essere considerate variabili significative della spesa sanitaria e la vicinanza alla morte aumenta il potere di spiegare spesa sanitaria. Al contrario, Dormont *et al.* (2006) analizzare le determinanti della spesa sanitaria in Francia (1.992–2.000) scoprendo che l'invecchiamento della popolazione impatto sulla spesa è relativamente piccolo. Infine, lo studio di Mosca (2007), sulla base di un campione di 20 paesi dell'OCSE, indica che il decentramento influenza positivamente la spesa sanitaria.

Dal lato dell'offerta, invece, i recenti sviluppi della ricerca politico-sanitaria (Carinci *et al.*, 2015) hanno messo in evidenza l'importanza di valutare le prestazioni dei servizi sanitari attraverso una serie il più possibile completa ed esaustiva di output/outcome. Le attuali tendenze e nuove risultanze della ricerca invitano a valutare attentamente gli effetti del settore sanitario sullo stato di salute dell'individuo. La transizione tecnologica ha portato ad una serie di strumenti di prevenzione, terapeutici e diagnostici innovativi che possono ridurre l'impatto delle malattie (Okunade & Murthy, 2002) sui sistemi sanitari. In questo contesto, ci sembra più cogente domandarsi se l'offerta di servizi medici e di prevenzione sia conforme o meno alla domanda locale da un punto di vista territoriale.

2.4 DOMANDA E OFFERTA: UNA CRITICA DELLE PROXY ESISTENTI

I sistemi sanitari nazionali, basati su regole di finanziamento capitarie, ovvero funzione del numero di pazienti, sono basati sull'ipotesi che il sistema è in equilibrio e la domanda di prestazioni sanitarie dei cittadini-pazienti è perfettamente soddisfatta; tuttavia la realtà è ben diversa e vi possono essere forti disallineamenti tra domanda e offerta. La valutazione di un sistema sanitario pubblico è, quindi un compito molto difficile dato il gran numero di aspetti - la salute della popolazione, gli esiti derivanti dal trattamento, la qualità clinica e l'appropriatezza delle cure, la reattività della domanda, l'equità e la produttività - spesso dipendenti tra loro.
E' possibile, comunque in via preliminare, individuare tre evitabili ampie fonti di disuguaglianze sanitarie, che in primo luogo possono essere causate da *(1)* variazioni nella qualità dei servizi sanitari, *(2)* un diverso accesso ai servizi sanitari, vale a dire dalla diversa possibilità di raggiungere ed ottenere servizi sanitari adeguati in situazioni di bisogno o da *(3)* fattori che esulano dal controllo diretto del sistema sanitario, come la ricchezza, lo stile di vita, la genetica ed altri fattori legati a caratteristiche ambientali.

In particolare un migliore o peggiore accesso alle cure sanitarie solitamente dipende dalle caratteristiche individuali, dagli ambienti sociali e fisici nonché dalle caratteristiche dei sistemi sanitari, degli organizzazioni e dei fornitori.
Pertanto, dal lato dell'offerta occorre considerare le caratteristiche dei sistemi e delle organizzazioni sanitarie, mentre dal lato della domanda le caratteristiche delle popolazioni. I fattori che influiscono sul processo produttivo sembrano essere invece legati essenzialmente alle modalità di accesso.
Pertanto, l'accesso è la caratteristica chiave che lega la possibilità di identificare i bisogni di salute, individuare i servizi sanitari, raggiungere facilmente le strutture sanitarie, di ottenere o utilizzare servizi di assistenza sanitaria e, in ultima istanza, di usufruire di servizi adeguati alle proprie esigenze. Accessibilità è, dunque non solo la disponibilità delle strutture sanitarie (la densità, la concentrazione o la loro distribuzione sul territorio) nei contesti urbani rispetto agli utenti, ma è anche un concetto legato all'assenza di costi privati aggiuntivi per coloro i quali non possono fruire di un'offerta adeguata da un punto di vista territoriale; l'accesso è limitato, in altri termini, se le risorse disponibili non sono distribuite equamente in tutto un paese.
Questi tre fattori potrebbero anche non avere un impatto simultaneo su un determinato individuo; tuttavia ogni fattore comporta implicazioni specifiche per l'allocazione delle risorse.
Servizi di scarsa qualità sono, in linea di principio, un problema di cattiva gestione. Una corretta allocazione di risorse monetarie, infatti, può essere spesa non in modo ottimale perché l'organizzazione degli stessi servizi locali è sub-ottimale. In questo caso non vi è alcuna necessità di una modifica al sistema di finanziamento. Piuttosto, l'attenzione dovrebbe essere rivolta ad un miglioramento della gestione delle risorse. Anche per il secondo fattore - l'accesso - le tradizionali formule di allocazione delle risorse potrebbero rivelarsi inadeguate. La sfida più ardua nel disegno di un sistema sanitario è data dall'impatto dei fattori ambientali sulla fascia di popolazione più svantaggiata. Le disuguaglianze sanitarie possono derivare sia da fattori ambientali territoriali, ma anche da fattori legati al singolo individuo come quelli genetici, di reddito, di stile di vita. E' necessario quindi scendere moltissimo di granularità per tenere conto di tali fattori; è quello che in questa versione del rapporto SaniRegio è per la prima volta introdotto, nel capitolo 7, con l'intento di evidenziare come informazioni maggiormente disaggregate possano suggerire evidenze nuove ed inedite impossibili da individuare a livello regionale.

2.5 LOCALIZZAZIONE DELLA DOMANDA E DELL'OFFERTA

Proporre un modello applicativo volto alla misurazione dei Fabbisogni Standard in Italia è un compito complesso in quanto spesso risulta difficile distinguere variabili di domanda da fattori di offerta.
I principali modelli proposti in letteratura (si veda ad esempio Atella *et al.* (2014)) possono

risultare incompleti sotto tre aspetti principali: *(1)* nel non riuscire a catturare correttamente il processo di domanda di assistenza sanitaria, *(2)* nel non modellare correttamente l'interazione tra risorse sanitarie ed offerta e *(3)* nel non trattare correttamente differenze territoriali dovute a vincoli amministrativi o a fattori di rischio interni ad una stessa Regione.
Queste difficoltà sono imputabili più che all'imperizia dei ricercatori, quanto al processo di selezione delle variabili e alla scelta territoriale entrambi limitati dalla (scarsa) disponibilità dei dati. Questa ricerca vuole superare anche solo parzialmente tali difficoltà dedicando una particolare attenzione agli aspetti di localizzazione della domanda e offerta (pur in mancanza di informazioni ufficiali a livello di piccole aree geografiche).
In letteratura, sono stati proposti diversi schemi concettuali per modellare i fabbisogni sanitari e la relativa variabilità geografica; la maggior parte di questi modelli presenta una base comportamentale.
Il primo modello comportamentale teorico, apparso alla fine del 1960 (Andersen, 1995), offre due quadri concettuali delle determinanti che sono stati sviluppati in seguito da Andersen & Newman (1973) e quasi venti anni più tardi da Evans & Stoddart (1990).
Andersen & Newman (1973) modellano il fabbisogno e l'utilizzo di servizi sanitari come un processo comportamentale influenzato da tre insiemi di determinanti (caratteristiche individuali, risorse ambientali e necessità) e dai rapporti interattivi tra queste determinanti. Evans & Stoddart (1990), cercano invece di cogliere la natura dinamica e ricorsiva dei servizi sanitari, introducendo gli effetti sulla salute e il concetto di soddisfazione individuale. Il modello proposto successivamente da Carr-Hill *et al.* (1994), è stato utilizzato come riferimento nel contesto di allocazione delle risorse nelle piccole aree (Rice *et al.*, 1999; Gravelle *et al.*, 2003).

Quest'ultima analisi è particolarmente rilevante ai fini del presente lavoro. Infatti nell'attuale versione di SaniRegio l'attenzione - oltre che alla stima macro del Fabbisogno Standard per Regione - è rivolta alla stima della domanda di assistenza sanitaria e l'offerta a livello di *small area*, nell'ipotesi che in un sistema sanitario pubblico le scelte istituzionali ed organizzative delle Regioni hanno una più forte influenza rispetto alla scelta dei singoli individui (Alexander *et al.*, 1999).
Gli studi geografici e di ricerca operativa hanno offerto una prospettiva più pragmatica rispetto ai modelli di interazione spaziale (SIMs) (più comunemente sotto forma di modelli gravitazionali) usati per descrivere e prevedere flussi di utilizzo delle risorse sanitarie da parte della popolazione locale. Tali modelli considerano piccole aree geografiche e la relativa domanda e offerta locale è modellizzata in termini di punti di fornitura. Tuttavia, tali SIMs sembrano non tenere in dovuto conto dei processi di domanda di assistenza ospedaliera (come ad esempio l'interazione tra ospedale e cure primarie). Sono sicuramente molto utili a descrivere il sistema (Cho, 1998), ma insufficienti a prevedere i flussi di utenti (Porell & Adams, 1995). Per queste ragioni, uno degli obiettivi di questa ricerca è quello di fare un passo ulteriore nel calcolo di un indice sintetico che migliora la precedente comprensione delle interazioni dei fabbisogni nonché del livello di offerta a livello comunale.

Si osserva che differenze nell'accessibilità dovute a fattori geografici e territoriali influenzano la domanda (Santana, 1999) e il livello di accessibilità è solitamente più basso nelle aree

più povere del paese (Santana, 1993). La cura che un individuo riceve è pertanto funzione delle caratteristiche demografiche, sociali ed economiche della famiglia, ma anche delle caratteristiche dell'ambiente in cui vive (Andersen, 1995). Questo fattore è ancor più rilevante in un paese come l'Italia dove esistono forti discontinuità e differenze orografiche che hanno un diretto effetto sul costo opportunità della domanda della salute (correlati con un maggior tempo e spese di viaggio). Ostacoli derivano dalla posizione delle fonti di offerta di servizi sanitari che influenzano direttamente le distanze ed i tempi di percorrenza. A sua volta la diversa orografia può interagire direttamente con l'accessibilità dello stesso trasporto influenzando le possibilità di accesso al servizio sanitario. Infine, le stesse caratteristiche organizzative legate a fattori territoriali possono avere un impatto sull'accesso. In Italia, il settore sanitario si avvale di un sistema di gatekeeper per l'accesso al settore ospedaliero in quanto sono i centri di assistenza primaria ad essere destinati a fornire il punto di primo accesso al sistema sanitario (in una sorta di prefiltro). Tuttavia, se i punti di accesso non sono distribuiti equamente, allora la possibilità di scelta e l'accesso effettivo possono essere parzialmente limitati.

2.6 NON SOLO DOMANDA ED OFFERTA: EFFICACIA ED EFFICIENZA NEI SISTEMI SANITARI NAZIONALI

I risultati di un'attività economica volta alla fornitura di un servizio ad un pubblico di utenti, sia essa pubblica o privata, oltre che sull'effettivo soddisfacimento della domanda, possono essere valutati sotto altri due aspetti, spesso confusi o utilizzati come sinonimi: quello dell'efficacia e quello dell'efficienza.

Un'attività utilizza risorse per fornire un prodotto attraverso il quale raggiungere uno scopo. L'efficacia si misura con riferimento al raggiungimento dello scopo. L'efficienza si misura con riferimento al rapporto tra risorse impiegate e prodotto ottenuto. Nel settore privato la valutazione dell'efficacia di un'attività viene effettuata dal mercato: è il fatto che un prodotto sia acquistato dai consumatori, che un'azienda consegua un profitto e che conquisti quote di mercato a fornire una misura osservabile dell'efficacia dell'attività svolta. Nel settore pubblico, invece, non esistono in generale indicatori specifici e osservabili dell'efficacia diversi dalla sanzione elettorale. La costruzione di indicatori di efficacia nel settore Pubblico incontra, infatti, molteplici difficoltà, legate soprattutto alla definizione degli obiettivi e alla verifica del loro conseguimento. Fra le criticità applicative possono essere elencate le seguenti:

- i servizi pubblici spesso perseguono più obiettivi simultaneamente rendendo necessario un criterio di ponderazione o una scala di priorità condivisa (l'istruzione pubblica, ad esempio, mira a fornire nozioni, creare professionalità, formare cittadini, facilitare l'inserimento nel mondo del lavoro, ecc.);

- non sempre ad ogni obiettivo corrisponde un risultato "misurabile" (l'obiettivo di un

sistema sanitario può essere l'allungamento della vita media e il miglioramento della sua qualità, ma come si misura quest'ultima?);

- i risultati possono prodursi su un arco di tempo prolungato (l'attività di ricerca svolta oggi può produrre risultati tra molti anni);
- gli utenti dei servizi non sono omogeneamente distribuiti - fattori di domanda differenti (i pazienti trattati da un ospedale possono essere mediamente più "gravi" di quelli trattati da un altro).

In alcuni casi si è tentato di ovviare utilizzando indagini sul grado di soddisfazione dei cittadini come succedaneo del giudizio del mercato. Tuttavia tali indagini presentano diverse difficoltà come ad esempio il fatto che i giudizi possono non essere formulati sulla base di esperienze dirette. Inoltre, essi si fondano sul confronto tra risultati e attese e la disomogeneità di queste ultime può rendere problematico l'utilizzo delle indagini sul grado di soddisfazione per il confronto internazionale fra sistemi e modelli organizzativi dei servizi pubblici.

La valutazione dell'efficienza, seppure non priva di difficoltà specifiche, appare meno complessa. Per valutare se e quanto un'impresa, pubblica o privata, sia efficiente occorre verificare se non sia possibile produrre di più con le stesse risorse, o utilizzare meno risorse per ottenere la stessa quantità di prodotto. Ciò richiede la specificazione di funzioni di produzione o di costo, o almeno l'individuazione delle variabili di input e di output al fine di pervenire ad una misurazione dello scarto fra ciò che si osserva e i valori ottimi teorici.

La valutazione e il miglioramento continuativo delle prestazioni in termini di output e di costo dei sistemi sanitari nazionali sono diventate un argomento chiave nelle politiche dei paesi maggiormente sviluppati; negli ultimi anni, sono state create autorità nazionali, come ad esempio lo United Kingdom National Health Service (NHS) e il Canadian Institute for Health Information. Inoltre l'analisi di ricerca applicata di singoli ricercatori o centri di ricerca (si veda ad esempio Jencks *et al.*, 2000, Kwon, 2003 o Nuti *et al.*, 2011 e Pammolli & Salerno, 2011 per il sistema sanitario nazionale italiano) è divenuta progressivamente più estesa e puntuale.

Nella letteratura internazionale la valutazione delle prestazioni, dell'efficienza o della sola analisi dell'offerta è stata effettuata a livello internazionale, nazionale, regionale, locale o istituzionale (Ibrahim, 2001). La valutazione di un sistema sanitario può essere progettata secondo una pluralità di soggetti ed obiettivi: l'ampiezza di indicatori di performance, di spesa o di servizio, quindi, può essere considerevole spaziando dall'analisi dello stato dell'intero sistema sanitario di una nazione per riflettere le esperienze dei singoli pazienti.
Si pone, quindi, una domanda chiave, specialmente in un sistema complesso quale quello sanitario: cosa deve essere misurato? In generale, i sistemi sanitari possono essere valutati rispetto a fattori quali la qualità/quantità di cure, l'accesso alle cure stesse o al costo/spesa; queste dimensioni di analisi, se pur di forte interesse non sono le uniche dimensioni ana-

litiche: Paakkonen & Seppala (2014), ad esempio, valutano nella loro analisi l'accessibilità, l'efficienza e l'uguaglianza di trattamento.

Ma la vera sfida nei modelli di valutazione, almeno da un punto di vista statistico ed econometrico, è la *multidimensionalità* dei sistemi sanitari nazionali; tale natura dell'oggetto di valutazione, da un punto di vista applicativo, è strettamente connesso con lo sviluppo e l'impiego di tecniche robuste ed affidabili che misurino in modo sintetico misure complesse (si veda il capitolo 9); al fine, ad esempio, del confronto e del miglioramento delle prestazioni, della qualità o del livello del servizio offerto è fondamentale disporre di: i) una serie di indicatori affidabili e confrontabili costruiti su un buon sistema informativo e ii) metodi robusti ed affidabili che permettano di integrare l'informazione in un'unica misura anche al fine di indicare parametri di riferimento o benchmark.

Più in particolare, Smith (2002a) evidenzia tre aspetti critici da un punto di vista metodologico relativamente agli indicatori compositi nelle analisi del settore sanitario: i) il calcolo del set dei pesi, ii) la sterilizzazione degli effetti delle variabili esogene sulle prestazioni del sistema e iii) le ipotesi alla base dei modelli di calcolo; su questi aspetti, egli osserva, inoltre, che non vi è un ampio consenso né una metodologia condivisa soprattutto sul metodo per individuare un set ottimale di pesi da utilizzare per formare l'indice composito.

Da ultimo, osservano Bankauskaite & Dargent (2007), le misure aggregate relative all'offerta o alla domanda di prestazioni sanitarie mancano di "*precision and combine uncertain weighting systems, imprecision arising from the potential non-comparability of component measures, and misleading reliability in the form of whole-population averages that mask distribution issues*".

Date queste premesse, nell'ambito della presente ricerca verranno proposti metodi di stima robusti dell'offerta e dell'efficienza tecnica, affinché le stime non siano affette dalla presenza di dati anomali e i risultati ottenuti non siano legati alle scelte del singolo ricercatore rispetto ai pesi da assegnare alle singole dimensioni di analisi.

3 SANIREGIO2016

Indice

3.1 SaniRegio1 e 2: cosa abbiamo imparato 27
3.1.1 SaniRegio1 27
3.1.2 SaniRegio2 28
3.1.3 SaniRegio2015 30
3.2 Obiettivi e struttura di SaniRegio2016 31

3.1 SANIREGIO1 E 2: COSA ABBIAMO IMPARATO

Al fine di cogliere al meglio le differenze e le similitudini tra l'analisi condotta in SaniRegio2016 rispetto a quella condotta nelle precedenti versioni, ripercorriamo brevemente gli aspetti principali delle versioni 1 e 2 del SaniRegio e di SaniRegio2015.

3.1.1 SaniRegio1

Il primo rapporto denominato SaniRegio (SaniRegio1, https://ideas.repec.org/p/ern/wpaper/02-2009.html), ha rappresentato un primo tentativo di analizzare i determinanti della spesa sanitaria pubblica nelle Regioni italiane.

Il modello originale - basato su un modello panel a effetti fissi finalizzato al benchmarking regionale - ha stimato una relazione funzionale media che lega la spesa sanitaria alle variabili esplicative selezionate tra quelle riguardanti il contesto socio-economico, la struttura demografica, gli stili di vita e la dotazione di capitale fisico e umano. Il confronto tra la spesa effettiva, ossia quella che entra in contabilità nazionale, e la spesa spiegata dalla relazione funzionale media, ha fornito scostamenti regionali che possono essere imputati a inefficienza, a differenze qualitative delle prestazioni offerte sul territorio o al mix delle due. Il panel ha fornito, per ogni regione, lo scostamento rispetto alla spesa standard, misurato in differenza rispetto alla regione con lo scostamento minore.

Alla definizione del panel è seguito un secondo step che è consistito nella costruzione della frontiera di spesa efficiente che ha contrapposto due dimensioni spesa-qualità grazie a un set ampio di variabili esplicative quali outcome del sistema sanitario, grado di soddisfazione del SSR da parte dei cittadini, mobilità interregionale e rischio di ospedalizzazione nelle regioni.

Ciò ha consentito di valutare quanto degli scostamenti della regione dalla sua spesa standard trovino giustificazione in una qualità delle prestazioni relativamente più elevata. La frontiera è determinata da quelle regioni che presentano gli scarti minori dal proprio livello di spesa standard e gli indici di qualità più alti. La frontiera di spesa efficiente, così costruita, ha permesso di valutare quanto di questo scostamento poteva o non poteva attribuirsi a miglior qualità delle prestazioni.

Le Regioni che riescono a utilizzare meglio le risorse sono state identificate in Friuli Venezia Giulia, Lombardia, Umbria, Piemonte, Marche e Toscana. La distanza di una regione dalla frontiera ha fornito una misura del grado di inefficienza. Questa distanza può essere misurata sia in verticale (output oriented approach) sia in orizzontale (input oriented approach). Si è preferita la misura orizzontale, che permette, a parità di qualità, di definire la spesa efficiente per ottenere un dato livello di qualità.

In conclusione l'analisi ha evidenziato differenze marcate tra regioni, con quasi tutte quelle del Sud e Isole che spendono molto di più di quanto dovrebbero alla luce dei bassi livelli di qualità che riescono a erogare. Nove Regioni dovrebbero ridurre la spesa a doppia cifra, e queste appartengono quasi tutte al Sud e Isole, con il Lazio che a tutti gli effetti è accomunabile al gruppo in esame. Spiccano le correzioni di spesa che sarebbero richieste alla Campania, quasi il 32%, alla Sicilia, il 24,7%, e alla Puglia, il 23%. Il dato significativo è che, delle prime 8 Regioni per ampiezza della correzione di spesa necessaria, 6 erano al tempo coinvolte in piani di rientro per il contenimento e la stabilizzazione della spesa sanitaria: Campania, Sicilia, Lazio, Liguria, Abruzzo, Molise.

Escludendo i casi di Trentino Alto Adige e Liguria, che presentano una buona qualità ma con spesa elevata, il Centro-Nord fa registrare livelli di spesa non troppo lontani dall'efficienza, con correzioni di spesa sempre inferiori al 3%, e in alcuni casi inferiori al punto percentuale. All'opposto del Sud e Isole, il Friuli Venezia Giulia e l'Umbria potrebbero permettersi un aumento di un paio di punti percentuali, dal momento che sono outlier in positivo per come spendono e tutelano la qualità.

Una maggiore spesa quindi si accompagna a una minore qualità delle prestazioni. Alla luce dei risultati, se tutte le Regioni si fossero posizionate sulla frontiera efficiente nel biennio 2007-2008 si sarebbero risparmiate risorse pari a 1,5 punti percentuali di PIL.

3.1.2 SaniRegio2

Il modello SaniRegio2 (https://ideas.repec.org/p/ern/wpaper/02-2011.html), costruito su dati aggiornati ed alcune innovazioni di rilievo nella struttura e nell'interpretazione, ha avuto il pregio di evidenziare quanta parte della spesa sanitaria LEA possa trovare giustificazione, da un lato, nelle caratteristiche demografiche, economiche e sociali e nella dotazione di capitale fisico e umano e, dall'altro, nel livello qualitativo delle prestazioni e dei servizi erogati ai cittadini.

Nel passaggio tra SaniRegio1 e SaniRegio2, è stata rivista e ampliata la base dati in modo

che il database fosse più completo e coprisse un numero maggiore di anni; è inoltre cambiata la specificazione del modello in termini di variabili esplicative e sono state proposte quattro differenti specificazioni dello stesso.

L'obiettivo, come in SaniRegio1, era quello di ottenere una misura di sovraspesa a livello regionale, ovvero una quantificazione della spesa sanitaria regionale che non è spiegata dall'impatto delle variabili esplicative valorizzate secondo coefficienti omogenei per tutte le regioni. Diversamente da SaniRegio1, invece, sono state avanzate diverse ipotesi inglobate in quattro modelli alternativi. In particolare: nel modello 1 e nel modello 2 si è tenuto conto solo delle variabili dal lato della domanda: variabili socio-economiche, di struttura demografica e degli stili di vita. Nel modello 3 sono state prese in considerazione anche tre variabili dal lato dell'offerta ed il Pil pro capite che risulta una variabile molto importante nello spiegare la variabilità regionale della spesa; nel modello 4, infine, è stato escluso il Pil pro capite con un peggioramento della stima della varianza between spiegata. In conclusione gli scostamenti dalla spesa standard sono stati calcolati a partire dal modello 3.

Una ulteriore innovazione di rilievo, rispetto a SaniRegio1 ha riguardato la costruzione di un Indicatore Sintetico di Qualità (ISQ) relativo ad ogni sistema sanitario regionale. L'ISQ è costruito come media pesata di quattro indicatori riassuntivi: outcome del sistema sanitario, grado di soddisfazione del SSR da parte dei cittadini, mobilità interregionale e rischio di ospedalizzazione.

Dal confronto dei risultati ottenuti in SaniRegio2 rispetto a quelli riportati in SaniRegio1 ne emergono due principali:

- si radicalizza in qualche modo la territorializzazione dei gap di spesa e di qualità, con il Mezzogiorno che, per raggiungere l'Umbria, deve compiere, assieme a correzioni di spesa della stessa entità di quanto emergeva in SaniRegio1, anche poderosi innalzamenti della qualità;

- il ranking delle Regioni non cambia, ma viene meglio dettagliato. Un esempio viene dalle posizioni di Campania e Calabria. SaniRegio1 richiedeva una riduzione di spesa del 31,9% alla prima, e dell'11,1% alla seconda, ragionando a parità di qualità. Adesso SaniRegio2, abbinando sempre spesa e qualità nel vettore verso l'Umbria, chiede alla Campania di ridurre la spesa del 33,1% e di aumentare la qualità dell'88,7%, mentre chiede alla Calabria di ridurre la spesa del 15,1% e di aumentare la qualità di ben il 132,6%.

In conclusione, dal confronto tra SaniRegio1, SaniRegio2 e il benchmarking tra Regioni sulla base dei profili di spesa pro capite per fasce di età si giunse alle seguenti conclusioni che rappresentano il punto di partenza per la nuova analisi. I profondi gap di efficienza e di qualità tra Regioni hanno natura strutturale e trovano conferma anche attraverso l'utilizzo di metodologie di analisi diverse. Il Mezzogiorno è staccato dal resto d'Italia di più ordini di grandezza che dimostrano l'urgenza di cogenti riforme. Sono cinque le Regioni per le quali il gap di efficienza e di qualità risulta particolarmente acuto: Campania, Sicilia, Puglia,

Calabria e Lazio. Gli aggiustamenti devono avvenire sia per la spesa che per la qualità, e compiersi durante una transizione che, per essere credibile e irreversibile, deve essere circoscritta nel tempo.

3.1.3 SaniRegio2015

Il modello predisposto per SaniRegio2015, nonostante abbia presentato lo stesso obiettivo finale e utilizzato, in gran parte, la stessa base dati delle precedenti versioni, è risultato notevolmente ampliato nel ventaglio dei risultati e delle tecniche econometriche impiegate.
In SaniRegio2015, in particolare, accanto all'analisi della spesa, si sono considerate per la prima volta in modo esplicito e dettagliati i livelli delle prestazioni erogate dai sistemi sanitari regionali italiani con tre principali obiettivi: calcolare, in primo luogo, il livello di spesa efficiente di ogni sistema regionale al fine di definire i fabbisogni per l'intero sistema sanitario nazionale; misurare, successivamente, i livelli qualitativi e quantitativi delle prestazioni erogate in modo da poter fornire un supporto metodologico e informativo volto alla definizione dei livelli essenziali delle prestazioni che risultino in linea con i livelli di spesa efficiente precedentemente definiti; individuare, da ultimo, un sistema di benchmark regionali e nazionali definiti in modo endogeno sulla base dei risultati del modello, attraverso i quali stimolare un processo di efficientamento.

Il modello si è articolato lungo quattro fasi principali: la prima relativa al calcolo dell'efficienza tecnica, volta a ottenere un indice che misurasse la capacità di ogni Regione di utilizzare in modo efficiente i propri input; la seconda incentrata sul calcolo dei livelli quantitativi delle prestazioni, condotta attraverso la stima di una funzione di domanda dei servizi sanitari; la terza fase riguardante, invece, la determinazione dei fabbisogni standard, calcolati attraverso la stima di una funzione di spesa del servizio sanitario con lo scopo di mettere in relazione la spesa con le variabili di contesto (in primis la distribuzione della popolazione per profili di età), in linea con le attuali tecniche di standardizzazione; l'ultima fase, in conclusione, rivolta alla revisione della spesa, al fine di determinare il livello di spesa efficiente (regionale e nazionale) compatibile, sia con la quantità storica dei servizi sia con i livelli essenziali delle prestazioni.

I risultati finali hanno mostrato, tra il 1998 il 2002 un incremento dell'inefficienza tecnica che spiega, da sola, parte del forte incremento della spesa registrato in quegli anni. Dal 2002 in poi, invece, si è registrata una costante riduzione dell'inefficienza tecnica grazie a un trend costante nel livello dell'input lavoro per abitante, un trend decrescente nel livello delle dotazioni strumentali per abitante e un incremento dell'output per abitante, soprattutto a partire dal 2008, anno dal quale si consolidano sia i maggiori recuperi di efficienza, sia le prime riduzioni nel livello della spesa corrente in termini reali. Conseguentemente, gli incrementi di spesa osservati tra il 2002 e il 2006 possono essere attribuiti in modo prevalente al livello dei salari e dei prezzi d'acquisto dei beni e servizi intermedi a causa dalla dinamica dei rinnovi contrattuali superiore all'inflazione programmata e dalla mancanza di controllo

sugli ordinatori di spesa.

Dalla distribuzione territoriale dell'inefficienza tecnica è emerso, negli anni dell'analisi, un forte divaricamento in termini di efficienza tra le regioni del Centro-Nord e quelle del Sud confermando i risultati prodotti nelle precedenti versioni di SaniRegio. Mentre nel 1998 le regioni risultavano tutte molto vicine, nel corso degli anni si assiste ad una forte recupero di efficienza nelle regioni del Centro e del Nord a fronte di un sostanziale immobilismo delle regioni meridionali che sembrano offrire progressivamente meno servizi con un livello di inefficienza più alto rispetto al resto del Paese. Nel 2010 le regioni del Sud, incluse le Isole, presentano valori di inefficienza quasi il doppio più alti rispetto alle regioni del Centro e del Nord (incluse le regioni a statuto speciale).

L'evidenza empirica ha mostrato che le cause del mancato recupero di efficienza osservato nelle regioni meridionali era da ricercarsi non solo nei più bassi livelli qualitativi e quantitativi dei servizi offerti, ma anche nella non corretta proporzione tra gli input che, nelle regioni meridionali, risulta (in termini pro capite) sbilanciata nei confronti del capitale rispetto a quanto osservato nelle regioni più efficienti. Da ultimo, tra le cause del ritardo in termini di efficienza delle regioni del Sud un ruolo non secondario è da attribuirsi all'impatto generato dall'adozione dei piani di rientro che hanno interessato principalmente le regioni meridionali[1].

L'acuirsi, in termini di efficienza nella fornitura dei servizi, del divario tra il Sud e il Centro-Nord del Paese in concomitanza dell'introduzione dei piani di rientro è stata una evidenza empirica che, sia pur da verificare attraverso analisi più dettagliate, ha generato il sospetto che i piani di rientro abbiano migliorato i saldi di finanza pubblica agendo principalmente sul versante dei prezzi degli input senza apportare, però, miglioramenti nell'efficienza tecnica e soprattutto senza stimolare un livello delle prestazioni più alto in termini di quantità e qualità. Da qui nel rapporto si evidenziava la necessità di riformare parzialmente tale strumento di governance sviluppando sistemi di monitoraggio che ponessero maggiore enfasi sul livello dei servizi erogati a parità di input al fine di garantire che, a fronte di un miglioramento dei saldi finanziari, non registrando - al contempo - un deterioramento delle prestazioni.

3.2 OBIETTIVI E STRUTTURA DI SANIREGIO2016

Il presente rapporto SaniRegio2016 riprende le innovazioni e le fasi principali di ricerca proposte in SaniRegio2015 con il duplice obiettivo di *(i)* aggiornare le principali stime all'annualità 2014 e *(ii)* di proporre un primo tentativo di tradurre le stime di domanda a livello

1 Lazio, Abruzzo, Molise, Campania, Calabria e Sicilia hanno avviato il piano di rientro nel periodo 2006-2010 e sono ancora in vigenza di piano; Piemonte e Puglia sono, invece, soggette a un piano di rientro "leggero" a partire dall'anno 2010.

regionale in stime a livello territoriale più fine.

La base dati di riferimento, in linea con le versioni precedenti, è stata costruita come un panel regionale che ha incluso una serie di informazioni relative alle 20 regioni italiane[2] lungo un arco temporale di diciassette anni dal 1998 al 2014. La principale fonte è rappresentata dal database *Health for All* (HFA)[3] dell'ISTAT, mentre la seconda fonte è rappresentata dai rapporti redatti dal Ministero della Salute in merito alle Schede di Dimissione Ospedaliera (SDO).

Il database HFA, nella versione del 31 luglio 2016, contiene 4.000 indicatori a livello regionale, provinciale e di ripartizione dal 1980 al 2015. Il database ha il vantaggio di produrre in forma sinottica indicatori provenienti da fonti eterogenee (Ministero della salute, MEF, ISTAT, Lega Tumori, etc.) e che sono strutturati in modo da essere interrogati da un software standard HFA fornito dall'Organizzazione Mondiale della Sanità e adattato alle esigenze nazionali. Gli indicatori sono suddivisi in nove gruppi tematici: contesto socio-demografico, mortalità per causa, stili di vita, prevenzione, malattie croniche e infettive, disabilità e dipendenze, condizioni di salute e speranza di vita, assistenza sanitaria, attività ospedaliera per patologia, risorse sanitarie. Il database HFA è stata la fonte principale per la costruzione delle variabili di spesa, output (in termini quantitativi, ad esempio degenza e qualitativi come la mobilità, input (personale tecnico, personale medico, posti letto, macchinari) e variabili di contesto (struttura della popolazione, stili di vita, spesa privata).

Il rapporto SDO fornisce una fotografia delle attività di ricovero degli ospedali italiani, pubblici e privati, in modo affidabile e completo. La scheda di dimissione ospedaliera è lo strumento di raccolta delle informazioni relative ad ogni paziente dimesso dagli istituti di ricovero pubblici e privati in tutto il territorio nazionale. La banca dati SDO assicura una copertura della rilevazione nel 2015 superiore al 99% degli ospedali pubblici e privati accreditati. Essi rappresentano i principali strumenti operativi per monitorare l'appropriatezza delle prestazioni erogate e seguire l'obiettivo del Ministero della Salute di migliorare e uniformare ad alti standard qualitativi l'attività ospedaliera in tutto il territorio nazionale ed in particolare nelle regioni in piano di rientro. Il rapporto SDO fornisce la principale fonte per la costruzione delle variabili relative ai saldi della mobilità sanitaria inter-regionale.

L'impianto metodologico del modello proposto in questa versione di SaniRegio - similmente all'edizione SaniRegio2015 - si basa su quattro pilastri fondamentali:

- il primo pilastro è rappresentato dalle tecniche non parametriche per il calcolo dell'efficienza tecnica (si veda l'Appendice, capitolo 9, per una più esauriente trattazione), la cui adozione rappresenta da un lato un elemento di continuità rispetto alle versioni precedenti di SaniRegio, dall'altro lato rappresenta un elemento di innovazione grazie sia all'utilizzo delle più recenti tecniche di tipo non parametrico, sia per il focus esplicito sul calcolo dell'inefficienza tecnica, ovvero sulla capacità dei singoli sistemi sanitari

2 Le province autonome di Trento e Bolzano sono state fuse insieme ai fini dell'analisi.
3 http://www.istat.it/it/archivio/14562.

regionali di combinare nel miglior modo possibile gli input al fine di raggiungere il massimo livello di output;

- il secondo pilastro è rappresentato dalle tecniche di aggregazione delle variabili elementari di input e output in indicatori compositi al fine di ottenere delle variabili rappresentative della funzione di produzione (si veda l'Appendice, capitolo 9, per una più esauriente trattazione) consentendo di specificare in modo più corretto la funzione di produzione del servizio sanitario e allo stesso tempo di monitorare l'evoluzione dei livelli di input ed output di ogni Regione lungo un arco temporale quasi ventennale;
- il terzo pilastro è rappresentato dalla regressione lineare multipla applicata a modelli panel per la stima dei livelli di spesa standard (si vedano i paragrafi 4.2.1 e 4.2.2). Questo punto, nonostante rappresenti un importante elemento di continuità rispetto alle versioni SaniRegio1 e SaniRegio2, introduce allo stesso tempo vari elementi innovativi. In primo luogo la regressione multipla è utilizzata anche per stimare il livello standard delle prestazioni, in secondo luogo il modello della funzione di spesa è stato arricchito includendo tra i regressori sia il livello di output prodotto in più o in meno rispetto allo standard, sia il livello di input considerato inefficiente. In questo modo è possibile individuare due componenti importanti della spesa storica che non devono essere inclusi nella spesa standard. La prima è la quota di spesa relativa all'inefficienza tecnica, ovvero la percentuale di spesa riconducibile agli input in eccesso rispetto a quelli compatibili con una produzione efficiente degli attuali livelli di servizio; la seconda è la quota di spesa attribuibile ai livelli - quantitativi e qualitativi - di output realizzati in misura superiore o inferiore rispetto ad un valore di output standard definito attraverso la stima di una funzione di domanda. Queste due componenti di spesa, se correttamente individuate, consentono di affiancare al calcolo dei coefficienti di riparto delle risorse del comparto sanitario, una misura del livello ottimale dei fondi pubblici oggetto del riparto;
- l'ultimo pilastro - originale della presente versione - è l'utilizzo di tecniche di disaggregazione spaziali derivate dai lavori su serie storiche di Chow & Lin (1971) presentato nel capitolo 7; tali tecniche permettono di derivare alcune misure a livello territorialmente disaggregato dai corrispondenti indicatori di livello aggregato rendendo possibile - pur sotto opportune ipotesi - lo studio della domanda e offerta a livello di *small area*. A livello di *small area*, infatti, diversi sottoinsiemi di popolazione competono in modo più efficace per i servizi ed il loro diverso utilizzo relativo può meglio contribuire a spiegare meglio la variabilità territoriale dei fabbisogni sanitari. L'utilizzo di tali informazioni più dettagliate - se non ricavabili da fonte ufficiale -, dunque, può essere in ultima analisi molto più informativo del solo criterio di allocazione delle risorse dallo Stato alle Regioni fornendo una base comune per la costruzione di politiche programmatorie che poggino su analisi coerenti e condivise tra i diversi sistemi regionali.

4 L'IMPIANTO METODOLOGICO

Indice

4.1 Stima del livello delle prestazioni, del livello degli input e dell'efficienza tecnica . . . 35
4.1.1 Gli indicatori compositi del livello delle prestazioni offerte e degli input impiegati . . . 35
4.1.2 L'efficienza tecnica . . . 37
4.2 La funzione di domanda e la funzione di spesa . . . 37
4.2.1 La funzione di domanda . . . 38
4.2.2 La funzione di spesa . . . 40

L'impianto metodologico proposto si snoda attraverso quattro fasi distinte.
Il primo passo è il calcolo del livello aggregato delle prestazioni e degli input impiegati nel sistema sanitario di ogni regione in ogni anno oggetto di analisi. Successivamente, nella seconda fase, s'individua il livello di efficienza tecnica sulla base degli input e degli output impiegati. La terza fase è rivolta alla stima della funzione di domanda in forma ridotta che chiameremo anche *Funzione di output*. Più specificatamente, l'obiettivo di questa fase è calcolare il livello standard dei servizi che, una volta confrontati con il livello storico offerto, consente di individuare l'output-gap di ogni regione, ovvero una misura di come ogni sistema regionale risulti in grado di soddisfare la propria domanda. Alcune regioni producono servizi con un livello qualitativo e quantitativo superiore a quello standard e presentano, di conseguenza, un output-gap positivo altre, invece, sono caratterizzate da un output-gap negativo in quanto producono prestazioni inferiori rispetto a quelle compatibili con la domanda potenziale del proprio territorio. L'ultima fase della nostra analisi è la stima della funzione di costo in forma ridotta, che prende il nome di *Funzione di spesa* attraverso la quale si individua il fabbisogno standard di ogni regione. Tra i regressori della funzione di spesa verranno inclusi il livello di inefficienza tecnica, l'output-gap e gli effetti fissi del modello che verranno utilizzati per ottenere una stima dell'inefficienza relativa ai prezzi dei fattori produttivi. In questo modo è possibile calcolare la spesa standard di ogni regione isolando la quota di spesa storica inefficiente dalla quota di spesa necessaria a colmare il deficit di prestazioni.

4.1 STIMA DEL LIVELLO DELLE PRESTAZIONI, DEL LIVELLO DEGLI INPUT E DELL'EFFICIENZA TECNICA

4.1.1 Gli indicatori compositi del livello delle prestazioni offerte e degli input impiegati

L'approccio metodologico scelto per il calcolo del livello delle prestazioni offerte e del livello degli input impiegati è quella degli indicatori compositi (*CI's methods*), essendo al momento la tecnica più avanzata per trattare il caso di strutture produttive multidimensionali come quella del settore sanitario.

Gli indicatori compositi sono uno strumento di analisi e comunicazione sempre più usato nello studio di fenomeni economici e sociali in una pluralità di ambiti del settore pubblico e privato quali, ad esempio, la competitività industriale, lo sviluppo sostenibile, la qualità della vita, la globalizzazione e l'innovazione (Nardo *et al.*, 2005). Queste metodologie di aggregazione consentono di trovare una tendenza comune tra molti indicatori separati, come ad esempio molteplici misure di output, dimostrandosi particolarmente utili per l'analisi comparativa delle prestazioni tra paesi, regioni ed unità territoriali.

La definizione di indicatore composito da parte dell'OECD (Nardo *et al.*, 2005) sottolinea in modo evidente il rapporto con il fenomeno da misurare in quanto definisce un indicatore come un "*insieme di indicatori semplici aggregati in base di un modello che ricalca le caratteristiche multidimensionali del fenomeno che viene misurato*".

Se la precedente definizione può, in prima analisi, essere accettata, meno semplice sembra essere l'esplicitazione concreta di una serie di step individuati dall'OECD; il Joint Research Centre della Commissione Europea, infatti, afferma che *"no uniformly agreed methodology exists to weight individual indicators before aggregating them into a composite indicator"*[1].

Se una metodologia più affidabile delle altre non esiste, l'OECD (Nardo *et al.*, 2005) raccomanda, comunque, di seguire una pluralità di passi nel costruire l'indicatore in esame al fine di garantire una maggiore robustezza ai risultati ottenuti: in particolare, anche ai fini della presente ricerca, sembrano fondamentali i passi relativi all'*investigating the structure* degli indicatori semplici tramite l'utilizzo di tecniche statistiche multivariate, il trattamento dei *missing data*, il problema del riporto ad un'unica unità di misura (*normalization*) e per ultimo la scelta di un modello appropriato di *weighting and aggregation*.

Tra i dieci passi attraverso i quali l'OECD raccomanda di dividere il processo di analisi e costruzione di un Composite Indicator (CI), la fase di ponderazione e di aggregazione (*weighting and aggregation*) sembra in generale la fase più controversa, in quanto non esiste una metodologia universalmente accettata e riveste un'influenza notevole sul risultato finale, in particolare quando non sia prevista una priorità specifica tra indicatori.

Nel prosieguo dell'analisi il calcolo degli indicatori compositi di input e output sarà effettuato utilizzando la procedura Benefit of Doubt (BoD) considerata da più parti (si consideri, ad esempio, Sorensen, 2014 e Lauer *et al.*, 2004 per un'applicazione al campo sanitario) una

1 https://composite-indicators.jrc.ec.europa.eu/?q=content/step-6-weighting

delle tecniche di aggregazione più interessanti sviluppate nell'ultimo ventennio, soprattutto grazie alle proprietà intrinseche del metodo.
Diversamente da metodi di *weighting* basati su misure medie, in tale metodo, infatti, lo schema dei pesi relativo agli indicatori elementari è scelto endogeneamente[2] in modo tale che l'indicatore risultante sia il più alto possibile per ogni unità; tale proprietà è particolarmente "*useful in policy arena, since policy-makers could not complain about unfair weighting: any other weighting scheme would have generated lower composite scores*" (Nardo *et al.*, 2005).
Più specificatamente il CI di tipo BoD soddisfa le seguenti proprietà desiderate: i pesi sono determinati **endogeneamente** attraverso le performance osservate e il benchmark non è basato su vincoli o scelte teoriche, ma su combinazioni lineari di performance osservate; il CI è **monotono debole** e **invariante** rispetto ad un ridimensionamento proporzionale degli indicatori semplici ed infine, come affermato in precedenza, lo schema dei pesi risulta per la singola unità il **più alto possibile**.
Tale metodo di stima consente, quindi, di includere una pluralità di indicatori di servizio facendo sì che nessun indicatore possa da solo o in modo preponderante catturare la complessità del fenomeno da studiare permettendo una migliore descrizione dei molteplici aspetti relativi alla fornitura di un servizio.
Infine l'aver adottato un metodo che fornisce una misura composita tramite metodi di benchmarking ci permetterà non solo di avere a disposizione una misurazione composita, ma fornirà ad ogni regione target di riferimento basati su performance reali (e non teoriche) per quanto riguarda il livello del servizio erogato.
Per i dettagli tecnici relativi alla metodologia BoD si rimanda alla sezione 9 dell'Appendice.

In conclusione è importate richiamare l'attenzione brevemente sui pro ed i contro comuni a qualsiasi esercizio di stima di un indicatore composito. Tra gli aspetti positivi si elencano: la possibilità di riassumere questioni complesse o multi-dimensionali con lo scopo di supportare i decisori politici, la formazione di una *big picture* del fenomeno analizzato, la possibilità di attirare l'interesse pubblico, fornendo un'indicazione di sintesi con cui confrontare le prestazioni di diverse realtà territoriali ed i loro progressi nel tempo, da ultimo la capacità di ridurre le dimensioni di una pluralità di indicatori ad un unica dimensione. Invece, tra gli aspetti negativi da tenere in debita considerazione rientrano: (i) il fatto che il processo di aggregazione possa fornire conclusioni di policy non robuste soprattutto se mal costruito e/o mal interpretato[3], (ii) i semplici risultati relativi alla *big picture* potrebbero portare a formulare policy semplicistiche e da ultimo (iii) che il risultato finale possa essere influenzato in misura rilevante dalla scelta del modello di aggregazione e dei relativi pesi oltre che dal trattamento dei dati anomali o mancanti.

2 Ovvero non in modo arbitrario dato che non vi è alcun motivo ex-ante per assegnare a ciascun indicatore lo stesso peso o più in generale ad ogni unità lo stesso peso.
3 Si pensi, ad esempio, alla difficoltà di includere informazioni affidabili riguardanti i livelli qualitativi dei servizi forniti.

4.1.2 L'efficienza tecnica

Gli indicatori compositi, calcolati con le metodologie descritte in precedenza, relativi all'output prodotto e agli input impiegati dai governi regionali sono utilizzati per misurare, in un secondo step, il livello di efficienza tecnica raggiunto da ogni regione nell'offerta dei servizi sanitari. L'inefficienza tecnica è qui intesa come misura di Farrell (1957), ovvero come rapporto tra l'output osservato e l'output massimo nell'ipotesi di input fissi, o, in alternativa, come rapporto tra l'input osservato e l'input minimo sotto l'ipotesi di output fissi.

Da un punto di vista applicativo, per efficienza tecnica, infatti, si intende:

- nello spazio degli output, data la tecnologia e gli input, la corrispondenza tra gli output effettivamente prodotti e gli output massimi potenzialmente producibili (*output-approach*);
- nello spazio degli input, data la tecnologia e il livello degli output, la corrispondenza tra la quantità di input utilizzato e la quantità minima potenzialmente utilizzabile (*input-approach*).

Nell'analisi sviluppata in SaniRegio2016, dovendo raggiungere obiettivi di revisione della spesa, l'inefficienza sarà misurata dal lato degli input. In linea, quindi, con la strategia metodologica di tipo *input-approach* l'enfasi sarà posta sul livello di comprimibilità degli input a parità di output prodotti.

Dal punto di vista applicativo, infine, esistono due filoni principali per misurare l'efficienza: (i) gli approcci non parametrici che non formulano ipotesi sulla forma funzionale della frontiera produttiva e (ii) gli approcci parametrici che assumono invece che la frontiera segua una determinata classe funzionale particolare (Cobb-Douglas, CES (Constant Elasticity of Substitution), translogaritmica, etc.) e mirano all'individuazione di una specifica frontiera di produzione mediante la stima dei parametri che la caratterizzano.

Nel prosieguo della nostra analisi e nell'applicazione pratica, pur consci dei limiti teorici descritti più accuratamente nel capitolo 9, verranno utilizzati modelli non parametrici di tipo *"Data Envelopment Analysis"* (Data Envelopment Analysis (DEA)) per le minori ipotesi alla base del modello e per la maggior flessibilità in fase di stima rispetto a quanto richiesto da un approccio di tipo parametrico.

4.2 LA FUNZIONE DI DOMANDA E LA FUNZIONE DI SPESA

Il modello utilizzato in SaniRegio2016 è incentrato sulla stima della funzione di domanda e di offerta dei servizi sanitari riportate, rispettivamente, nelle equazioni (1) e (2) che seguono.

$$q = d(R, D, c) \tag{1}$$

$$c = s(P, S, q) \tag{2}$$

Dove: q = servizi offerti; c = costo unitario dei servizi; R = reddito (PIL); D = determinanti della domanda (ad esempio la struttura della popolazione); P = prezzi degli input; S = variabili ambientali dell'offerta (ad esempio la spesa delle famiglie).

4.2.1 La funzione di domanda

Al fine di semplificare il processo di stima si procede, inizialmente, alla definizione della funzione di domanda in forma ridotta ottenuta sostituendo l'equazione (2) nella (1). La funzione di domanda in forma ridotta corrisponde, quindi, al modello riportato di seguito che chiameremo funzione di output:

$$q = h(R, D, P, S) \tag{3}$$

Il passaggio dal modello teorico al modello empirico della funzione di output richiede, in primo luogo, la definizione della variabile dipendente q che deve racchiude al suo interno il livello dei servizi offerti dal sistema sanitario di ogni regione. A tal fine, per ogni Regione i ed anno t, è stato calcolato un indicatore composito CI_{it} dei servizi offerti in modo da cogliere il carattere multi-output della funzione di produzione tipica del settore sanitario.

Successivamente, sfruttando la struttura panel del dataset, il modello empirico di riferimento corrisponde a un modello panel lineare ad effetti fissi come riportato nell'equazione (4) che segue:

$$CI_{it} = \alpha_i + \eta_t + \beta_1 R_{it} + \beta_2 D_{it} + \beta_3 S_{it} + \epsilon_{it} \tag{4}$$

dove: i = indice di regione; t = indice degli anni; CI_{it} = indicatore composito di output; β = coefficienti; α_i = effetti fissi regionali; η_t = effetti fissi annuali; ϵ_{it} = errore idiosincratico. E' importante sottolineare che, in assenza di informazioni dettagliate relative ai prezzi degli input (costo del lavoro e costo del capitale) il loro impatto sulla spesa è approssimato dagli effetti fissi regionali.

Il modello riportato nell'equazione (4) è poi modificato inserendo tra i regressori una serie di variabili strutturali. In primo luogo la variabile M_{it} che cattura i flussi di mobilità in termini di saldo netto tra i pazienti in entrata ed uscita tra le regioni, in modo da misurare quanto output di un sistema sanitario regionale sia rivolto a soddisfare la domanda interna e quanto, invece, sia rivolto a soddisfare la domanda esterna, ovvero proveniente da pazienti delle altre regioni. Questo indicatore può anche essere interpretato, pur in maniera indiretta, quale proxy della qualità del servizio sanitario di una regione. In secondo luogo sono state inserite le dummy z_{it} volte ad identificare l'effetto medio dei piani di rientro sui livelli dei

servizi offerti delle regioni che li hanno adottati dal 2006 al 2010 (ultimo anno del nostro campione di regressione)[4].

Il modello finale della funzione di output, quindi, diventa quello riportato di seguito nell'equazione (5).

$$CI_{it} = \alpha_i + \eta_t + \beta_0 M_{it} + \sum_{t=2006}^{2010} \beta_t z_{it} + \beta_1 R_{it} + \beta_2 D_{it} + \beta_3 S_{it} + \epsilon_{it} \tag{5}$$

Dopo aver stimato il modello (5) con lo stimatore *Within-the-Group* (WG), i valori attesi $\hat{C}I_{it}$ sono stati definiti come *Output standard* di ogni regione. Più specificatamente, nel calcolo dell'output standard - equazione (6) - si considera come regione benchmark quella con l'effetto fisso più alto, ovvero quella che a parità di variabili di contesto produce il maggior livello di prestazioni per abitante. Inoltre, nel computo di tale valore atteso si esclude l'effetto esercitato dalla mobilità e dai piani di rientro.

$$\hat{C}I_{i,t} = E[CI_{i,t}|\alpha_{max}, \eta_t, R_{i,t}, D_{i,t}, S_{i,t}] \tag{6}$$

In seconda battuta si procede al calcolo dell'*Output-gap* (w_i) complessivo della regione i-esima nei vari anni dell'analisi, corrispondente alla differenza tra l'output storico CI e l'output standard come riportato nell'equazione seguente:

$$w_i = CI_{i,t} - \hat{C}I_{i,t} \tag{7}$$

Da ultimo dalla stima dell'output gap è importate isolare la quota relativa alla mobilità dei pazienti tra le regioni (w_{it_mob}). In questo modo, per le regioni caratterizzate da un saldo di mobilità positivo, è possibile individuare in modo più preciso l'output gap "proprio" ($w_{it_proprio}$) ovvero quello effettivamente riferito al soddisfacimento della domanda dei propri residenti. Il calcolo delle suddette quantità viene riportato nelle equazioni (8) e (9) che seguono.

$$w_{it_mob} = E[CI_{i,t}|M_{i,t}] \tag{8}$$

$$w_{it_proprio} = CI_{i,t} - \hat{C}I_{i,t} - w_{it_mob} \tag{9}$$

4 Le regioni interessate dai piani di rientro nel periodo di analisi sono: Lazio, Abruzzo, Molise, Campania, Calabria e Sicilia.

4.2.2 La funzione di spesa

In modo similare al caso della funzione di domanda, al fine di semplificare il processo di stima, si procede alla definizione della funzione di costo dei servizi sanitari in forma ridotta sostituendo l'equazione (1) nella (2); la funzione di costo in forma ridotta, corrispondente al modello (10) riportato di seguito, prende il nome in letteratura di *Funzione di spesa*.

$$c = f(P, S, R, D) \tag{10}$$

La struttura longitudinale del nostro dataset fa si che il modello empirico di riferimento corrisponda ad un modello panel lineare ad effetti fissi come riportato nell'equazione (11) che segue:

$$H_{it} = \phi_i + \tau_t + \gamma_1 S_{it} + \gamma_2 R_{it} + \gamma_3 D_{it} + u_{it} \tag{11}$$

Dove: i = indice di regione; t = indice degli anni; H_{it} = spesa sanitaria corrente; γ = coefficienti; ϕ_i = effetti fissi regionali; τ_t = effetti fissi annuali e u_{it} = errore idiosincratico. Anche in questo caso, come nella funzione di domanda, l'assenza di informazioni dettagliate relative ai prezzi degli input (costo del lavoro e costo del capitale) ci impone di approssimarne l'impatto relativo sulla spesa attraverso gli effetti fissi regionali.

Sino a questo punto, la funzione di spesa specificata nel modello (11) risulta molto simile a quella stimata nelle versioni precedenti di SaniRegio. Come discusso già nelle pagine precedenti, una delle principali novità introdotte in SaniRegio2015 e poi in SaniRegio2016 è l'inclusione tra i regressori della funzione di spesa di due nuove variabili: θ_{it} corrispondente all'indice di inefficienza tecnica stimata in precedenza attraverso tecniche DEA e w_{it} corrispondente all'output-gap stimato dalla funzione di domanda. Conseguentemente il modello finale della funzione di spesa diventa quello specificato nella equazione (12) nel quale vengono aggiunte le dummy z_{it} incluse nella funzione di domanda per misurare l'apporto dei piani di rientro sulla spesa.

$$H_{it} = \phi_i + \tau_t + \delta_1 \theta_{it} + \delta_2 w_{it} + \sum_{t=2006}^{2010} \gamma_t z_{it} + \gamma_1 S_{it} + \gamma_2 R_{it} + \gamma_3 D_{it} + \psi_{it} \tag{12}$$

Il primo obiettivo della nostra analisi è quello di ottenere delle stime consistenti e non distorte dei seguenti coefficienti: δ_1 che ci consente di misurare la quota di spesa assorbita dall'inefficienza tecnica; δ_2 attraverso cui è possibile identificare l'ammontare di spesa necessario a colmare il deficit di prestazioni osservato nelle diverse regioni; γ_t che misura l'efficacia dei piani di rientro quali misure di contenimento della spesa a parità del contesto socio-economico e di servizi offerti.

Successivamente, attraverso la stima degli effetti fissi regionali ϕ_i è possibile misurare l'inefficienza di prezzo di ogni sistema regionale calcolando la distanza tra gli effetti fissi di ogni regione rispetto al valore minimo, di conseguenza la regione che presenterà l'effetto fisso più piccolo verrà considerata il benchmark di riferimento.

Il valore complessivo dell'inefficienza I_{it} di ogni sistema sanitario regionale avrà, dunque, come riportato nell'equazione (13), due componenti: la prima componente (θ_{it}) misura il livello di inefficienza tecnica, a partire dalla quale è possibile capire se la regione presenti margini di risparmio possibili a seguito di un miglior utilizzo degli input; la seconda componente $[\hat{\phi}_i - min(\hat{\phi}_i)]$ misura l'inefficienza di prezzo, ovvero la possibilità di impiegare uno stesso quantitativo di fattori produttivi ad un costo unitario più basso. E' importante notare che la stima dell'inefficienza tecnica è per costruzione variabile nel tempo, mentre la stima dell'inefficienza di prezzo no; per tale motivo l'inefficienza di prezzo deve essere interpretata come una media riferita all'intero periodo di analisi.

$$I_{it} = \hat{\delta_1}\theta_{it} + [\hat{\phi}_i - \hat{\phi}_{min})] \tag{13}$$

L'ultima fase della nostra analisi è il calcolo della spesa standard che può essere ottenuta secondo due specificazioni: la prima depurando la spesa storica di riferimento soltanto dall'inefficienza globale I_{it}. Questa prima misura della spesa standard, indicata con il simbolo $\hat{H}^a_{it}$, non tiene conto delle diversità dei livelli di servizi offerti lungo il territorio riconoscendo ad ogni regione una spesa standard compatibile con la quantità di servizi storicamente offerta.

La seconda specificazione prendendo in considerazione, oltre all'inefficienza, anche la quota di spesa necessaria a colmare l'output gap ($\hat{\delta_2}w_{it}$); si otterrà, quindi, una seconda misura della spesa standard che, contraddistinta dal simbolo $\hat{H}^b_{it}$, attribuisce ad ogni sistema regionale un fabbisogno di spesa tale da poter offrire ad un costo efficiente il massimo livello dei servizi necessari a soddisfare appieno la domanda del territorio. Le due versioni di spesa standard sono riportate di seguito rispettivamente nelle equazioni (14) e (15).

$$\hat{H}^a_{it} = E[H_{i,t}|\phi_{min}, \delta_2 w_{it}, \tau_t, R_{i,t}, D_{i,t}, S_{i,t}] \tag{14}$$

$$\hat{H}^b_{it} = E[H_{i,t}|\phi_{min}, \tau_t, R_{i,t}, D_{i,t}, S_{i,t}] \tag{15}$$

5 L'EFFICIENZA TECNICA PRODUTTIVA PER LE AMMINISTRAZIONI REGIONALI

Indice

5.1 Il livello delle prestazioni sanitarie regionali 42
5.1.1 Costruzione dell'indicatore composito di output 45
5.2 Analisi degli input del servizio 45
5.2.1 Costruzione dell'indicatore composito di input 53
5.3 Stima dell'efficienza tecnica produttiva 60
5.4 Un esercizio di simulazione per gli anni 2013-2015 62

5.1 IL LIVELLO DELLE PRESTAZIONI SANITARIE REGIONALI

Scopo di tale sezione è la stima del livello delle prestazioni sanitarie regionali in termini di output prodotti negli anni 1998-2012 dalle singole Regioni italiane; tale analisi è fortemente vincolata ai dati di base a disposizione, alla qualità e l'uniformità degli stessi in tutto il territorio nazionale.

Ancora più importante ci sembra sottolineare la distinzione, soprattutto in tale ambito, tra output ed outcome del servizio sanitario, ovvero definire in modo netto ciò che valutiamo e ciò che non valutiamo tra beni e servizi forniti in termini quantitativi (output) e fine ultimo del servizio pubblico; il focus di tale analisi, essendo propedeutica alla stima dell'efficienza tecnica produttiva delle Regioni, verterà, quindi, sugli aspetti quantitativi ovvero sul numero di degenze o sulle giornate di cura piuttosto che sugli esiti quali i tassi di mortalità o natimortalità.

Più specificatamente l'indicatore composito di output proposto vuole essere più precisamente un indicatore della "quantità" del servizio erogato; non vuole essere un indicatore di appropriatezza del servizio o legato agli esiti delle cure offerte. Questo essenzialmente per tre ordini di motivi: *(i)* l'introduzione di funzioni di valore o di appropriatezza per i singoli indicatori avrebbe introdotto un'inaccettabile rischio di distorsione dei risultati finali quando invece il principale vantaggio del metodo di weighting proposto si basa sull'essere il più oggettivo possibile; *(ii)* la definizione di appropriatezza delle cure non può essere introdotta dal ricercatore, ma dovrebbe primariamente essere definita dal legislatore in base a criteri medici e/o epidemiologici essenzialmente per non confondere il livello storico con il livello desiderato (o ottimale a seconda del criterio di valore scelto); *(iii)* l'indicatore di

output all'interno di un modello di spesa e/o di costo deve riflettere le "quantità" di output prodotte indipendentemente da altri criteri (appropriatezza, sostenibilità o rilevanza sociale) che devono muovere altri tipi di misurazione.

Il primo passo nell'analisi dei livelli quantitativi delle prestazioni sanitarie regionali è una verifica puntuale della qualità del dato relativo all'output prodotto soprattutto in termini dinamici; era importante cioè verificare che nei dati di base non ci fossero carenze informative in un determinato anno né che ci fossero forti dissonanze tra anni vicini.
Sono stati dunque analizzati gli andamenti di ogni singola variabile di analisi, verificando che la qualità del dato di base (non erano presenti dati mancanti), utilizzato per la stima dell'output ed in generale per la fase di stima della funzione di produzione, fosse molto alta[1].

Il secondo passo ha riguardato la selezione delle variabili di output. Dovendo calcolare un livello composito del servizio offerto, era necessario non includere più volte la stessa informazione di base ovvero verificare le correlazioni tra le variabili di output; avendo ottenuto un'alta correlazione tra i differenti aspetti scelti per la stima del livello composito, si è reso necessario calcolare dei fattori indipendenti ed incorrelati tramite un'analisi fattoriale quale l'Analisi in Componenti Principali (ACP).
Essendo tale analisi fortemente dipendente dall'ordine di grandezza delle variabili elementari, le variabili di output scelte sono state normalizzate per il numero dei residenti nella Regione per l'anno di riferimento ed è stata verificata la polarità di ogni variabile[2].

La tabella 3 riporta i fattori principali ortogonali ed indipendenti[3] per le variabili elementari di output, riuscendo a spiegare con i primi tre autovalori il 77% della varianza totale, permettendo, abbastanza agevolmente, di spiegare i fattori alla luce delle variabili più correlate con essi: il fattore 1 risulta essere il fattore dimensionale, legato cioè sia alle giornate di degenza sia al numero di degenze indipendentemente dalla tipologia; il fattore 2 sembra essere più legato a fattori di qualità del servizio essendo maggiormente correlato con la mobilità interregionale e l'assistenza extra ospedaliera, mentre il terzo fattore è legato maggiormente alla lungodegenza ed alla riabilitazione.

1 Solamente in due casi si era in presenza di un dato anomalo o fuori scala e si è deciso, in un caso di sostituire il dato interno alla serie con la media dell'anno precedente e successivo, mentre nell'altro, essendo relativo al primo anno e non avendo un trend con il quale confrontarlo, il dato non è stato eliminato né corretto. Più specificatamente si è corretto il dato del Molise, anno 2004, relativo alle camere iperbariche, ai tavoli operatori extra ospedalieri ed agli apparecchi per anestesia extraospedaliera con le relative medie per gli anni 2003 e 2005, mentre non è stato corretto il dato della Campania per l'anno 1998 relativo alla degenza media lungodegenza e riabilitazione.

2 Sembra un'operazione semplice attribuire una polarità ad ogni indicatore scelto rispetto all'obiettivo che si sta misurando, ma ci sono casi in cui la scelta, se non si esplicitano bene i criteri di valutazione, non risulta di facile applicazione: si pensi ad esempio al numero dei ricoveri ospedalieri; in un'ottica di miglioramento della qualità del servizio verso il paziente che favorisca cure domiciliari tale variabile dovrebbe avere polarità negativa ovvero premiare le Regioni che si pongono target decrescenti negli anni, mentre dovendo valutare esclusivamente il carico di lavoro - come nel nostro caso - tale indicatore deve essere posto con polarità positiva.

3 I valori stampati sono moltiplicati per 100 e arrotondati all'intero più vicino. I valori maggiori di 60 sono contrassegnati da un '*'. I valori minori di 30 non sono stampati; Metodo rotazione: Varimax con normalizzazione di Kaiser.

Tabella 3: Analisi in componenti principali - fattori di output (pro capite), anni 1998-2012

Variabili Elementari	Fattore1		Fattore2		Fattore3	
DEGENZE_TOT	97	*	.		.	
DEGENZE_ACUTI	96	*	.		.	
DEGENZE_PRIV	95	*	.		.	
GIORNATE_DEG_ACUTI	95	*	.		.	
GIORNATE_DEG	94	*	.		.	
DEGENZE_PUB_ACUTE	94	*	.		.	
DEGENZE_PUB	94	*	.		.	
DEVENZE_PRIV_ACUTE	93	*	.		.	
GIORNATE_DEG_ACUTI_PRIV	92	*	.		.	
GIORNATE_DEG_ACUTI_PUB	92	*	.		.	
GIORNATE_DEG_PUB	91	*	.		.	
GIORNATE_DEG_PRIV	89	*	.		.	
DEGENZE_RIABILITZ_PRIV	82	*	37		.	
GIORNATE_DEG_RIABILITAZ	80	*	42		.	
DEGENZE_RIABILITZ	80	*	51		.	
GIORNATE_DEG_RIABIL_PRIV	72	*	.		38	
GIORNATE_DEG_RIABILITAZ_PUB	72	*	58		.	
DEGENZE_RIABILITZ_PUB	71	*	58		.	
MOBILITA_PRC	.		71	*	.	
MOBILITA2_PRC	.		71	*	.	
ASSIST_PER_MEDICO	.		53		.	
ASSIST_PEDIATRA	.		52		.	
TASSPOSTI_LETTOSP	.		46		.	
DEGENZA_MEDIA_RIABILITAZ	.		.		91	*
DEGENZA_MEDIA_RIABIL_PRIV	.		.		82	*
DEGENZA_MEDIA_RIABIL_PUB	.		.		74	*
DEGENZA_MEDIA_PRIV	.		50		62	*
DEGENZA_MEDIA_ACUTI_PRIV	.		45		58	

Le figure 4, 5 ed 6, infine, permettono di valutare gli andamenti regionali negli anni per singolo fattore: si noti in particolare come l'andamento del fattore 1 risulti molto stabile negli anni, mentre il secondo fattore presenti una dinamica più accentuata nella quasi totalità delle Regioni. Il fattore 3 sembra più legato a specifiche scelte regionali.

Da ultimo si è verificata che la correlazione tra i fattori e la spesa corrente fosse positiva.

5.1.1 Costruzione dell'indicatore composito di output

Al fine di ottenere un'unica misura del livello del servizio offerto negli anni dalle singole Regioni, si è passati alla fase di costruzione dell'indicatore composito di output CI_{out} con tecniche di frontiera (BoD) robuste, come presentate in Appendice - capitolo 9, ponendo una particolare attenzione a verificare la robustezza territoriale dei differenti ranking ottenuti.

Tale indicatore composito dell'output è costruito[4] a partire dai soli fattori 1 e 2; il fattore 3 non è stato considerato in questa fase perché troppo legato a dinamiche interne alle Regioni stesse, ma è stato utilizzato successivamente per testare le differenze in termini di funzione di spesa.

La figura 7 riporta la distribuzione dell'indicatore composito di output sul territorio nazionale per l'anno 2012; ricordando che per costruzione tale indice è rapportato alla popolazione residente, si osserva un livello delle prestazioni molto più elevato nelle regioni settentrionali rispetto alle regioni centro-meridionali con un valore massimo in Lombardia.

http://www.cermlab.it/saniregio2016/mappa_ci_out.html

La figura 8 permette, invece, di apprezzare per singola Regione la dinamica di tale indicatore robusto negli anni (dal 1998 al 2012); si osserva nella maggior parte delle regioni, seppur con dinamiche differenti, un incremento medio del livello del servizio. Le regioni con il livello delle prestazioni più elevato e con incremento costante risultano essere Lombardia, Veneto, Emilia Romagna e Piemonte.

Le dinamiche, evidenziate nelle figure precedenti, sono state verificate tramite un'analisi di sensitività, ovvero facendo variare il modello di weighting; i risultati ottenuti si sono dimostrati essere molto robusti anche al variare delle ipotesi alla base del criterio di aggregazione, come mostrato in figura 9.

5.2 ANALISI DEGLI INPUT DEL SERVIZIO

Una volta analizzato il livello del servizio sanitario offerto nelle Regioni si è passato a stimare, con un processo logico simile, anche la parte relativa agli input alla base di tale processo

4 Per una maggior chiarezza riguardo le variabili utilizzate si faccia riferimento alla tabella 14 riportata in Appendice.

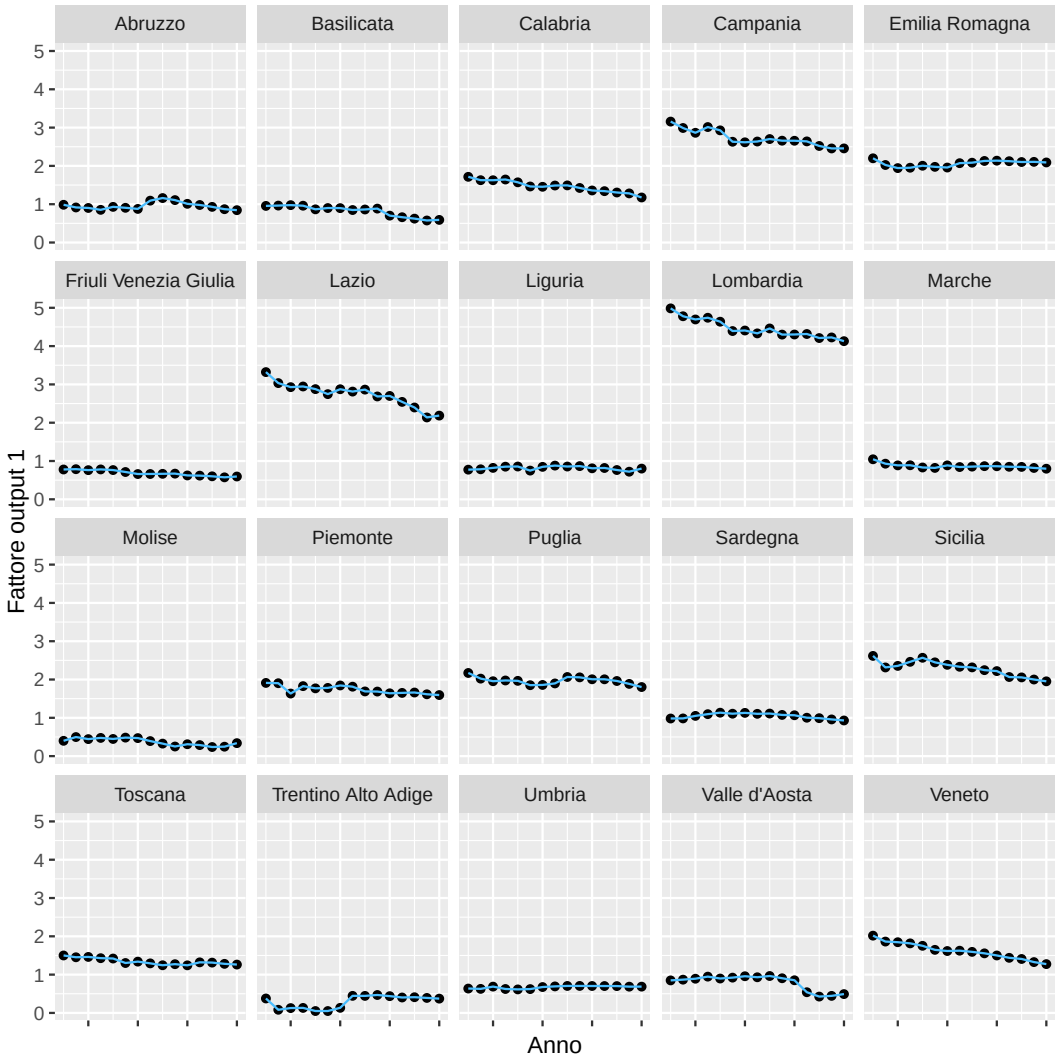


Figura 4: Dinamica del fattore di output 1 (dimensionale) per Regione ed anno, anni 1998-2012

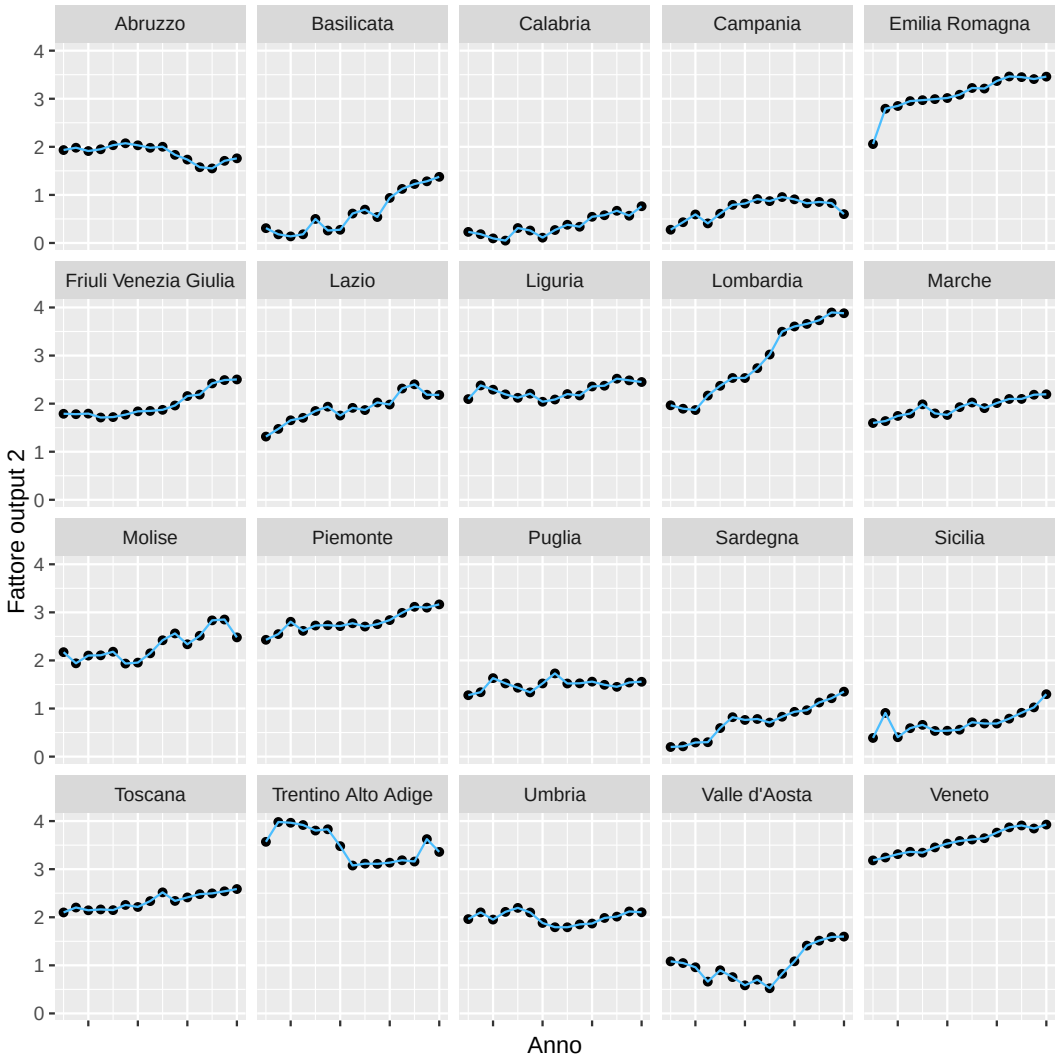


Figura 5: Dinamica del fattore di output 2 (qualità) per Regione ed anno, anni 1998-2012

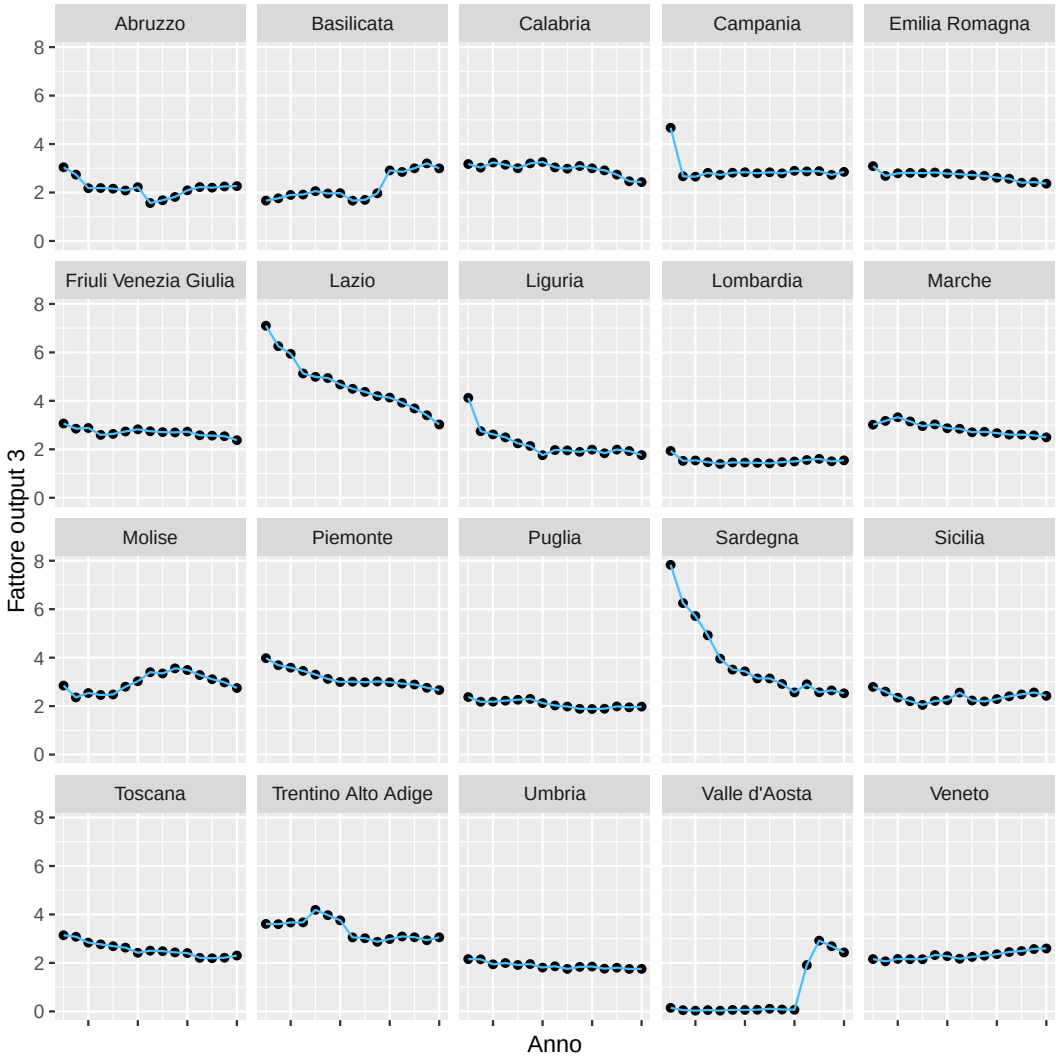


Figura 6: Dinamica del fattore di output 3 (lungodegenza e riabilitazione) per Regione ed anno, anni 1998-2012

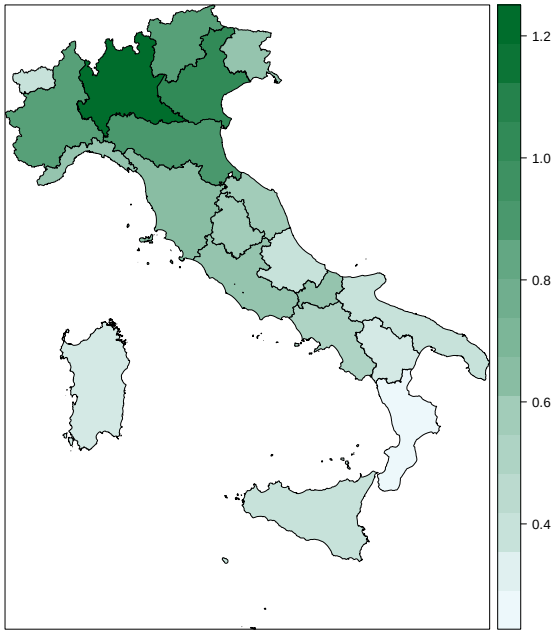


Figura 7: Distribuzione dell'indicatore composito robusto di output per Regione - anno 2012

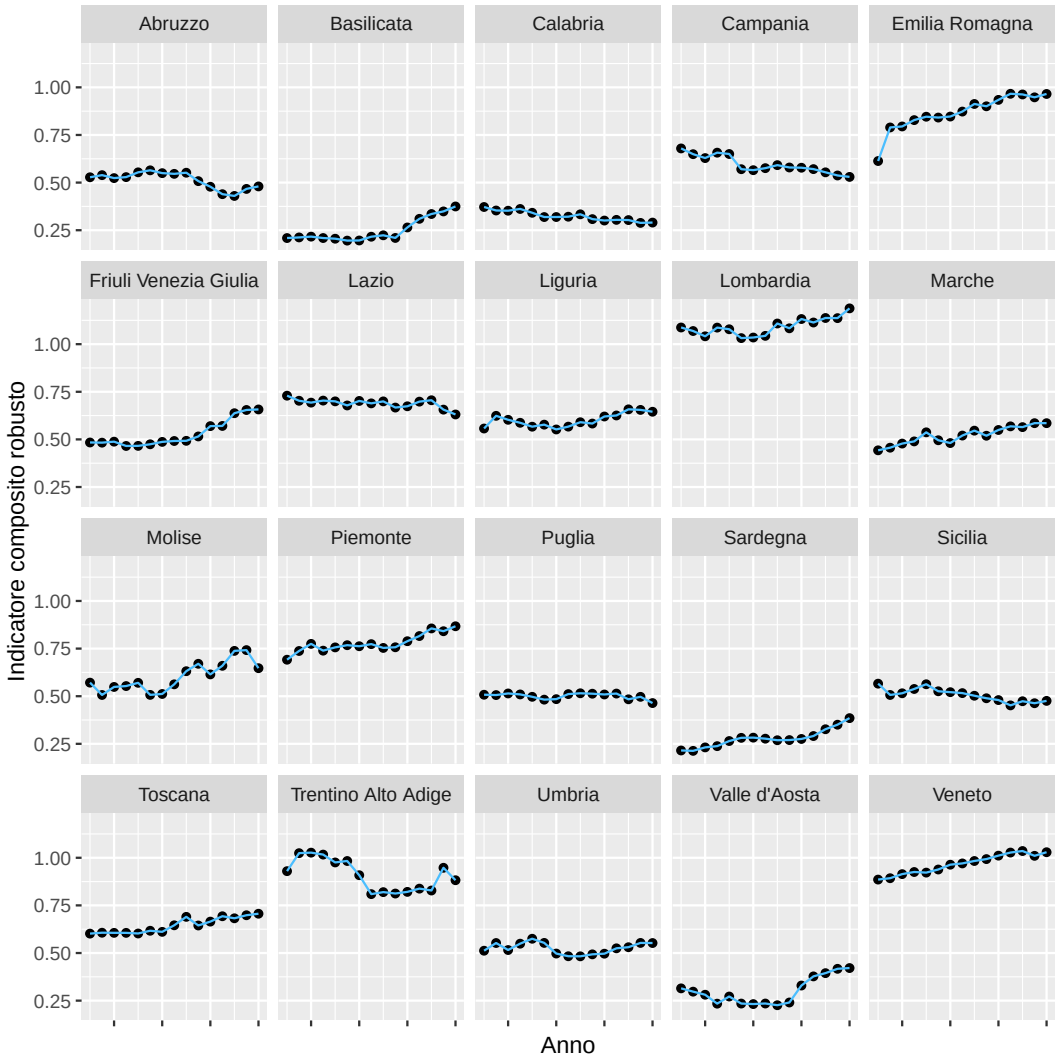


Figura 8: Dinamica dell'indicatore composito robusto di output per Regione e per anno, anni 1998-2012

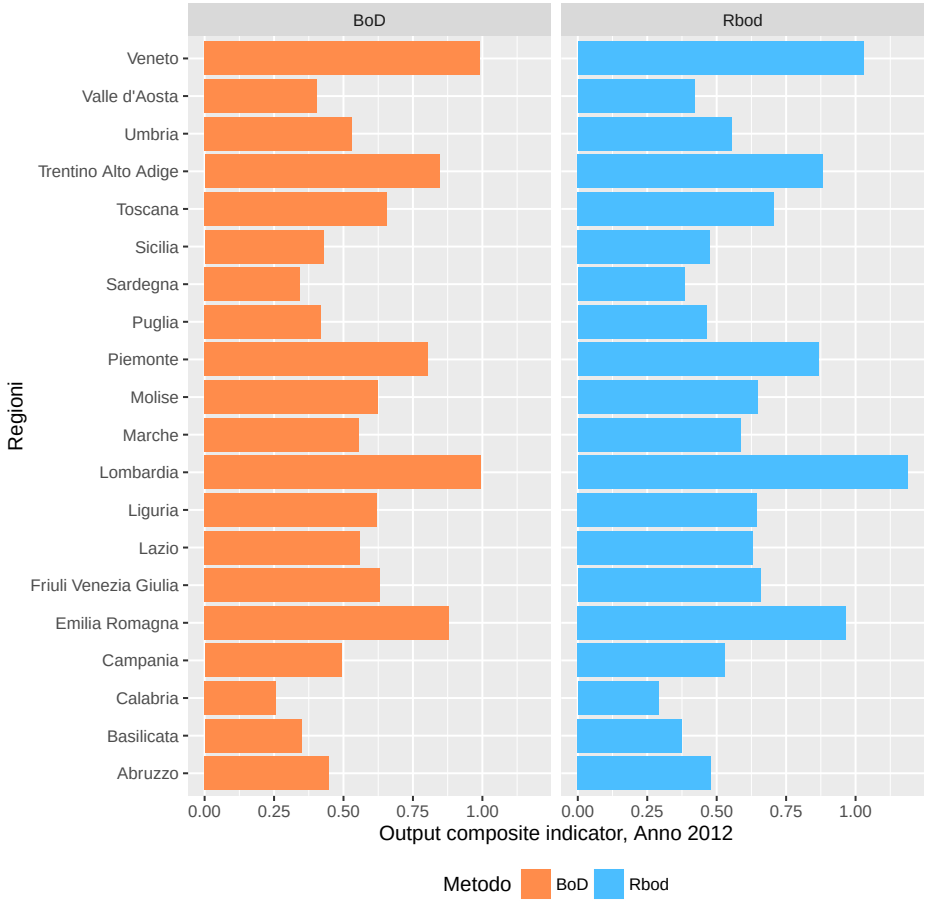


Figura 9: Distribuzione dell'indicatore composito robusto di output per Regione al variare del metodo (BoD e RBoD) - anno 2012

produttivo; in tabella 4 vengono riportate le stime relative ai fattori principali[5] degli input.

È prioritario comunque fare alcune considerazioni sui beni strumentali scelti per l'analisi: come evidenziato in molte analisi i beni strumentali, specie di tipo ospedalieri o comunque specialistico, non sono facilmente comparabili sia in termini qualitativi sia in termini di sostituibilità tra beni nuovi e vecchi (un ospedale ad esempio può non impiegare un bene più vecchio perché ne utilizza uno più moderno). Sarebbe quindi auspicabile poter introdurre nell'analisi un sottoinsieme di dotazione di beni industriali comparabili e condivisi tra gli attori regionali che da un punto di vista "industriale" individuino univocamente gli ospedali con dotazione maggiore; tale approfondimento esula dalla finalità della presente analisi, ma può rappresentare un utile strumento per migliorare l'affidabilità delle stime ottenute.

Per le finalità della presente analisi, si sono scelti beni strumentali disponibili in tutte le Regioni essenzialmente di tipo ospedaliero.

Tabella 4: Analisi in componenti principali - fattori di input (pro capite), anni 1998-2012

Variabili elementari	Fattore 1		Fattore 2		Fattore 3		Fattore 4	
TECNICI_DIP_PRIV	88	*	.		.		.	
INFERMIERI_DIP_PUB	86	*	.		.		.	
TECNICI_DIP_PUB	86	*	.		.		.	
INFERMIERI_DIP_PRIV	84	*	.		.		.	
RIABILITAZ_DIP_PUB	81	*	.		-42		.	
RIABILITAZ_DIP_PRIV	74	*	.		-39		.	
INFERMIERI_DIP_SSN	72	*	.		.		.	
TAC	.		77	*	.		.	
CONTAGLOBULI	.		75	*	.		.	
ANNESTESIA_EXTRA	.		70	*	.		.	
TAVOLI_RADIO	.		69	*	.		.	
TAVOLI_EXTRA	.		69	*	.		.	
IPERBARICHE	-38		50	*	.		.	
MEDICI_DIP_PUB	44		.		76	*	.	
MEDICI_DIP_SSN	.		.		74	*	.	
MEDICI_DIP_PRIV	.		36		64	*	.	
POST_LET_DAY_PRC	.		.		53	*	38	
POSTI_LETTO_PRC3	.		.		.		92	*

Se i fattori indipendenti ottenuti spiegano il 70% della varianza totale, le variabili elementari permettono anche in questo caso, abbastanza agevolmente, di spiegare i fattori alla luce delle variabili più correlate con essi; più nello specifico:

- il fattore 1 è correlato con il numero pro capite di infermieri, tecnici e personale riabilitativo;
- il fattore 2 è correlato con il numero pro capite di beni strumentali ospedalieri;

5 I valori stampati sono moltiplicati per 100 e arrotondati all'intero più vicino. I valori maggiori di 60 sono contrassegnati da un '*'. I valori minori di 30 non sono stampati. Per una maggior chiarezza riguardo le variabili utilizzate si faccia riferimento alla tabella 14 riportata in Appendice.

- il fattore 3 è correlato con il numero di medici pro capite sia pubblici che privati;
- il fattore 4 è correlato con il numero di posti letto pro capite sia in regime di day hospital che ordinario.

Le figure 10, 11, 12 e 13, pur con le specifiche differenze regionali, evidenziano alcune regolarità: per quanto riguarda i due fattori di input lavoro (fattore 1 e 3) il fattore 1 (infermieri e tecnici) presenta un andamento leggermente crescente, mentre il fattore 3 (medici) appare più stabile nel tempo soprattutto a partire dal 2004, anno nel quale è stato introdotto il blocco del turnover; per quanto riguarda i due fattori di tipo capitale (fattore 2 e 4) si assiste ad un andamento decrescente soprattutto del fattore 4 (posti letto). Su questi risultati sembrano aver inciso, da un lato l'adozione dei piani di rientro a partire dal 2007 e dall'altro lato i provvedimenti relativi al blocco del turnover a partire dal 2004.

Blocco del turnover e piani di rientro

Il blocco del turnover (regolato della legge 30 dicembre 2004, n. 311 e dal decreto-legge 31 maggio 2010 n. 78 (convertito nella L. 122/2010)) prevede il blocco automatico delle assunzioni fino al 31 dicembre del secondo anno successivo a quello in corso qualora i provvedimenti necessari per il ripianamento del disavanzo di gestione non vengano adottati dal commissario ad acta entro il 31 maggio con riferimento agli anni di imposta 2006 e successivi. Sono vietate, inoltre, le assunzioni di personale a qualsiasi titolo e con qualsivoglia tipologia contrattuale negli enti nei quali l'incidenza delle spese di personale è pari o superiore al 40% delle spese correnti.

I piani di rientro sono strumenti attraverso cui le Regioni, che presentano deficit sanitari strutturali, stabiliscono, di concerto con i Ministeri della Salute e dell'Economia, gli obiettivi e le azioni strategiche finalizzate al recupero dell'equilibrio finanziario e alla rimozione delle determinanti strutturali del disequilibrio. Le regioni interessate ai piani di rientro sono: Sicilia, Lazio, Campania, Molise, Abruzzo a partire dal 2007; Liguria e Sardegna dal 2007 al 2010; Calabria dal 2009; Puglia e Piemonte dal 2010.

5.2.1 Costruzione dell'indicatore composito di input

Per ottenere una migliore leggibilità economica dei fattori di input a partire dai fattori principali individuati sono stati costruiti due indicatori compositi di input:

- Indicatore composito dell'input lavoro composto dal fattore degli input 1 (infermieri e tecnici) e dal fattore 3 (medici);
- Indicatore composito dell'input capitale composto dal fattore 2 (beni strumentali) e dal fattore 4 (posti letto).

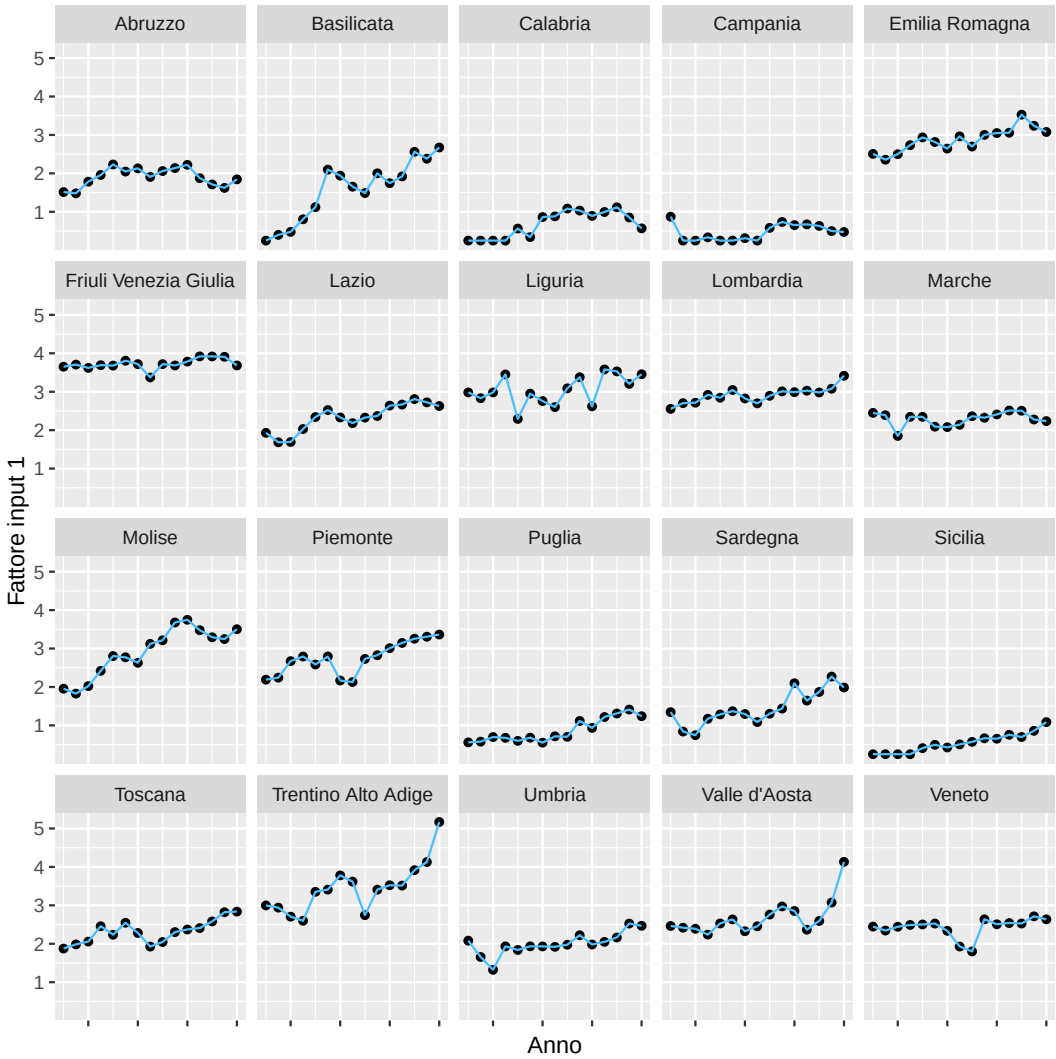


Figura 10: Dinamica del fattore di input 1 (infermieri e tecnici) per Regione ed anno, anni 1998-2012

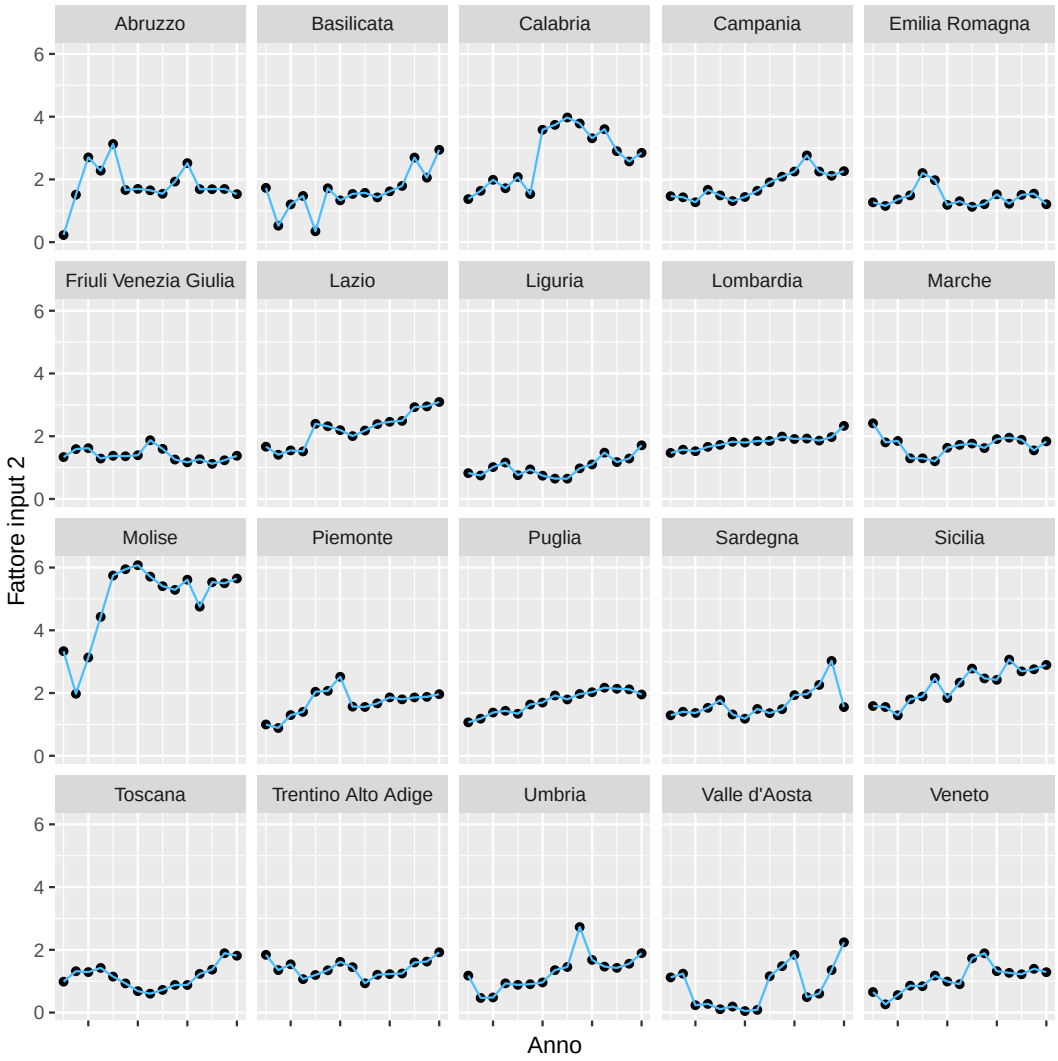


Figura 11: Dinamica del fattore di input 2 (beni strumentali) per Regione ed anno, anni 1998-2012

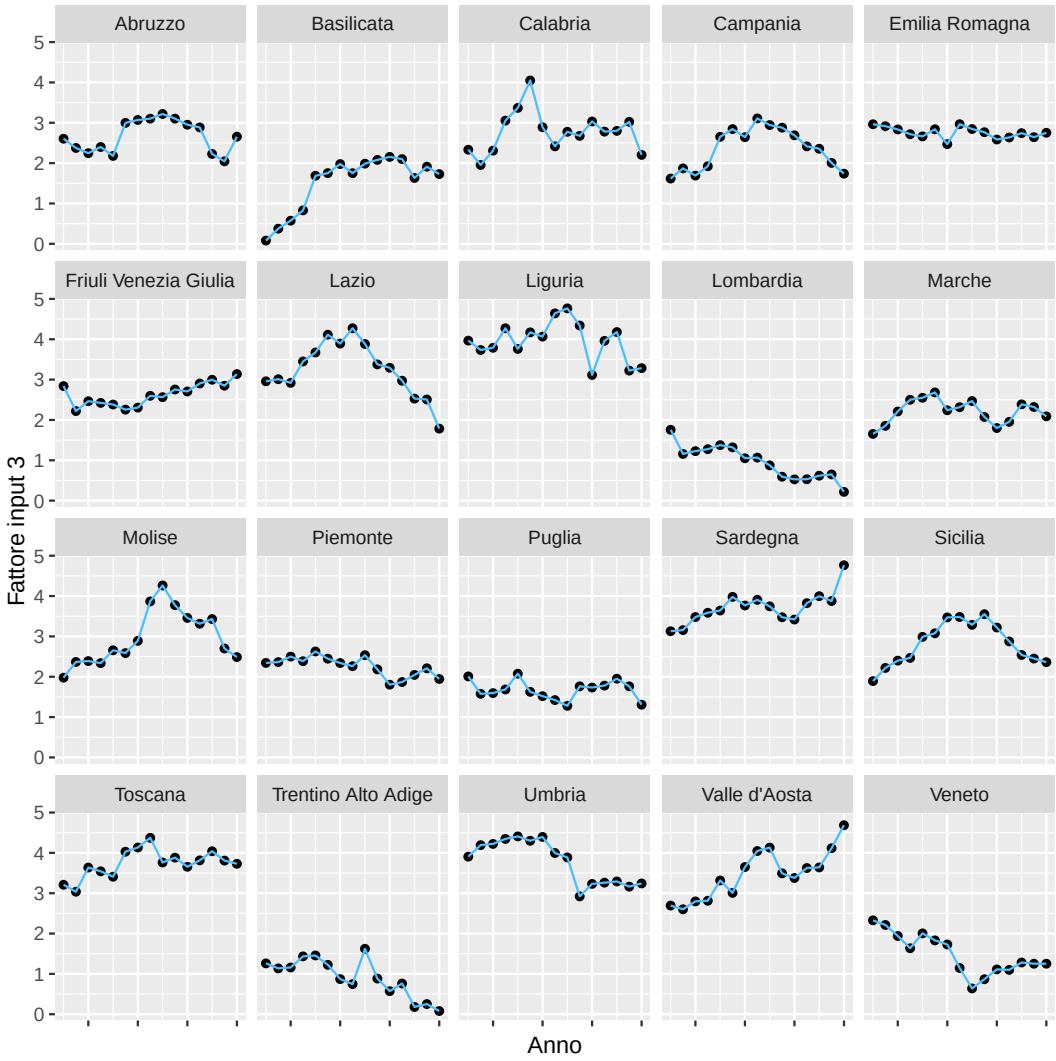


Figura 12: Dinamica del fattore di input 3 (medici) per Regione ed anno, anni 1998-2012

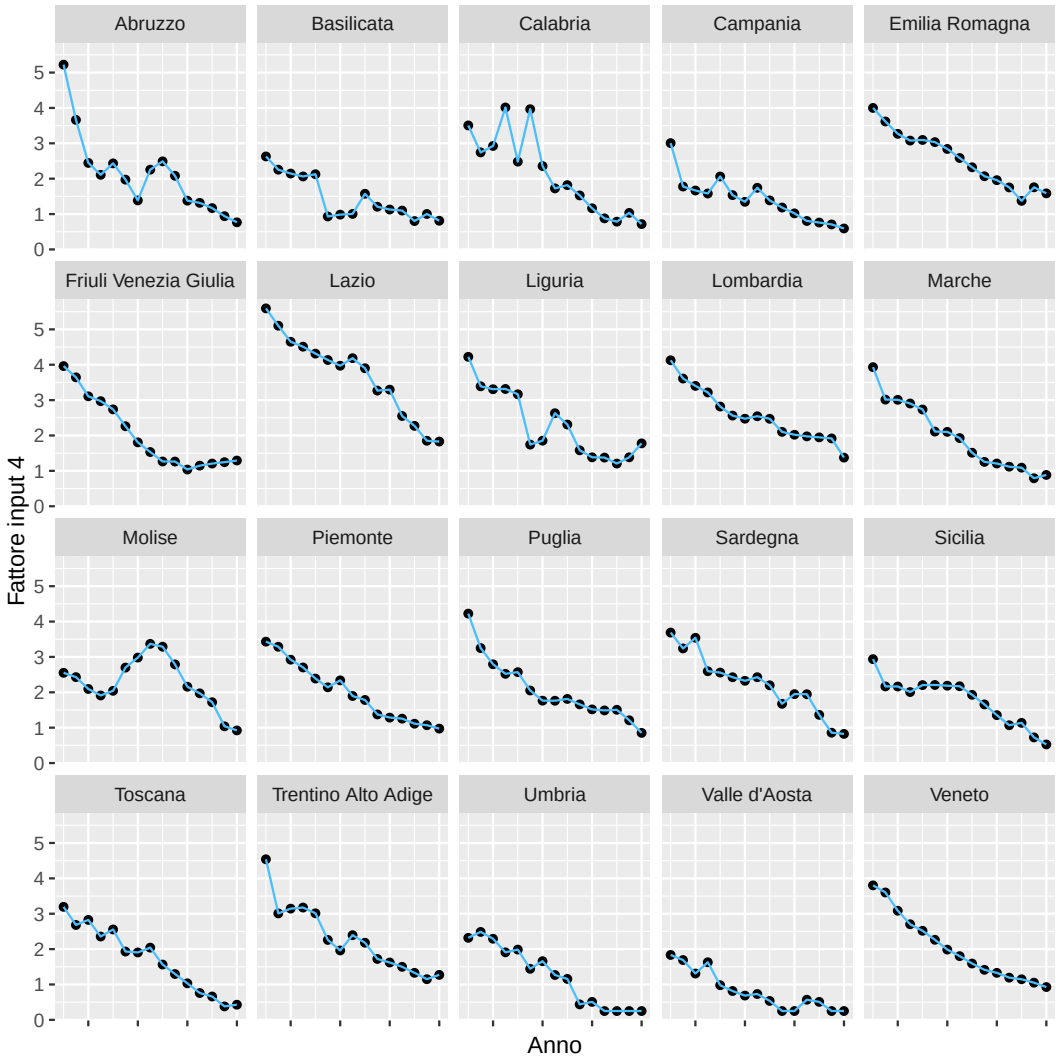


Figura 13: Dinamica del fattore di input 4 (posti letto) per Regione ed anno, anni 1998-2012

Dinamica per Regione ed anno: http://www.cermlab.it/saniregio2016/dinamica.html

In figura 14 viene riportata la distribuzione degli indicatori compositi del lavoro e del capitale per regione per l'anno 2012; si osserva la prevalenza del fattore lavoro, che in generale risulta mediamente superiore nelle regioni settentrionali, mentre il capitale sembra distribuirsi più uniformemente sul territorio con punte massime in Lombardia, Lazio e Molise.

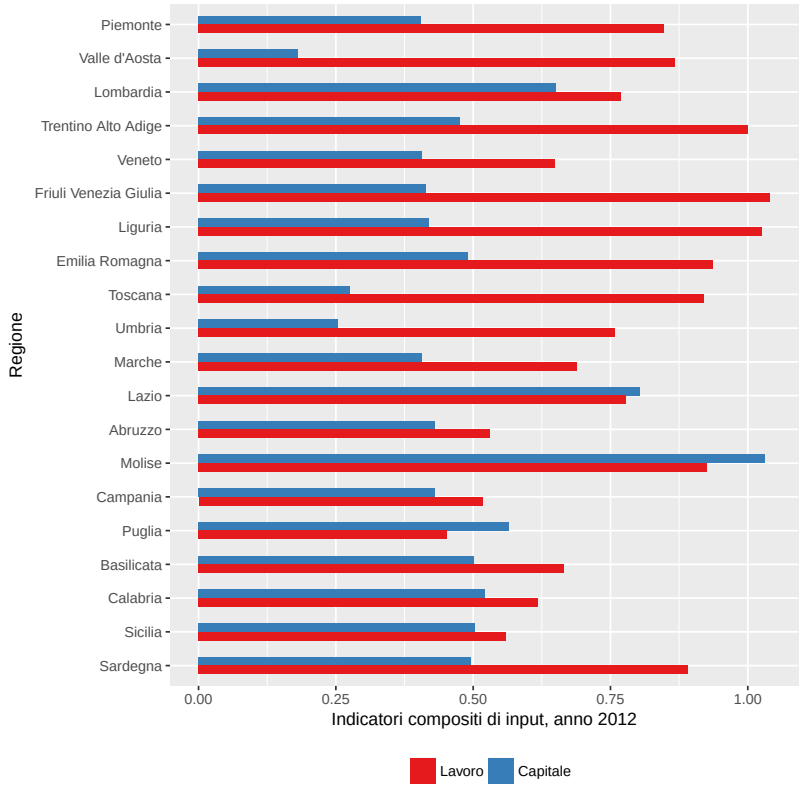


Figura 14: Distribuzione degli indicatori compositi di input Lavoro e Capitale per Regione, anno 2012

Molto esplicativa risulta anche essere la figura 15, nella quale viene riportata la distribuzione degli indicatori compositi del lavoro e del capitale media tra le regioni per gli anni 1998-2012; si osserva un forte incremento dell'input lavoro sino al 2003, dopodiché si registra un livello costante, mentre l'input capitale mostra una costante riduzione, soprattutto a partire dal 2003.

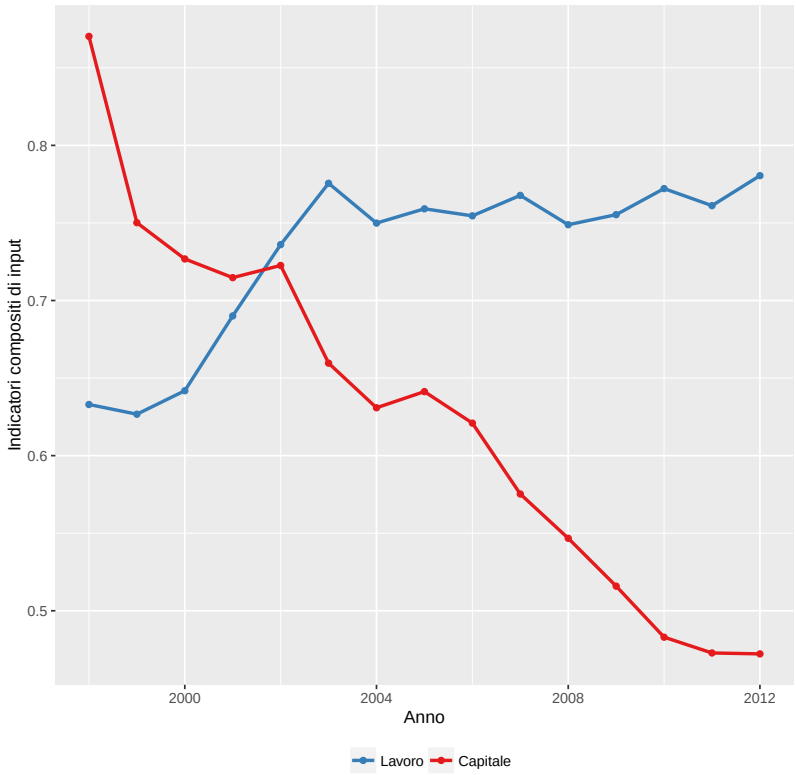


Figura 15: Dinamica media degli indicatori compositi di input Lavoro e Capitale per anno

Dinamica per anno: http://www.cermlab.it/saniregio2016/dinamica_media_in_out.html

5.3 STIMA DELL'EFFICIENZA TECNICA PRODUTTIVA

Dopo aver analizzato il livello dei servizi (output) e le risorse (input) impiegate nel processo produttivo, può essere calcolata per ogni Regione e per ogni anno di riferimento un indice di efficienza tecnica relativamente alle regioni benchmark, ricordando che in un modello non parametrico di tipo DEA o Robust Nonparametric Efficiency Analysis - Order-m (Order-m) tali benchmark non vengono individuati a priori, ma sono uno dei risultati del problema di ottimizzazione relativo alla base di questi modelli.

Tale indice o punteggio di efficienza tecnica, dunque, come illustrato nei paragrafi precedenti, è stato calcolato mettendo a confronto un unico indicatore composito di output che include due fattori:

- fattore dimensionale (degenza per tipologia, giornate e numero) che spiega il 79% della varianza dell'output,
- fattore qualitativo (saldi di mobilità) che spiega il 21% della varianza dell'output,

e due indicatori compositi di input (che includono a loro volta due fattori):

- indicatore composito del lavoro (personale tecnico e personale medico) che spiega il 57% della varianza degli input;
- indicatore composito del capitale (macchinari quali TAC, camere iperbariche ecc.. e posti letto) che spiega il 43% della varianza degli input.

Tale calcolo è stato condotto sia con tecniche parametriche classiche, quali la DEA, sia con tecniche robuste (di tipo Order-m), ottenendo comunque un'ottima correlazione tra i due punteggi.

Mappa dell'efficienza per Regione: http://www.cermlab.it/saniregio2016/mappa_eff.html

In figura 16 viene riportata la dinamica dell'efficienza tecnica DEA per regione dal 1998 al 2012; si osserva da un lato un incremento medio dell'efficienza, soprattutto dopo il 2003, mentre per alcune regioni, come la Toscana, il Veneto, la Lombardia, l'Emilia Romagna e il Piemonte tale incremento è particolarmente pronunciato sia in termini di miglioramento negli anni che di posizionamento relativo nell'ultimo anno di analisi.

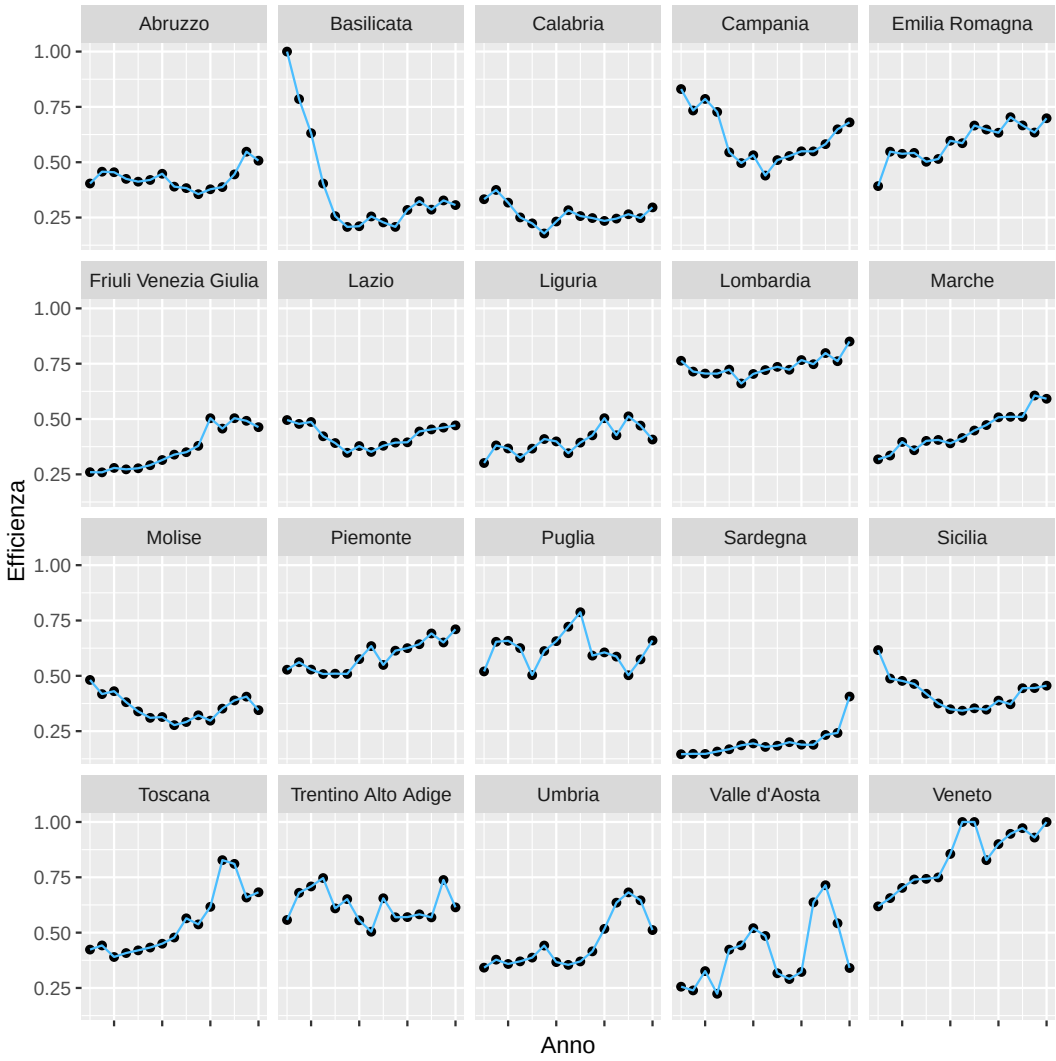


Figura 16: Dinamica dell'efficienza tecnica produttiva per Regione ed anno

5.4 UN ESERCIZIO DI SIMULAZIONE PER GLI ANNI 2013-2015

I risultati presentati nei paragrafi precedenti, basandosi sui dati ufficiali *Health for All* di fonte ISTAT, forniscono una stima dell'efficienza produttiva per gli anni 1998-2012; a partire da tali risultati, quindi, può essere utile, sia come esercizio di simulazione che come stima di possibili scenari, predisporre previsioni sulle principali stime ottenute in questa sezione per gli anni 2013, 2014 e 2015.

Non avendo ulteriori informazioni cui legare anche in modo indiretto il trend dell'efficienza e degli indicatori compositi di input e di output è stato necessario costruire un modello di previsione che tenesse in considerazione la sola serie storica per definire i valori attesi nei tre anni successivi.

Le figure 8, 12 e soprattutto 16 evidenziano dinamiche differenziate per Regione e fortemente non lineari; tali evidenze ci hanno spinto a non effettuare previsioni del trend come effetto medio costante nell'intero periodo (trend lineare), ma ad utilizzare metodi non parametrici più flessibili che, pur insistendo sull'intero periodo 1998-2012, tenessero in considerazione maggiormente il trend degli ultimi 4-5 anni.

Più in particolare proponiamo di utilizzare un approccio di tipo Generalized Additive Model (GAM) per la stima del trend medio e per la successiva previsione combinando la semplice struttura additiva dei modelli di regressione parametrica (si tratta pur sempre di un modello additivo) con la flessibilità che caratterizza l'approccio non parametrico, dato che nel modello GAM non si impone alcuna restrizione riguardo la forma delle funzioni che determinano come la variabile indipendente X influenza il valore atteso della variabile dipendente Y.

Un modello GAM, dunque, stima la variabile di risposta Y utilizzando una somma di funzioni *smooth*[6] delle variabili dipendenti (nel nostro caso gli anni) X_j for $j = 1998, \ldots, 2012$. In un contesto regressivo il modello GAM può essere espresso come:

$$\mu = E(Y|\mathbf{X} = \mathbf{x}) = \alpha + \sum_{j=1}^{p} f_j(X_j), \tag{16}$$

dove $f_j(\cdot)$ rappresentano le funzioni *smooth*.

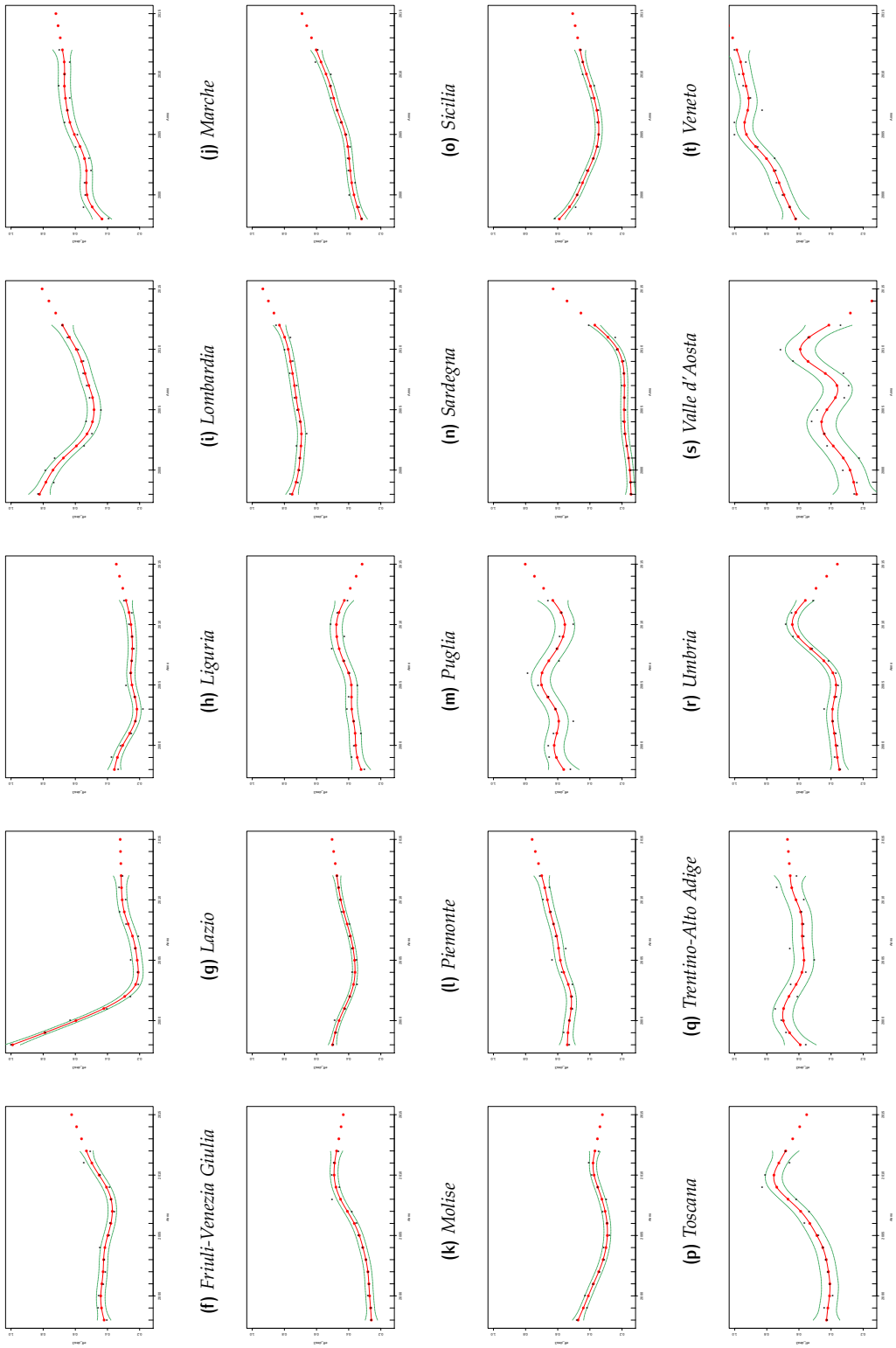
In tabella 5 viene riportata, distintamente per ogni Regione, sia la stima GAM sull'efficienza tecnica per gli anni 1998-2012 (linea rossa e relativi intervalli di confidenza in verde), sia le stime per i tre anni successivi (asterischi in rosso).

Efficienza per Regione: `http://www.cermlab.it/saniregio2016/stima_efficienza.html`

6 Il parametro di *smooth*, inoltre, permette di tenere in considerazione un differente numero di anni nella stima di forecast e può essere quindi un utile parametro di tuning.

Per semplicità espositiva non vengono riportate in figura le previsioni riguardanti gli indicatori compositi di input (lavoro e capitale) e di output, stime ottenute con la stessa tecnica descritta in precedenza.

Tabella 5: Stima dell'andamento dell'efficienza tecnica produttiva per gli anni 2013-2015



6 VALORIZZAZIONE DELLA SPESA SANITARIA STANDARD E ANALISI DELLE CAUSE DI SCOSTAMENTO DALLA SPESA STORICA

Indice

6.1 La stima della funzione di domanda e dell'output-gap 65
6.2 Stima della funzione di spesa 69
6.3 Applicazione dei modelli per la valutazione della spesa standard 73
6.4 Riassunto dei risultati e confronti con la spesa storica 79

6.1 LA STIMA DELLA FUNZIONE DI DOMANDA E DELL'OUTPUT-GAP

Il modello illustrato nella sezione 4.2.1 è stato impiegato per stimare la funzione di domanda dei servizi sanitari e la sua forma ridotta, quest'ultima denominata funzione di output. Come variabile dipendente si è utilizzato l'indicatore composito dell'output calcolato nella sezione 5.1.1. In linea con il modello empirico di riferimento (equazione (5)), tra i regressori della funzione di output sono stati inclusi i seguenti gruppi di variabili: *(i)* come variabili strutturali i saldi di mobilità ospedaliera interregionale; *(ii)* come variabili di contesto legate alla domanda il PIL regionale per abitante, che rappresenta una proxy del reddito, la struttura demografica suddivisa in classi per fascia di età ed alcune variabili legate alla condizione di salute; *(iii)* come variabili di contesto legate all'offerta (utilizzate come variabili strumentali della spesa nella stima della funzione di domanda) la spesa sanitaria delle famiglie, un indicatore di progresso tecnologico nell'offerta medica ed altri fattori legati alle condizioni di vita; *(iv)* da ultimo come determinanti istituzionali la presenza di piani di rientro suddivisi per anno di introduzione.

Grazie alla struttura longitudinale del nostro dataset, che in relazione alle 20 regioni abbraccia un arco temporale di 15 anni dal 1998 al 2012 per un totale di 300 osservazioni, si è potuto stimare un modello panel lineare di tipo "fixed effect" utilizzando lo stimatore *within-the-group* in modo da controllare le specificità regionali e l'effetto del tempo includendo tra i regressori gli effetti fissi regionali e gli effetti fissi degli anni in linea con quanto riportato nell'equazione 5.

Il campione di regressione si restringe da 300 a 284 osservazioni a causa della mancanza di informazioni in relazione ad alcune variabili di contesto, di input e di output che l'ISTAT non aveva ancora aggiornato al momento di inizio delle nostre elaborazioni per gli anni successivi al 2012. Inoltre, dal campione di regressione sono state escluse le seguenti regioni in quanto hanno riportato dei valori anomali in relazione ai livelli di spesa e/o inefficienza tecnica: Abruzzo (1998, 2004), Basilicata (1998, 2010), Friuli Venezia Giulia (2005, 2008), Lombardia (1998), Molise (1998, 1999), Puglia (1998), Toscana (2009), Valle d'Aosta (1998, 1999, 2004, 2007, 2010).

La tabella 6 mostra le stime puntuali dei coefficienti della funzione di output e della funzione di domanda. La colonna 1, riporta le stime ottenute con riferimento al modello completo della funzione di output che rappresenta il nostro modello di riferimento per tutte le elaborazioni successive, mentre le colonne da 2 ad 7 mostrano le stime riferite ai modelli parziali in modo da valutare l'apporto delle singole categorie delle variabili indipendenti. Al fine di verificare l'aderenza del modello empirico utilizzato al framework teorico di riferimento, le colonne 8 e 9 della tabella 6 mostrano le stime riferite alla funzione di domanda che, diversamente dalla funzione di output, vede al posto delle variabili di contesto dell'offerta la spesa pro capite come proxy del costo unitario di fornitura del servizio. In prima battuta la stima dei coefficienti della funzione di domanda è stata effettuata senza strumentare la spesa spesa corrente pro capite (colonna 8), successivamente le stime sono ripetute utilizzando le variabili di contesto legate all'offerta come variabili strumentali per la spesa corrente pro capite (colonna 9) in modo da poterne testare l'endogeneità.

Se si guarda la tabella 6, i risultati mostrano diversi aspetti di rilievo: *(i)* in primo luogo si riscontra come la mobilità tra le regioni risulti uno dei fattori più significativi nell'analisi del livello di domanda di servizi che ogni sistema sanitario deve fronteggiare, *(ii)* la presenza di specifici piani di rientro ha influito negativamente sul livello dell'output soprattutto nel medio termine, da cui si evince come una parte dei risparmi di spesa derivanti da queste misure siano dovuti non a recuperi di efficienza, ma ad una riduzione nell'erogazione dei servizi;*(iii)* le principali variabili di contesto legate alla domanda (come le variabili legate alla condizione di salute e le variabili riferite alle fasce di popolazione) mostrano una relazione con la spesa in linea con le aspettative; *(iv)* in particolare, i coefficienti stimati per le fasce di età della popolazione residente evidenziano come la domanda di servizi sanitari risulti concentrata in tre fasce di età: tra 0-4 anni; tra 45-54 anni e nell'ultima fascia oltre il 75 anni di età; *(v)* la spesa privata delle famiglia (così come le altre variabili di offerta) presentano i segni attesi in termini di impatto sulla domanda, nonostante il basso livello di significatività statistica individuale dei singoli coefficienti. Si deve notare come il PIL per abitante risulte avere, quando statisticamente significativo, un impatto negativo sul livello di output esercitando quindi un effetto di prezzo sull'output piuttosto che un effetto di reddito.

Le stime puntuali dei coefficienti, ottenute con riferimento al modello base della funzione di output, riportato nella colonna 1 della tabella 6, si sono dimostrati robusti al cambiare della specificazione del modello, come mostrato nelle colonne da 2 a 8. In particolare i coefficienti per i quali si registra una significatività statistica elevata (p-value inferiore a

Tabella 6: Funzione di domanda per modello di stima, variabile dipendente = indicatore composito di output, anno 2014.

Variabili	**Funzione di output**							**Funzione di domanda**	
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)	OLS (7)	OLS (8)	2SLS (9)
VARIABILI STRUTTURALI									
Saldi mobilità ospedaliera interregionale (attivi-passivi) - per 10.000 abit.	1.369	1.314	1.315	1.356	1.482	1.800	1.538	1.437	1.834
	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***	[0.000]***
PIANI DI RIENTRO									
Piano di rientro (t+1)	-0.0110		-0.0213					-0.00615	0.0102
	[0.395]		[0.105]					[0.556]	[0.627]
Piano di rientro (t+2)	-0.0291		-0.0393					-0.0271	-0.0220
	[0.092]*		[0.011]**					[0.035]**	[0.289]
Piano di rientro (t+3)	-0.0361		-0.0657					-0.0402	-0.0447
	[0.081]*		[0.002]***					[0.029]**	[0.030]**
Piano di rientro (t+4)	-0.0109		-0.0569					-0.0190	-0.0333
	[0.595]		[0.050]**					[0.318]	[0.124]
VARIABILI DI CONTESTO LEGATE ALLA DOMANDA									
Popolazione residente 0-4 M+F - % pop totale	0.241				0.234			0.232	0.315
	[0.001]***				[0.008]***			[0.012]**	[0.000]***
Popolazione residente 5-14 M+F - % pop totale	0.201				0.153			0.187	0.216
	[0.000]***				[0.077]*			[0.007]***	[0.000]***
Popolazione residente 15-24 M+F - % pop totale	0.0283				0.0494			0.0356	0.0937
	[0.363]				[0.101]			[0.211]	[0.002]***
Popolazione residente 25-34 M+F - % pop totale	0.121				0.0853			0.121	0.129
	[0.000]***				[0.067]*			[0.005]***	[0.000]***
Popolazione residente 45-54 M+F - % pop totale	0.122				0.0972			0.0974	0.102
	[0.000]***				[0.008]***			[0.003]***	[0.000]***
Popolazione residente 55-64 M+F - % pop totale	0.103				0.127			0.105	0.157
	[0.003]***				[0.006]***			[0.006]***	[0.000]***
Popolazione residente 65-74 M+F - % pop totale	0.0881				0.0920			0.0850	0.128
	[0.002]***				[0.008]***			[0.006]***	[0.000]***
Popolazione residente 75+ M+F - % pop totale	0.172				0.106			0.157	0.206
	[0.000]***				[0.000]***			[0.002]***	[0.000]***
% grandi fumatori 15+ M+F	0.000666					0.000622		0.00118	0.00130
	[0.645]					[0.749]		[0.448]	[0.470]
Tasso abortività volontaria 15-49	0.00412					0.00310		0.00330	0.00186
	[0.508]					[0.733]		[0.665]	[0.733]
% persone consumano verdura almeno 1 volta al giorno 3+ M+F	-0.000551					-0.00199		-0.000814	-0.00164
	[0.549]					[0.171]		[0.396]	[0.156]
% persone consumano pesce qualche volta la settimana 3+ M+F	-0.00255					-0.00196		-0.00252	-0.00300
	[0.002]***					[0.281]		[0.025]**	[0.007]***
% persone consumano carni bovine qualche volta la settimana 3+ M+F	0.000923					0.00192		0.00123	0.000869
	[0.150]					[0.097]*		[0.160]	[0.493]
% persone consumano formaggio almeno 1 volta al giorno 3+ M+F	0.000190					0.000368		-0.000856	-0.000608
	[0.866]					[0.836]		[0.411]	[0.635]
% persone colazione adeguata 3+ M+F	0.000763					0.00208		0.00149	0.00213
	[0.506]					[0.094]*		[0.307]	[0.166]
% persone pasto principale cena 3+ M+F	-0.00155					-0.000133		-0.00109	-0.000679
	[0.397]					[0.950]		[0.431]	[0.688]
Tasso incidenza tumori maligni 0-84 M	0.00199					-0.00545		0.00321	0.00598
	[0.285]					[0.001]***		[0.373]	[0.025]**
Tasso incidenza tumori maligni 0-84 M (quadrato)	-0.00000312					0.00000401		-0.00000429	-0.00000646
	[0.095]*					[0.007]***		[0.208]	[0.006]***
VARIABILI DI CONTESTO LEGATE ALL'OFFERTA (utilizzate come variabili strumentali della spesa nella funzione di domanda)									
PIL ai prezzi di mercato - (€) reali pro capite	-0.00252			0.00139					
(differenza dalla media, base 2005)	[0.026]**			[0.532]					
Spesa sanitaria delle famiglie (%)	0.00385						0.0101		
	[0.164]						[0.037]**		
Indicatore di progresso tecnologico nell'offerta medica	0.000943						-0.000448		
	[0.118]						[0.651]		
Tasso attività 15+ F	-0.00387						0.00661		
	[0.204]						[0.146]		
% pop. con titolo universitario M+F	-0.00604						-0.00981		
	[0.352]						[0.239]		
% famiglie che lamentano inquinamento acustico	-0.00111						-0.00157		
	[0.646]						[0.331]		
% famiglie che lamentano sporcizia stradale	-0.00165						-0.00380		
	[0.093]*						[0.020]**		
% famiglie che lamentano inquinamento aria	0.000474						0.000315		
	[0.808]						[0.854]		
Speranza di vita 75 M	0.00712						0.00641		
	[0.602]						[0.862]		
Speranza di vita 75 F	-0.00332						-0.00626		
	[0.828]						[0.873]		
VARIABILI DI SPESA									
Spesa sanitaria pubblica corrente pro capite								0.0000336	-0.000355
								[0.561]	[0.006]***
Numero di osservazioni	283	284	284	284	284	283	284	283	283
Effetti fissi regionali	Si	Si	Si	Si	Si	Si	Si	Si	Si
Effetti fissi annuali	Si	Si	Si	Si	Si	Si	Si	Si	Si
R^2	0.746	0.321	0.360	0.330	0.623	0.533	0.451	0.696	
Adjusted R^2	0.695	0.283	0.314	0.290	0.590	0.488	0.400	0.648	
F test for excluded instruments (pvalue)									0,068
Overidentification test, Sargan Statistics $\chi^2(X)$ (pvalue)									0,312

Numero di osservazioni = 284; s.e. clusterizzati a livello di regione, p-value riportati tra parentesi. * significatività inferiore al 10%, ** significatività inferiore al 5%, *** significatività inferiore all'1%.

0,10) presentano valori comparabili lungo le diverse specificazioni del modello base. Se si considerano i coefficienti stimati della funzione di domanda riportati nelle colonna 9, è molto interessante notare come la robustezza delle stime aumenti dopo aver strumentato la spesa pro capite con le variabili di contesto dell'offerta, evidenza che porta alla luce i problemi di endogeneità presenti nella funzione di domanda. Questi risultati confermano, quindi, da un lato la necessità di utilizzare un modello a due stadi per correggere l'endogeneità della spesa rispetto all'output, dall'altro lato convalidano la scelta di usare come punto di riferimento la forma ridotta della funzione di domanda in quanto non solo consente di superare i problemi di endogeneità, ma al tempo stesso consente di non legare direttamente il livello di servizio standard ad una stima della spesa standard garantendo un alto livello di semplificazione. Nel modello in cui si utilizzano le variabili strumentali (colonna 9) la spesa pro capite mostra un impatto negativo e significativo sulla domanda di servizi corroborando la validità del framework teorico sottostante basato sulla domanda e l'offerta del servizio sanitario. I test relativi alla bontà delle variabili strumentali, riportati in coda alla colonna 9 della tabella 6, mostrano sia la buona correlazione tra le variabili strumentali e la spesa (*F-test for excluded instruments*) sia la sostanziale esogeneità delle variabili strumentali (*Overidentification text*).

Una volta stimati i valori attesi del livello dei servizi offerti[1], grandezza che chiameremo anche *output standard*, è stata calcolata la differenza tra output storico e output standard per ogni regione in ogni anno dell'analisi, in modo da ottenere un valore riferito all'*output-gap* secondo la formula riportata nell'equazione (7). E' importante sottolineare che, nella valorizzazione dell'output standard (e quindi dell'output gap), gli effetti negativi dei piani di rientro siano stati neutralizzati ponendo, quindi, pari a zero i valori di tutte le dummy, in modo da non ridurre il livello di output standard delle regione soggette ai piani di rientro, essendo tale riduzione un effetto collaterale di queste misure di riordino finanziario.

L'output gap, quindi, identifica, per ogni regione, l'adeguatezza dei servizi offerti avendo come benchmark il sistema sanitario regionale che a parità di variabili di contesto riesce a massimizzare il livello di prestazioni per abitante. Nel nostro modello la regione benchmark è la Lombardia (in quanto è, tra le grandi regioni, quella che presenta i livelli di soddifacimento della domanda e di efficienza più alti), di conseguenza per le altre regioni l'output gap mostra il livello dei servizi pro capite che si sarebbe potuto produrre se si fosse operato come la Lombardia a parità di variabili che misurano il contesto socio-economico di riferimento. E' importate, a tal proposito, precisare che il livello di output standard ottimale di ogni regione, anche considerando l'output gap rispetto alla regione benchmark, non sarà necessariamente uguale a quello della Lombardia, in quanto il livello di servizio ottimale dipende in primis dalle caratteristiche di ogni territorio tra cui, ad esempio, la struttura della popolazione.

La stima dell'output gap di ogni regione è il risultato di due componenti: 1) l'output gap interno o proprio della regione, che misura quanto una regione riesce a soddisfare la domanda dei propri cittadini rispetto al benchmark di riferimento; 2) l'output gap legato alla mobilità. Quest'ultimo quando assume valore negativo aggrava l'output gap interno e

1 Il modello di riferimento è quello completo della funzione di output riportato nella colonna 1 della tabella 6.

potrebbe anche non essere necessario isolarlo, viceversa quando assume un valore positivo va identificato separatamente e sottratto dall'output gap complessivo al fine di individuare correttamente l'output gap proprio di queste regioni. Il calcolo delle due componenti di output gap è stato effettuato utilizzando la stima del coefficiente sui saldi di mobilità riportato nella colonna 1 della tabella 6 e le formule riportate nelle equazioni (7), (8) e (9) .

6.2 STIMA DELLA FUNZIONE DI SPESA

In modo speculare alla analisi svolta in merito alla funzione di domanda, il modello illustrato nella sezione 4.2.2 è stato impiegato per stimare la funzione di costo dei servizi sanitari e la sua forma ridotta, che abbiamo chiamato funzione di spesa. Come variabile dipendente è stata utilizzata la spesa sanitaria corrente per abitante pubblicata dall'ISTAT nell database "*Health for All*". In linea con il modello di funzione di spesa aumentato riportato nell'equazione (12), tra i regressori sono state incluse (i) come variabili strutturali l'indice di inefficienza tecnica calcolato nella sezione 5.3 e l'output-gap stimato attraverso la funzione di output descritta sopra nella sezione 6.1; (ii) come variabili di contesto legate alla domanda (utilizzate come variabili strumentali dell'output nell'ambito della funzione di costo) il PIL regionale per abitante - proxy del reddito - la struttura demografica suddivisa in classi per fascia di età ed alcune variabili legate alla condizione di salute, (iii) come variabili di contesto legate all'offerta la spesa sanitaria delle famiglie, un indicatore di progresso tecnologico nell'offerta medica ed altri fattori legati alle condizioni di vita ed infine (iv) come determinanti istituzionali la presenza dei piani di rientro suddivisi per anno di introduzione. E' importante notare, quindi, come ad eccezione delle variabili strutturali, i regressori utilizzati per specificare la funzione di spesa sono identici a quelli utilizzati per specificare la funzione di output con l'unica differenza rappresentata dalle variabili strutturali, questo mette in evidenza la specularità dei due modelli secondo quanto previsto dal framework teorico di riferimento.

Anche qui, grazie alla struttura longitudinale del nostro dataset, che in relazione alle 20 regioni abbraccia un arco temporale di 15 anni dal 1998 al 2012 per un totale di 300 osservazioni, si è potuto stimare un modello panel lineare di tipo "fixed effect" utilizzando lo stimatore *within-the-group* in modo da controllare le specificità regionali e l'effetto del tempo includendo tra i regressori gli effetti fissi regionali e gli effetti fissi degli anni in linea con quanto riportato nell'equazione 12.

Come per la stima della funzione di output, anche per la stima della funzione di spesa il campione di regressione si restringe da 300 a 284 osservazioni a causa della mancanza di informazioni in relazione ad alcune variabili di contesto, di input e di output che l'ISTAT non aveva ancora aggiornato al momento di inizio delle nostre elaborazioni per gli anni successivi al 2012. Inoltre, dal campione di regressione sono state escluse le seguenti regioni in quanto hanno riportato dei valori anomali in relazione ai livelli di spesa e/o inefficienza tecnica: Abruzzo (1998, 2004), Basilicata (1998, 2010), Friuli Venezia Giulia (2005, 2008), Lombardia

(1998), Molise (1998, 1999), Puglia (1998), Toscana (2009), Valle d'Aosta (1998, 1999, 2004, 2007, 2010). Quindi, il modello della funzione di spesa è speculare a quello della funzione di output anche per quanto riguarda la struttura del campione di regressione.

La tabella 7 mostra le stime puntuali dei coefficienti della funzione di spesa e della funzione di costo. La colonna 1, riporta le stime ottenute con riferimento al modello completo della funzione di spesa che rappresenta il nostro modello di riferimento per tutte le elaborazioni successive, mentre le colonne da 2 ad 7 mostrano le stime riferite ai modelli parziali in modo da valutare l'apporto delle singole categorie delle variabili indipendenti. Al fine di verificare l'aderenza del modello empirico utilizzato al framework teorico di riferimento, le colonne 8 e 9 della tabella 7 mostrano le stime riferite alla funzione di costo che, diversamente dalla funzione di spesa, vede al posto delle variabili di contesto della domanda un polinomio di secondo grado dell'indicatore composito di output (la variabile dipendente dalla funzione di domanda) individuato come descritto nella sezione 5.1.1. In prima battuta la stima dei coefficienti della funzione di costo è stata effettuata senza strumentare le variabili di output (colonna 8), successivamente le stime sono ripetute utilizzando le variabili di contesto legate alla domanda come variabili strumentali per l'indicatore composito di output (colonna 9) in modo da poterne testare l'endogeneità.

I risultati riportati nella tabella 7 mostrano come la maggior parte dei coefficienti, indipendentemente dalla specificazione del modello, presentano i segni attesi e, per le variabili più rilevanti, anche alti livelli di significatività statistica. In particolare è importante mettere i risalto i seguenti risultati: *(i)* l'indice di inefficienza tecnica e l'output-gap presentano entrambi segno positivo e un elevato grado di significatività statistica, evidenziando come una parte non trascurabile della spesa storica sia spiegata dalla presenza di input in eccesso rispetto all'output prodotto e dall'offerta di servizi non in linea rispetto alla standard di riferimento misurato in base alla domanda di ogni territorio; *(ii)* la presenza di specifici piani di rientro corrisponde mediamente ad una contrazione della spesa, ma non prima di due anni dalla loro applicazione, inoltre i coefficienti non appaiono statisticamente significativi corroborando i dubbi sulla piena efficacia di questi strumenti, soprattutto se questo risultato viene letto congiuntamente all'evidenza empirica ottenuta con la stima della funzione di domanda dove, invece, i piani di rientro sembrano aver influito negativamente sui livelli dei servizi offerti dalle regioni teatro di queste misure; *(iii)* le principali variabili di contesto legate alla domanda (quali il PIL per abitante e le variabili legate alla condizione di salute e le variabili riferite alle fasce di popolazione) mostrano una relazione con la spesa in linea con le aspettative; *(iv)* in particolare, i coefficienti stimati per le fasce di età della popolazione residente evidenziano l'atteso andamento ad U della spesa rispetto al crescere dell'età. Le fasce di età che assorbono la quota principale di spesa sono quella tra 0 e 4 anni e quella tra 55 e 64 anni; *(v)* da ultimo, nel gruppo delle variabili di contesto legate all'offerta, la spesa privata delle famiglia è la variabile che oltre a mostrare l'atteso segno negativo, il linea con il principio di sostituzione tra spesa privata e spesa pubblica, presenta anche un alto livello di significatività, mentre le altre variabili pur presentano i segni attesi, nella maggior parte dei casi sono contraddistinte, come nella stima della funzione di domanda, da un basso livello

Tabella 7: Funzione di spesa per modello di stima, variabile dipendente = spesa sanitaria pubblica corrente per abitante, anno 2014

Variabili	Funzione di spesa							Funzione di costo	
	OLS (1)	OLS (2)	OLS (3)	OLS (4)	OLS (5)	OLS (6)	OLS (7)	OLS (8)	2SLS (9)
VARIABILI STRUTTURALI									
Inefficienza tecnica	137.1	198.2	187.7	207.4	225.3	231.2	132.9	132.1	135.9
	[0.034]**	[0.005]***	[0.004]***	[0.003]***	[0.006]***	[0.001]***	[0.030]**	[0.093]*	[0.004]***
Output gap	405.0	556.1	553.0	571.9	577.9	597.2	475.3		
	[0.005]***	[0.000]***	[0.001]***	[0.000]***	[0.000]***	[0.000]***	[0.003]***		
PIANI DI RIENTRO (utilizzate come variabili strumentali dell'output nella funzione di costo)									
Piano di rientro (t+1)	58.12		63.89						
	[0.015]**		[0.003]***						
Piano di rientro (t+2)	24.95		44.28						
	[0.347]		[0.024]**						
Piano di rientro (t+3)	18.35		8.876						
	[0.545]		[0.634]						
Piano di rientro (t+4)	-26.88		-24.33						
	[0.356]		[0.280]						
VARIABILI DI CONTESTO LEGATE ALLA DOMANDA (utilizzate come variabili strumentali dell'output nella funzione di costo)									
PIL ai prezzi di mercato - (€) reali pro capite	-1.733			1.534					
(differenza dalla media, base 2005)	[0.262]			[0.242]					
Popolazione residente 0-4 M+F - % pop totale	188.9				219.8				
	[0.024]**				[0.004]***				
Popolazione residente 5-14 M+F - % pop totale	177.8				55.28				
	[0.014]**				[0.359]				
Popolazione residente 15-24 M+F - % pop totale	130.9				127.4				
	[0.006]***				[0.001]***				
Popolazione residente 25-34 M+F - % pop totale	76.68				20.45				
	[0.082]*				[0.616]				
Popolazione residente 45-54 M+F - % pop totale	21.63				34.30				
	[0.672]				[0.322]				
Popolazione residente 55-64 M+F - % pop totale	185.6				145.0				
	[0.001]***				[0.001]***				
Popolazione residente 65-74 M+F - % pop totale	126.3				103.5				
	[0.006]***				[0.006]***				
Popolazione residente 75+ M+F - % pop totale	141.0				123.2				
	[0.028]**				[0.007]***				
% grandi fumatori 15+ M+F	1.902					-1.273			
	[0.416]					[0.679]			
Tasso abortività volontaria 15-49	-3.080					6.780			
	[0.671]					[0.471]			
% persone consumano verdura almeno 1 volta al giorno 3+ M+F	-2.819					-3.217			
	[0.048]**					[0.144]			
% persone consumano pesce qualche volta la settimana 3+ M+F	-0.455					-1.428			
	[0.754]					[0.353]			
% persone consumano carni bovine qualche volta la settimana 3+ M+F	-0.745					-0.764			
	[0.606]					[0.681]			
% persone consumano formaggio almeno 1 volta al giorno 3+ M+F	-0.488					-1.838			
	[0.723]					[0.375]			
% persone colazione adeguata 3+ M+F	2.840					1.660			
	[0.151]					[0.366]			
% persone pasto principale cena 3+ M+F	1.156					0.697			
	[0.550]					[0.748]			
Tasso incidenza tumori maligni 0-84 M	9.341						2.222		
	[0.041]**						[0.277]		
Tasso incidenza tumori maligni 0-84 M (quadrato)	-0.00839						-0.00243		
	[0.036]**						[0.200]		
VARIABILI DI CONTESTO LEGATE ALL'OFFERTA									
Spesa sanitaria delle famiglie (%)	-10.86						-7.016	-16.10	-18.32
	[0.022]**						[0.045]**	[0.003]***	[0.000]***
Indicatore di progresso tecnologico nell'offerta medica	2.194						3.061	2.309	2.095
	[0.001]***						[0.000]***	[0.005]***	[0.000]***
Tasso attività 15+ F	15.26						13.38	4.690	3.628
	[0.002]***						[0.027]**	[0.339]	[0.244]
% pop. con titolo universitario M+F	7.885						1.659	3.429	4.109
	[0.425]						[0.880]	[0.730]	[0.536]
% famiglie che lamentano inquinamento acustico	2.044						1.582	-1.777	-1.320
	[0.322]						[0.530]	[0.525]	[0.481]
% famiglie che lamentano sporcizia stradale	0.525						1.270	3.891	3.976
	[0.652]						[0.375]	[0.006]***	[0.001]***
% famiglie che lamentano inquinamento aria	-2.953						-2.883	-0.866	-1.091
	[0.072]*						[0.109]	[0.586]	[0.453]
Speranza di vita 75 M	-24.27						-11.31	-1.032	6.161
	[0.416]						[0.639]	[0.958]	[0.819]
Speranza di vita 75 F	1.677						38.88	44.85	34.67
	[0.957]						[0.253]	[0.306]	[0.190]
VARIABILI DI OUTPUT									
Indicatore composito di output (BoD robusto)								813.7	1117.0
								[0.041]**	[0.000]***
Indicatore composito di output (BoD robusto) - quadrato								-295.7	-480.3
								[0.304]	[0.004]***
Numero di osservazioni	283	283	283	283	283	283	283	283	283
Effetti fissi regionali	Sì	Sì	Sì	Sì	Sì	Sì	Sì	Sì	Sì
Effetti fissi annuali	Sì	Sì	Sì	Sì	Sì	Sì	Sì	Sì	Sì
R^2	0.983	0.973	0.974	0.973	0.978	0.974	0.977	0.974	
Adjusted R^2	0.980	0.971	0.972	0.971	0.976	0.972	0.975	0.972	
F test for excluded instruments (pvalue)									0,005
Overidentification test, Sargan Statistics $\chi^2(X)$ (pvalue)									0,253

Numero di osservazioni = 283; s.e. clusterizzati a livello di regione, p-value riportati tra parentesi. * significatività inferiore al 10%, ** significatività inferiore al 5%, *** significatività inferiore all'1%. In tutti i modelli sono inclusi gli effetti fissi regionali e gli effetti fissi annuali dal 1998 al 2010.

di significatività statistica individuale dei singoli coefficienti.

Le stime puntuali dei coefficienti, ottenute con riferimento al modello base della funzione di spesa, riportato nella colonna 1 della tabella 7, si sono dimostrati robusti al cambiare della specificazione del modello, come mostrato nelle colonne da 2 a 8. In particolare i coefficienti per i quali si registra una significatività statistica elevata (p-value inferiore a 0,10) presentano valori comparabili lungo le diverse specificazioni del modello base, a conferma soprattutto della forte relazione positiva esistente tra inefficienza tecnica e spesa.

Se si considerano i coefficienti stimati della funzione di costo riportati nelle colonna 9 della tabella 7, è molto interessante notare che dopo aver strumentato le variabili di output sono confermate, da un lato la forte relazione positiva esistente tra output e spesa e dall'altro lato la presenza di economie di scala crescenti a favore, quindi, dei sistemi sanitari regionali di più ampie dimensioni.[2] Questi risultati corroborano la validità del framework teorico sottostante basato sulla domanda e l'offerta del servizio sanitario. Inoltre, con l'utilizzo delle variabili strumentali i coefficienti delle variabili di output pur mantenendo lo stesso segno mostrano una magnitudine più bassa, risultato che porta alla luce i problemi di endogeneità presenti nella funzione di costo confermando la necessità di utilizzare un modello a due stadi per correggere l'endogeneità dell'output rispetto alla spesa. In aggiunta, questi risultati convalidano anche la scelta di usare come punto di riferimento per la valutazione della spesa standardizzata la funzione di spesa, ovvero la forma ridotta della funzione di costo, in quanto non solo consente di superare i problemi di endogeneità, ma al tempo stesso consente valutare in termini monetari sia l'inefficienza tecnica che l'output gap e quindi il livello di adeguatezza dei servizi offerti da ogni sistema includendo questa valutazione nella costruzione del fabbisogno standard in modo semplice.

E' importante notare che la valutazione dei livelli di adeguatezza dei servizi offerti risulterebbe più difficile all'interno della funzione di costo a meno di non avere delle indicazioni precise circa il livello di output standard che ogni sistema sanitario regionale deve produrre in linea con i LEA. Implicitamente, l'utilizzo della funzione di spesa consente di superare questo problema, anche in assenza di LEA sui livelli complessivi di output, individuando come misura dell'output standard di ogni regione quelle mediamente offerto dagli enti con caratteristiche simili in relazione alle variabili di contesto e avendo come benchmark il sistema regionale che a parità di variabili di contesto massimizza la produzione di servizi. In presenza di LEA i risultati ottenuti con la funzione di spesa possono, comunque, essere utilizzati per introdurre una misura dell'adeguatezza dei servizi nella valutazione della spesa standard tarata sui LEA.

2 I test relativi alla bontà delle variabili strumentali, riportati in coda alla colonna 9 della tabella 7, mostrano sia la buona correlazione tra le variabili strumentali e l'output (*F-test for excluded instruments*) sia la sostanziale esogeneità delle variabili strumentali (*Overidentification text*).

6.3 APPLICAZIONE DEI MODELLI PER LA VALUTAZIONE DELLA SPESA STANDARD

In questa sezione si procederà ad applicare le stime discusse in precedenza in modo da poter valutare la spesa standardizzata regionale sulla base dell'adeguatezza dei servizi tenendo conto, allo stesso tempo, dei livelli di efficienza con cui gli stessi sono erogati nell'ambito dei veri sistemi regionali. Al fine di produrre una stima - la più attuale possibile - della spesa standard, delle possibili riduzioni di risorse associate all'inefficienza tecnica e delle risorse necessaria a colmare l'output-gap, i coefficienti stimati dei modelli base della funzione di output e della funzione di spesa riportati nella colonna 1 delle tabelle 7 e 6 sono applicati ai valori di output-gap e di inefficienza tecnica stimati per il 2014 con la metodologia descritta nella sezione 5.4 e ad un set di variabili di contesto aggiornate al 2014.

Le stime puntuali dei coefficienti della funzione di spesa riportate nella colonna 1 della tabella 7 consentono di ottenere una stima della quota di spesa corrente riferita dall'inefficienza tecnica il cui andamento può essere misurato oltre che a livello regionale anche lungo gli anni sui quali si estende la nostra analisi. Al riguardo, la figura 17 riporta la distribuzione dell'inefficienza tecnica totale in percentuale della spesa storica dal 1998 al 2012. Si può facilmente notare come l'inefficienza tecnica (in percentuale della spesa) mostri un trend decrescente soprattutto a partire dal 2003, passando dal 7% della spesa storica totale registrato nel 1998, al 3% registrato nel 2012. Questa ultima evidenza empirica mostra come gli interventi posti in essere soprattutto nella seconda metà degli anni 2000 abbiano consentito, oltre ad un rallentamento dei tassi di crescita della spesa corrente, anche un notevole recupero di efficienza grazie, in particolare, alla contrazione delle quantità degli input come riportato nella figura 14.

Successivamente, come descritto nella sezione 4.2.2, attraverso la stima degli effetti fissi regionali è possibile misurare l'inefficienza riferita ai prezzi dei fattori produttivi di ogni sistema regionale calcolando la distanza tra gli effetti fissi di ogni regione dall'effetto fisso della regione Umbria, che rappresenta il benchmark di riferimento per questa seconda componente di inefficienza in quanto per la regione Umbria è stato stimato l'effetto fisso più basso. Per costruzione la nostra misura di inefficienza di prezzo è time-invariant e quindi va interpretata come valore medio riferito a ogni sistema regionale lungo l'intero arco temporale della nostra analisi.

La figura 18 riporta la distribuzione dell'inefficienza sia tecnica che di prezzo in termini percentuali rispetto alla spesa storica dell'anno 2012 (ultimo anno della stima). L'inefficienza tecnica in termini monetari è, in media, pari al 3% della spesa storica ed è concentrata maggiormente nelle regioni meridionali. Lombardia e Veneto risultano le regioni più efficienti nel 2012 con valori di inefficienza tecnica prossimi allo zero; Calabria, Basilicata e Valle d'Aosta, invece, risultano le regioni più inefficienti con valori che superano il 5% della spesa. Per quanto riguarda invece l'inefficienza di prezzo la figura 18 mostra come tale quantità risulti essere in percentuale della spesa storica molto più importante di quella tecnica andando da

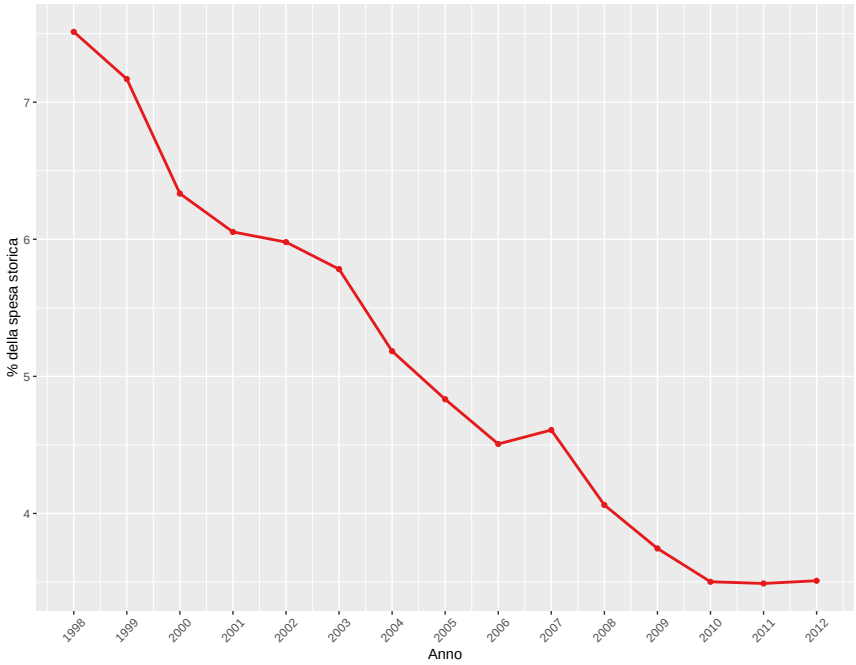


Figura 17: Distribuzione dell'inefficienza globale in percentuale della spesa storica.

valori che superano il 15% in quasi tutte le regioni meridionali, con punte del 25% in Calabria e Sardegna, a valori inferiori al 15% invece nelle regioni del centro-nord ad eccezione della Valle d'Aosta. Umbria e Marche sono le regioni che presenta valori prossimi allo zero in termini di inefficienza di prezzo e per questo motiva rivendica il suo ruolo di benchmark per le altre regioni. Da questi risultati è possibile concludere che per stimolare livelli di inefficienza più contenuti nella gestione dei servizi sanitari è necessario agire principalmente sul sistema dei prezzi piuttosto che sulle quantità degli input che risultano essere già oggi in proporzione adeguate rispetto agli output nella maggior parte dei sistemi sanitari regionali.

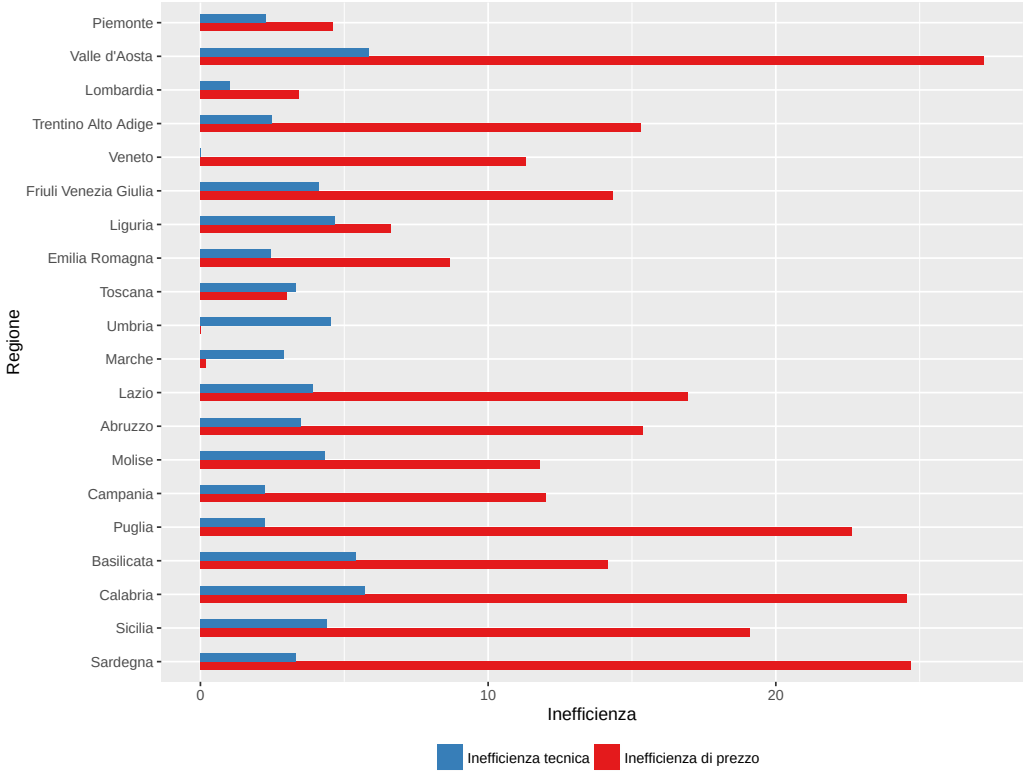


Figura 18: Inefficienza tecnica e di prezzo per Regione (% della spesa storica)

I coefficienti stimati della funzione di spesa riportata nella colonna 1 della tabella 7 possono essere utilizzati per valutare in termini monetari le due componenti dell'output gap complessivo relativo ad ogni sistema sanitario regionale oltre che alla sua evoluzione nel tempo.

La figura 19 mostra l'andamento nel tempo seguito dell'output-gap globale mediamente nelle regioni italiane in percentuale della spesa storica; si nota chiaramente un trend decrescente molto forte sino al 2006, che ha portato ad una stabile riduzione dell'output-gap passato dal 22% del 1998 al 13% del 2005. A partire dal 2006 questo trend decrescente sembra essersi arrestato essendo l'output gap rimasto costante ad un valore pari all'11% della spesa.

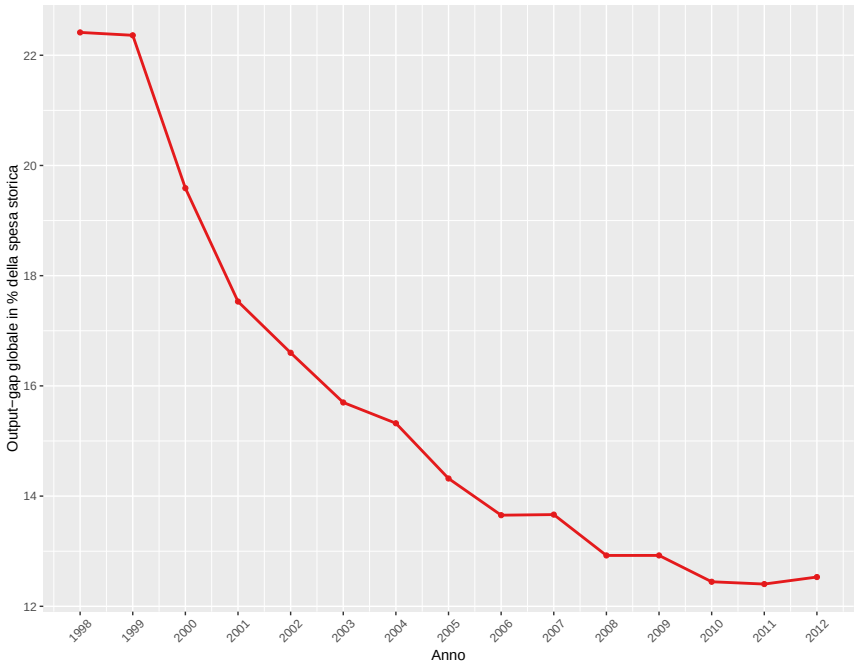


Figura 19: Distribuzione dell'output-gap globale in percentuale della spesa storica.

La figure 20 e 21 mostrano, rispettivamente in percentuale della spesa storica e in milioni di euro, la distribuzione dell'output-gap proprio e dell'output gap legato alla mobilità per regione nell'anno 2014 (ultimo anno rispetto al quale è possibile applicare le nostre stime). Le figure mostrano come, una volta isolato l'effetto dell'output gap legato alla mobilità, tutte le regioni presentano un output gap proprio negativo di una cerca consistenza, ad esclusione della Lombardia che presenta valori prossimi allo zero e che, per tale motivo, diventa la regione benchmark per misurare il livello dei servizi offerti. Questo mostra come nessuna regione, eccetto la Lombardia, soddisfi pienamente la domanda interna del proprio territorio erogando dei servizi sanitari ad un livello inferiore rispetto a quello compatibile con le determinanti della domanda interna avendo come punto di riferimento gli agli standard lombardi. In totale, con riferimento al 2012, la nostra simulazione valuta in oltre 10 miliardi di euro la

spesa necessaria a colmare l'output gap di tutte le regioni che può essere interpretata come una misura globale dell'ineguatezza dei servizi.

La valorizzazione dell'output-gap legato alla mobilità ricalca, invece, i flussi di mobilità inter-regionali che si muovono da sud verso nord.

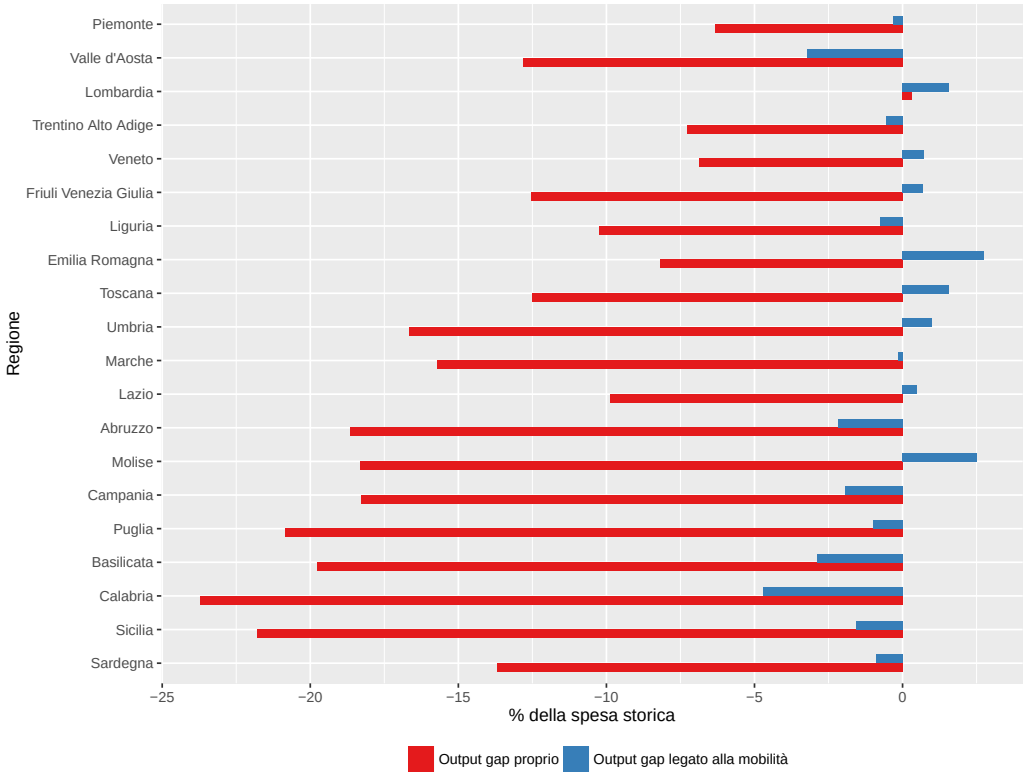


Figura 20: Output gap proprio e legato alla mobilità per Regione (% della spesa storica)

Sempre a partire dai risultati mostrati nella colonna 1 della tabella 7, a questo punto, si è calcolata la spesa standard per regione secondo le due specificazioni riportate nelle equazioni (14) e (15) della sezione 4.2.2: la prima - $\hat{H}^a_{it}$ - depurando la spesa storica di riferimento soltanto dall'inefficienza globale I_{it} senza tenere in considerazione l'adeguatezza dei servizi offerti lungo il territorio riconoscendo ad ogni regione una spesa standard compatibile con la quantità di servizi storicamente offerta; la seconda - $\hat{H}^b_{it}$ - prendendo in considerazione, invece, oltre all'inefficienza anche la quota di spesa necessaria a colmare l'output gap in modo da attribuire ad ogni sistema sanitario regionale un fabbisogno di spesa tale da poter

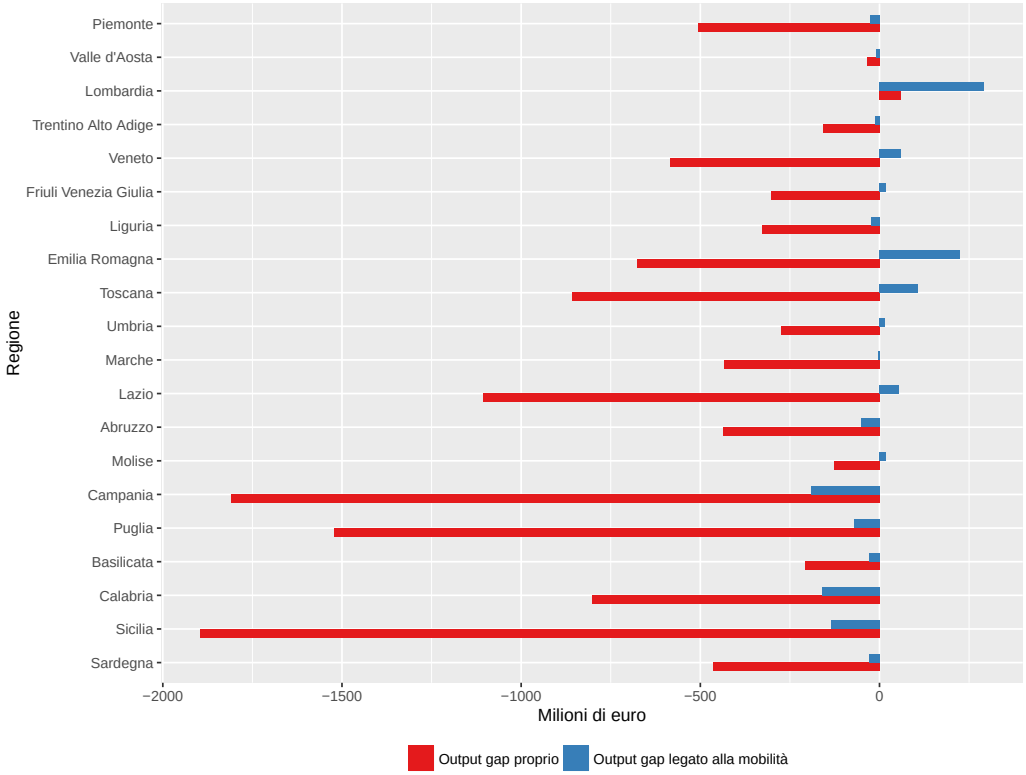


Figura 21: Output gap proprio e legato alla mobilità per Regione (mln di €)

offrire ad un costo efficiente il massimo livello dei servizi necessari a soddisfare appieno la domanda del territorio.

La figura 22 riporta di quanto le due versioni di spesa standard si discostano dalla spesa storica del 2014 in percentuale di quest'ultima. Le differenze sono maggiori e sempre negative quando nel computo della spesa standard si considera solo l'inefficienza globale, si riducono alla metà, invece, se nel calcolo della spesa standard si considera anche l'adeguatezza dei servizi misurata attraverso l'output gap totale. In questo secondo caso per Umbria, Marche, Toscana, Abruzzo, Campania e Basilicata la spesa standard supera la spesa storica grazie ad un bassa inefficienza combinata con un alto valore di output gap.

Da ultimo è importante notare come la dualità dell'economia italiana si rifletta pesantemente anche nelle differenze tra spesa storica e spesa standard dei sistemi sanitari regionali; le regioni del sud anche se contraddistinte da un output gap più marcato a causa dei più alti livelli di inefficienza mostrano sempre un differenziale tra spesa storica e spesa standard negativo più alto rispetto alle regioni del nord.

6.4 RIASSUNTO DEI RISULTATI E CONFRONTI CON LA SPESA STORICA

Le tabelle 8 e 9 che seguono, rispettivamente in termini assoluti e in percentuale della spesa storica di riferimento sintetizzano, in primo luogo, i risultati relativi alla stima della spesa standard e delle componenti relative all'inefficienza e all'output gap riferite al 2014, successivamente le due ipotesi di spesa standardizzata già discusse sopra sono confrontate con la spesa storica corrente dello stesso anno.

In estrema sintesi sono essenzialmente tre gli aspetti da mettere in risalto come risultati principali del nostro esercizio di standardizzazione. Il primo è legato alle due componenti che contribuiscono alla stima dell'inefficienza globale di ogni sistema regionale: al riguardo la colonna (C) riporta la spesa riferita al livello di inefficienza tecnica, che in totale ammonta a 3 mld di euro pari al 3,4% della spesa storica, mentre la colonna (D) mostra quella riferita al livello di inefficienza di prezzo, che in totale raggiunge i 12,5 mld di euro pari al 12,8%.

Il secondo aspetto si riferisce alle componenti che formano l'output gap, ovvero una misura dell'adeguatezza dei servizi erogati rispetto alla domanda di ogni territorio, che può essere anche interpretata come la capacità di ogni regione di massimizzare l'offerta dei servizi avendo come benchmark la regione Lombardia: al riguardo la colonna (F) riporta la spesa riferita l'output-gap proprio che in totale ammonta a 12,4 mld di euro corrispondente a circa il 13,7% della spesa storica, mentre la colonna (G) riporta quello legato alla mobilità che per costruzione come aggregato nazionale è molto vicino allo zero evidenziando un'ulteriore situazione deficitaria nell'erogazione dei servizi da parte delle regioni meridionali per oltre 600 milioni.

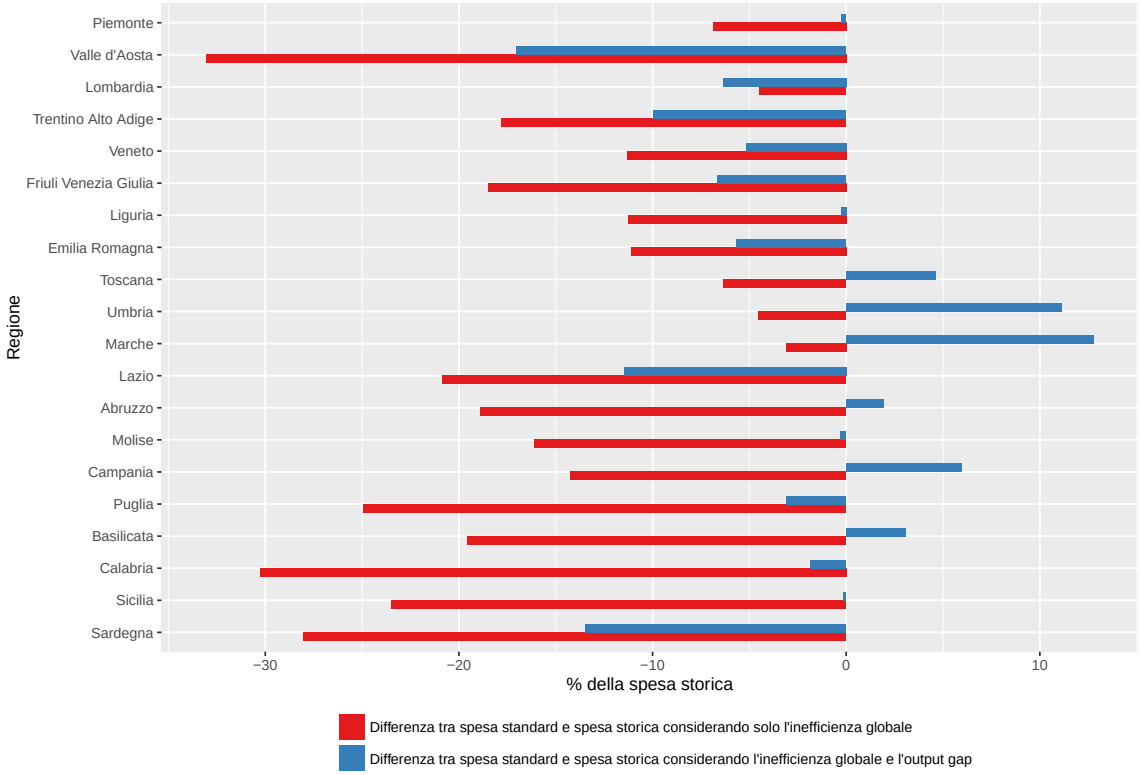


Figura 22: Differenze tra spesa standard e storica per Regione (% della spesa storica)

Il terzo ed ultimo aspetto riguarda la definizione di due ipotesi di spesa standardizzata. La prima ipotesi tiene conto solo del livello di inefficienza globale, ipotizzando che i servizi continuino ad essere erogati a livello storico senza alcuna correzione dell'output gap; in questo caso rispetto alla spesa storica di riferimento si ottiene (come riportato nella colonna (E)) una spesa comprimibile che supera i 15 miliardi pari al 16,2% della spesa storica. In questo caso i risparmi di spesa si concentrano prevalentemente nelle regioni meridionali caratterizzate dalle percentuali di inefficienza tecnica e di prezzo più alta, andando da una riduzione di spesa del 3% nelle Marche ad una riduzione di spesa che supera il al 30% in Calabria e Valle d'Aosta. Nella seconda definizione di spesa standard si considera, oltre al livello di inefficienza, anche il livello di adeguatezza dei servizi misurato in termini di output gap totale; in questo caso la quota di spesa comprimibile, riportata nella colonna (L), scende a 3 miliardi che corrispondono all'2,5% della spesa storica. Con questa seconda definizione di spesa standardizzata le differenze rispetto alla spesa storica, anche se continuano ad essere più concentrate nelle regioni meridionali, risultano meno difformi lungo la penisola.

Tabella 8: Spesa storica e standard per Regione, valori monetari (mld di €), anno 2014

Regione	Popolazione residente al 2014 (A)	Spesa storica 2014 (mld di €) (B)	Inefficienza tecnica (mld di €) (C)	Inefficienza di prezzo (mld di €) (D)	Inefficienza globale (mld di €) (E = C+D)	Output gap proprio (mld di €) (F)	Output gap legato alla mobilità (mld di €) (G)	Output gap totale (mld di €) (H = C+D)	Spesa standard al netto dell' inefficienza globale (mld di €) (I = B-E)	Spesa standard al netto dell' inefficienza globale considerando l'output gap (mld di €) (K = B-E-H)	Differenza tra spesa standard e spesa storica considerando l'inefficienza globale e l'output gap (mld di €) (L = K - B)
Piemonte	4.430.633	7.997	182	369	550	-506	-25	-531	7.447	7.978	-44
Valle d'Aosta	128.445	269	16	73	89	-34	-9	-43	180	223	-54
Lombardia	9.988.006	18.398	191	632	822	59	290	350	17.575	17.226	-882
Trentino Alto Adige	1.053.943	2.164	54	331	385	-157	-12	-170	1.779	1.948	-228
Veneto	4.927.207	8.504	0	963	963	-584	61	-523	7.541	8.064	-379
Friuli Venezia Giulia	1.228.243	2.407	100	346	445	-302	17	-285	1.962	2.247	-143
Liguria	1.587.601	3.197	149	211	360	-328	-24	-352	2.837	3.189	-32
Emilia Romagna	4.448.431	8.252	202	714	916	-676	226	-450	7.335	7.785	-241
Toscana	3.751.583	6.873	228	206	434	-859	107	-752	6.438	7.190	424
Umbria	895.752	1.657	75	0	75	-276	17	-260	1.582	1.842	201
Marche	1.551.967	2.764	80	5	86	-435	-4	-439	26.78	3.118	349
Lazio	5.881.438	11.198	440	1.897	2.336	-1.106	54	-1.052	8.862	9.914	-1.230
Abruzzo	1.332.757	2.336	82	359	441	-436	-51	-487	1.895	2.382	-5
Molise	314.037	699	30	82	113	-128	18	-111	586	697	15
Campania	5.865.747	9.907	221	1.191	1.412	-1.810	-191	-2.001	8.496	10.497	398
Puglia	4.090.186	7.293	165	1.653	1.817	-1.521	-72	-1.593	5.475	7.069	-296
Basilicata	577.505	1.056	57	150	207	-209	-30	-239	850	1.089	2
Calabria	1.978.582	3.385	194	831	1.025	-803	-159	-962	2.361	3.323	-222
Sicilia	5.093.509	8.695	382	1.661	2.042	-1.895	-136	-2.031	6.652	8.683	-147
Sardegna	1.663.573	3.395	113	838	951	-464	-30	-494	2.444	2.939	-487
Italia (totale)	**60.789.145**	**110.448**	**2.958**	**12.513**	**15.471**	**-12.470**	**45**	**-12.425**	**94.978**	**107.403**	**-3.000**

Tabella 9: Spesa storica e standard per Regione, valori percentuali (% della spesa storica), anno 2014.

Regione	Popolazione residente al 2014 (A)	Spesa storica 2014 (mld di euro) (B)	Inefficienza tecnica (C)	Inefficienza di prezzo (D)	Inefficienza globale (E = C+D)	Output gap proprio (F)	Output gap legato alla mobilità (G)	Output gap totale (H = C+D)	Differenza tra spesa standard e spesa storica considerando l'inefficienza globale e l'output gap (L = -E-H)
Piemonte	4.430.633	7.997	2,272	4,609	6,881	-6,327	-0,317	-6,644	-0,55
Valle d'Aosta	128.445	269	5,841	27,220	33,062	-12,824	-3,228	-16,052	-20,24
Lombardia	9.988.006	18.398	1,036	3,434	4,470	0,323	1,579	1,902	-4,79
Trentino Alto Adige	1.053.943	2.164	2,491	15,309	17,800	-7,274	-0,572	-7,847	-10,53
Veneto	4.927.207	8.504	0,000	11,324	11,324	-6,866	0,715	-6,151	-4,46
Friuli Venezia Giulia	1.228.243	2.407	4,135	14,354	18,489	-12,532	0,701	-11,831	-5,96
Liguria	1.587.601	3.197	4,664	6,608	11,271	-10,265	-0,745	-11,010	-1,01
Emilia Romagna	4.448.431	8.252	2,449	8,656	11,105	-8,188	2,735	-5,453	-2,92
Toscana	3.751.583	6.873	3,317	3,004	6,321	-12,497	1,560	-10,937	6,18
Umbria	895.752	1.657	4,529	0,000	4,529	-16,670	1,002	-15,668	12,14
Marche	1.551.967	2.764	2,907	0,194	3,101	-15,733	-0,158	-15,891	12,63
Lazio	5.881.438	11.198	3,925	16,937	20,862	-9,880	0,482	-9,397	-10,98
Abruzzo	1.332.757	2.336	3,490	15,386	18,876	-18,663	-2,168	-20,832	-0,21
Molise	314.037	699	4,316	11,793	16,110	-18,324	2,514	-15,810	2,21
Campania	5.865.747	9.907	2,226	12,022	14,248	-18,269	-1,933	-20,202	4,02
Puglia	4.090.186	7.293	2,258	22,661	24,919	-20,855	-0,989	-21,844	-4,06
Basilicata	577.505	1.056	5,394	14,177	19,571	-19,777	-2,872	-22,649	0,21
Calabria	1.978.582	3.385	5,717	24,548	30,265	-23,709	-4,701	-28,410	-6,56
Sicilia	5.093.509	8.695	4,392	19,099	23,491	-21,796	-1,563	-23,359	-1,70
Sardegna	1.663.573	3.395	3,334	24,683	28,018	-13,677	-0,886	-14,562	-14,34
Italia (media)			**3,435**	**12,801**	**16,236**	**-13,690**	**-0,442**	**-14,132**	**-2,546**

7 DISAGGREGAZIONE SPAZIALE DEI LIVELLI DI DOMANDA ED OFFERTA SANITARIA A LIVELLO COMUNALE

Indice

7.1 Domanda di ricerca 84
7.2 Metodi di disaggregazione spaziale di tipo Chow-lin 86
7.3 Una prima applicazione al caso italiano 89
7.3.1 Indicatore composito della domanda potenziale a livello provinciale 89
7.3.2 Identificazione e stima del modello a livello provinciale 91
7.3.3 Disaggregazione spaziale a livello comunale: un approccio di tipo Chow-Lin 94
7.3.4 Domanda ed offerta a livello comunale tra prossimità e specializzazione 95
7.4 Osservazioni conclusive 100

7.1 DOMANDA DI RICERCA

La presenza di disuguaglianze sanitarie all'interno di un paese può essere attribuito al prodotto dei differenziali di reddito e accesso ai servizi sanitari associato a disuguaglianze strutturali che influenzano gli ambiti nei quali le persone vivono e lavorano. Questo capitolo, originale rispetto alle precedenti versione del rapporto SaniRegio, sottolinea la necessità di studiare la domanda di salute a un livello disaggregato più basso rispetto a quello regionale al fine di avere uno strumento più adatto per riallineare la relativa offerta.
Il SSN, come già evidenziato nel paragrafo 2.2, ha subito importanti riforme nel corso degli ultimi due decenni, tra le quali un progressivo e significativo spostamento di giurisdizione dal Centro alle Regioni. Dall'inizio degli anni '90, ai governi regionali sono stati concessi ampi margini di discrezionalità nella pianificazione e nell'organizzazione dei servizi sanitari nel proprio territorio. L'autonomia amministrativa regionale ha provocato la nascita di 20 "sistemi regionali diversi" (Mapelli, 2012) piuttosto che un singolo servizio sanitario nazionale.
Negli ultimi anni, il SSN sta subendo un graduale processo di revisione della spesa, diverso da Regione a Regione, con l'intento di razionalizzare beni e servizi attraverso l'adozione di un modello di "quasi-market" nell'organizzazione. Con l'obiettivo di razionalizzare le spese, molte ASL sono stati fuse; si pensi, ad esempio, che nel 2017, dopo le ultime riforme, 97 ASL e 99 Aziende ospedaliere autonome (grandi ospedali, policlinici e IRCCS) erano presenti nei

territori rispetto ai 642 disponibili a metà degli anni '80. In particolare, la Toscana e la Lombardia nel 2015 hanno approvato due coraggiose riforme sanitarie riducendo drasticamente il numero di ASL, rispettivamente da 12 a 3 e da 15 a 8. Il Lazio è passato da 12 a 10 e il Veneto ha approvato il passaggio da 21 a 9 ASL. La Sardegna ha creato una sola ASL come fusione delle precedenti 8 così come la Liguria che similmente ha optato per una sola ASL per l'intera Regione.
Contemporaneamente a questo sforzo di razionalizzazione, però, tali riforme sono state attuate senza un adeguato sistema di monitoraggio dei bisogni locali di salute o l'identificazione di zone ottimali di offerta/domanda di assistenza sanitaria (Atella *et al.*, 2014); tutto ciò ha portato il sistema ad allontanarsi da un'allocazione efficiente delle risorse e dei servizi che possano garantire un uniforme livello di assistenza su tutto il territorio nazionale.
Tuttavia l'obiettivo di tale capitolo non è tanto quello di offrire un modello per il dimensionamento ottimale locale delle ASL (che in Italia rimane una prerogativa dei singoli governi regionali che come abbiamo detto seguono altre logiche) nè di stimare coefficienti intra-regionali di riparto delle risorse destinate dallo Stato alle Regioni. L'obiettivo di questa sezione è invece quello di evidenziare gli squilibri delle risorse territoriali tramite l'ausilio di indicatori specifici; da un lato, quindi, si propone un indicatore composito della domanda di assistenza sanitaria a livello provinciale, dall'altro, si stima un indicatore spazialmente disaggregato a livello comunale utilizzando driver della domanda locale con l'obiettivo di identificare le esigenze sanitarie e evidenziare eventuali squilibri territoriali tra l'offerta e la domanda locale.
In altri termini, l'obiettivo principale di questo capitolo è quello di individuare squilibri a livello di *small area* (Comuni) nella domanda di prestazioni sanitarie in relazione all'offerta. Questo livello di disaggregazione sembra in prima istanza essere il più corretto perché, come affermato da Royston *et al.* (1992), l'utilizzo di aree omogenee più grandi può comportare l'identificazione di relazioni spurie tra le variabili (*the ecological fallacy*); nel contempo, dati molto disaggregati ed adeguati alla nostra analisi non sono, di solito, disponibili da fonti ufficiali[1]. L'utilizzo di informazioni più dettagliate inoltre può essere molto più informativo anche rispetto al problema dell'allocazione delle risorse dallo Stato alle Regioni; un livello più alto (ad esempio regionale), può essere fuorviante perché in gran parte determinato dal lato della domanda. A livello di *small area*, invece, diversi sottoinsiemi di popolazione concorrono per i servizi ed il loro diverso utilizzo relativo (dopo aver normalizzato per i fattori disponibili) può meglio contribuire a spiegare la variabilità territoriale dei fabbisogni sanitari.

1 Anche se possono essere disponibili a livello amministrativo presso il Ministero della Salute, come ad esempio nel caso italiano.

7.2 METODI DI DISAGGREGAZIONE SPAZIALE DI TIPO CHOW-LIN

Come evidenziato nel paragrafo 4.1.1, gli indicatori compositi possono aiutare a studiare un fenomeno sociale (Nardo *et al.*, 2005) ad un determinato livello territoriale, ma sono necessariamente insufficienti per fornire informazioni ad una risoluzione più fine.
Al fine di derivare gli indicatori a livello territorialmente disaggregato dai corrispondenti indicatori di livello aggregato[2], si propone un modello basato sull'approccio di Chow & Lin (1971) che si fonda, come è noto, su tre assunzioni principali:

1. *Structural similarity*: il modello regressivo aggregato è strutturalmente simile a quello disaggregato; ciò implica che le relazioni tra le variabili osservate a livello aggregato sono equivalenti a quelle a livello disaggregato, cioè i parametri di regressione sono uguali o simili in entrambi i modelli.

2. *Error similarity*: gli errori spazialmente correlati presentano la stessa struttura sia a livello aggregato che a livello disaggregato; questo significa che la correlazione spaziale non è significativamente diversa tra i due modelli.

3. *Reliable indicators*: le covariate di regressione sono significative, ovvero hanno un sufficientemente grande potere predittivo in entrambi i modelli; in altri termini le misure di bontà della regressione (R^2 o F test) risultano significativamente diverse da zero.

Da notare che la violazione dell' ipotesi (1) porta ad ottenere stime sistematicamente distorte; la violazione dell'ipotesi (2), invece, comporta errori di valutazione circa i potenziali effetti di spillover che contribuiscono in larga misura alle stime, mentre il mancato rispetto dell'ipotesi (3) implica che le stime disaggregate rispecchieranno solo la semplice proporzione di quelle aggregate.
Modelli di questo tipo sono stati prevalentemente utilizzati per la costruzione di serie temporali mensili o trimestrali a partire da corrispondenti serie annuali, ma più recentemente sono stati impiegati anche in alcune applicazioni di disaggregazione spaziale. In Italia,Bollino & Polinori (2007), stimano il valore aggiunto a livello comunale in Umbria al fine di studiare la convergenza tra comuni suburbani e comuni urbani che beneficiano di una maggiore crescita spiegato da fattori di contiguità e da meccanismi di agglomerazione.

Più specificatamente, il modello base è caratterizzato sia da una relazione econometrica tra l'indicatore composito a livello provinciale e le rispettive covariate osservabili a livello disaggregato (e, ovviamente anche a livello aggregato) e da una metodologia di inferenza

2 Bene precisare che in questo capitolo il termine "indicatore aggregato" fa riferimento ad una misura di livello territorialmente superiore (regionale, ad esempio) rispetto ad analoga misura di livello inferiore (provinciale, ad esempio), e non deve pertanto essere confuso con il termine "indicatore sintetico" (o "composito"), che rappresenta invece il risultato di un'operazione di aggregazione di indicatori semplici considerati sempre allo stesso livello territoriale.

dei parametri incogniti a livello provinciale. Il modello si basa sul presupposto che, a livello disaggregato, sia fondata una relazione econometrica lineare (non direttamente stimabile perché non è noto y_d):

$$y_d = X_d \beta_d + \epsilon_d \tag{17}$$

dove il pedice d (Comuni nella presente applicazione, sezione 7.3) indica i termini disaggregati.

Denominando come C la matrice di dimensione $n * N$ (con n il numero di Comuni italiani e N il numero di Province) che permette di trasformare le osservazioni disaggregate in aggregate e con a il pedice che indica i termini aggregati, la relazione lineare aggregata tra la y e le covariate può essere scritta come:

$$y_a = CX_d \beta_a + \epsilon_a \tag{18}$$

In particolare, se si adotta l'operatore somma, le stime provinciali sono ottenute per somma dei corrispondenti livelli comunali ($y_a = \sum y_d$); in questo caso, l'elemento generico C_{ij} sarà costruito come:

$$C_{ij} = \begin{cases} 1, & \text{se il Comune } i \in \text{provincia } j, \\ 0, & \text{altrimenti.} \end{cases} \tag{19}$$

Se invece si adotta l'operatore media aritmetica C deve essere costruita come:

$$C_{ij} = \begin{cases} 1/k, & \text{se il Comune } i \in \text{provincia } j, \\ 0, & \text{altrimenti.} \end{cases} \tag{20}$$

dove k è il numero dei Comuni appartenenti alla Provincia j; in questo caso, le stime provinciali saranno ricostruite attraverso la media di quelle comunali ($y_a = E(y_d)$).

Date queste premesse e alla luce del vincolo $y_a = Cy_d$, della *structural similarity* ($\beta_d = \widehat{\beta}_a$) e dell'ipotesi di *error similarity* ($\sigma_a^2 = \widehat{\sigma}_d^2$), la previsione ottimale (Bollino & Polinori, 2007) di y_d può essere scritta come:

$$\widehat{y}_d = X_d \widehat{\beta}_a + GU \tag{21}$$

dove:

$$G = \widehat{\sigma}_a^2 C' (C \widehat{\sigma}_a^2 C')^{-1} \text{ e } U = (y_a - X_a \widehat{\beta}_a) \tag{22}$$

Recentemente Polasek & Sellner (2008) hanno proposto[3] una generalizzazione molto interessante del modello delineato nelle equazioni (21) e (22) introducendo un termine di auto-correlazione spaziale nell'equazione multivariata classica (17).

3 Gli autori ringraziano il Prof. Sellner per la sua routine in `Matlab`; il relativo codice `R` verrà reso disponibile dagli autori su richiesta.

Da un punto di vista applicativo, ciò significa che il valore della variabile dipendente in una determinata area dipende non solo dalle proprie variabili indipendenti, ma anche dal livello della variabile dipendente nelle aree limitrofe. Supponendo che esistano, e siano significativi, effetti di correlazione spaziale nei livelli competitivi tra Comuni, ma anche e soprattutto nelle Province, è possibile ipotizzare, a livello disaggregato, un modello di tipo "*mixed regressive spatial auto regressive*" (LeSage & Pace, 2009), (ancora non direttamente stimabile) dove W_N è la matrice dei pesi spaziali e $\rho \in [0, 1]$ un parametro di lag spaziale):

$$y_d = \rho_d W_N y_d + X_d \beta_d + \epsilon_d, \ \epsilon_d \sim N(0, \sigma_d I_N) \tag{23}$$

La forma ridotta consente di valutare al meglio la componente spaziale che impatta sui livelli di X_d:

$$y_d = (I - \rho_d W_N)^{-1} X_d \beta_d + (I - \rho_d W_N)^{-1} \epsilon_d \tag{24}$$

È quindi possibile riscrivere la forma ridotta dell'equazione (24) con $R_N = (I - \rho_d W_N)$:

$$y_d = R_N^{-1} X_d \beta_d + R_N^{-1} \epsilon_d, \ \epsilon_d \sim N(0, \Sigma_d) \tag{25}$$

e con la matrice di covarianza Σ_d pari a:

$$\Sigma_d = \sigma_d (R_N' R_N)^{-1} \tag{26}$$

I termini da stimare dei modelli a livello disaggregato sono ρ_d, β_d e la covarianza σ_d. Stimare questi termini, significa, date le ipotesi di base, stimare la relazione autoregressiva che intercorre tra y e X a livello aggregato:

$$y_a = \rho_a W_N y_a + C X_d \beta_a + \epsilon_a \tag{27}$$

con lo scopo di ottenere $\widehat{\rho}_a$ e $\widehat{\sigma}_a$.

Data l'ipotesi di *structural similarity* ($\rho_d = \widehat{\rho}, \beta_d = \widehat{\beta}_a$) e di *error similarity* ($\sigma_d = \widehat{\sigma}_a$) è possibile sostituire i parametri stimati nell'equazione (25) e (26).

La stima di β_a, similimente al metodo base di tipo Chow-Lin, può essere ottenuta come:

$$\widehat{\beta}_{a,GLS} = (X_a'(C\widehat{\Sigma}_d C')^{-1} X_a)^{-1} X_a'(C\widehat{\Sigma}_d C')^{-1} y_a \tag{28}$$

e la stima di y_d a livello disaggregato può essere dunque costruita come:

$$\widehat{y}_d = \underset{(1^{st}\text{ term})}{\widehat{R}_N^{-1} X_d \widehat{\beta}_a} \ \underset{(2^{nd}\text{ term})}{+\widehat{\Sigma}_d C'(C\widehat{\Sigma}_d C')^{-1}} \ \underset{(3^{rd}\text{ term})}{(y_a - C\widehat{R}_N^{-1} C' X_a \widehat{\beta}_a)} \tag{29}$$

Si noti come il primo termine dell'equazione (29), quindi, rappresenta la stima naive del vettore sconosciuto y_d, mentre il terzo termine dell'equazione rappresenta l'errore stimato a livello aggregato moltiplicato per il secondo termine, la "*gain projection matrix*" (Goldberger,

1962).
Tale correzione, dunque, dipende in modo cruciale dal lag spaziale $\widehat{\rho}_a$ a livello aggregato; si noti che se $\widehat{\rho}_a = 0$ la matrice $\widehat{\Sigma}_d$ è uguale alla matrice Identità e si riduce quindi alla trasposta della matrice di proiezione: $G = C'(CC')^{-1}$ come nel modello di base; il parametro ρ_d e la matrice W_N permettono quindi che la $1/N$ parte del residuo a livello aggregato non venga assegnato in parte uguale a tutti i Comuni, ma venga filtrato attraverso la distanza spaziale tra essi.

7.3 UNA PRIMA APPLICAZIONE AL CASO ITALIANO

7.3.1 Indicatore composito della domanda potenziale a livello provinciale

Al fine di stimare il livello di domanda a livello comunale è necessario procedere con la stima della domanda a livello provinciale per poi - con il metodo esposto nel paragrafo precedente -alla disaggregazione spaziale; in particolare, l'indicatore composito della domanda sanitaria a livello provinciale è stimato a partire dai dati "Health for All" di fonte ISTAT[4] per l'anno 2013.

La scelta delle dimensioni di analisi è purtroppo e chiaramente sempre limitata dalla disponibilità di dati statistici: in particolare in questa applicazione solamente i tassi standardizzati di mortalità (per 10.000 abitanti) - divisi per causa di morte - sono disponibili a livello provinciale. Date queste premesse, l'indicatore composito risultante deve essere letto più come un indicatore di "*rischio sanitario*" che come un indicatore di domanda di "assistenza sanitaria" tout court. Come meglio esplicitato in Appendice, all'interno del framework BoD l'indicatore composito è costruito come aggregazione lineare ponderata applicata agli indicatori semplici. L'ipotesi di linearità della forma aggregativa, tuttavia, può essere accolta positivamente soltanto seguendo il teorema di Krantz *et al.* (1971) (Nardo *et al.*, 2005, pp. 103) che dimostra come, date le variabili $(I_1, I_2, \ldots, I_n)$, esiste una funzione di aggregazione additiva se, e soltanto se, questi indicatori sono reciprocamente preferenzialmente indipendenti.
Da un punto di vista operativo, il teorema sottolinea che una funzione di aggregazione additiva consente la valutazione del contributo marginale di ciascuna variabile separatamente e il contributo marginale di ciascuna variabile può essere successivamente utilizzato per produrre il valore totale; questo, quindi, implica che l'aggregazione lineare è del tutto appropriata solo nel caso in cui ogni indicatore semplice non sia solo non collineare, ma anche indipendente dagli altri.
Date queste premesse, da un punto di vista statistico, questo implica che è necessario ricondurre gli indicatori a fattori principali soprattutto se gli indicatori semplici stessi sono fortemente correlati tra loro.

4 http://www.istat.it/it/archivio/14562.

La tabella 10 mostra l'ACP ruotata[5] relativa ai tassi di mortalità provinciali per tipologia di malattia.

Malattie	Fattore #1	Fattore #2	Fattore #3	Fattore #4
Tumori	89 *			
Polmonite	86 *			
Tumori apparato digerente	81 *			
Malattie infettive	74 *			
Tumori alla mammella	74 *			
Respiratorio	74 *	43		
Disturbi mentali	67 *			
Tumori al colon	66 *	47		
Sistema nervoso	58			
Sistema muscolare	55			
HIV/AIDS	53			
Pelle e tessuti	45			
Sistema circolatorio		84 *		
Ghiandole endocrine, nutrizione e metabolismo		81 *		
Disturbi circolatori cerebrali		79 *		
Diabete	-50	75 *		
Sistema genitale		68 *		
Disturbi cronici polmonari ostruttivi		64 *		
Cardiopatia ischemica		56		
Incidenti stradali			80 *	
Cause esterne di infortuni e avvelenamenti			71 *	
Suicidi ed auto lesioni	47		51	
Cirrosi e altre malattie croniche del fegato				77 *
Sistema digestivo	40			76 *
Sangue, organi ematopoietici e disturbi del sistema immunitario		42		50

Tabella 10: Indicatore composito di domanda a livello provinciale: analisi di tipo ACP

Quattro fattori principali risultano significativi (spiegando cumulativamente il 64% della varianza totale): in particolare, il fattore 1 può essere interpretato come correlato all'inquinamento ambientale e ai temi sociali; il fattore 2 legato ai disturbi alimentari e metabolici; il fattore 3 a cause non sanitarie (incidenti stradali e suicidi o auto lesioni), mentre il fattore 4 all'abuso di alcool.
La figura 23 (a) mette in evidenza evidenti regolarità spaziali nella direzione Nord-Sud, soprattutto nella zona Nord-Ovest tra Liguria e Piemonte[6] nonché per il fattore 2 (figura 23(b));

5 I valori in tabella sono moltiplicati per 100 e arrotondati al numero intero più vicino. I valori superiori a 60 sono contrassegnati con un '*'. Valori inferiori a 40 non vengono riportati; il metodo di rotazione è il Varimax con normalizzazione di Kaiser.

6 In questa zona, chiamata "*triangolo industriale*" l'industrializzazione dell'economia italiana si sviluppò su larga scala tra la fine del XIX e l'inizio del XX secolo.

gli altri due fattori non sembrano essere correlati a specifici fattori regionali, ma piuttosto con le caratteristiche di ogni singolo territorio.

Dopo aver ottenuto i fattori principali non correlati, la stima dell'indicatore composito è effettuata utilizzando il metodo BoD, il Robust Benefit of Doubt (RBoD) ed il metodo Method of Penalties for Coefficient of Variation (MPCV) (si veda l'Appendice per maggiori dettagli). Il coefficiente di correlazione di Spearman (tabella 11) tra le differenti stime mostra una buona stabilità nonostante le differenti proprietà alla base dei metodi proposti.

	BoD	Robust BoD	MPI
BoD	1.00	0.93	0.70
Robust BoD	0.93	1.00	0.72
MPI	0.70	0.72	1.00

Tabella 11: Indicatore composito di domanda a livello provinciale: metodi a confronto

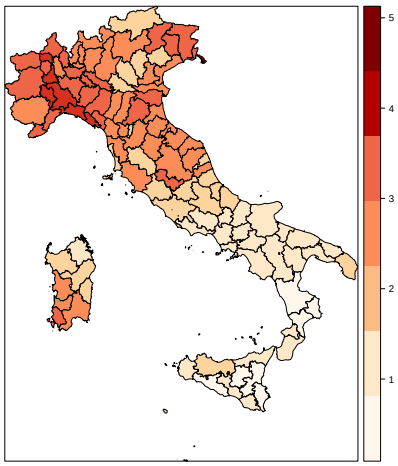
Dopo aver verificato la mancanza di variazioni significative nelle stime al variare del modello aggregativo, si è deciso di proseguire l'analisi utilizzando il metodo RBoD per le proprietà di robustezza alla base del metodo e per l'endogeneità dei pesi utilizzati. La distribuzione territoriale dell'indicatore RBoD non sembra mostrare una chiara regolarità Nord-Sud (figura 24), anche se si possono facilmente notare alcune regolarità locali trasversali alle Regioni (si veda ad esempio il quadrante Liguria - Piemonte, la Toscana, l'Appennino e la zona vicino a Venezia - petrolchimico di Porto Marghera). Poste queste premesse, è necessario disaggregare l'indicatore provinciale a livello comunale per meglio capire l'esistenza di aree locali omogenee ed analizzare le possibili determinanti.

7.3.2 Identificazione e stima del modello a livello provinciale

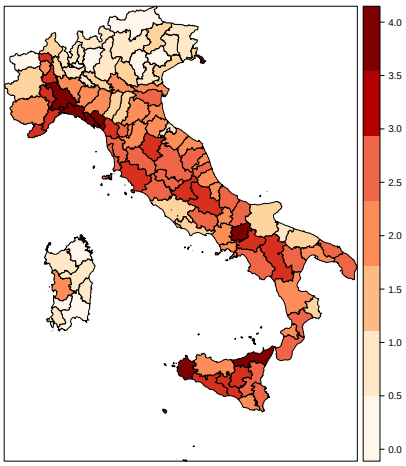
Come evidenziato nel paragrafo 7.2, nel framework di tipo Chow-lin una particolare enfasi è posta sulla parte deterministica del modello, mentre l'errore stocastico è ripartito tra i territori attraverso ρ e W.

La tabella12 presenta le stime OLS e SAR della relazione che intercorre tra l'indicatore robusto di tipo RBoD e le determinanti della domanda a livello provinciale: più in particolare, le colonne da (1) a (4) riportano le stime OLS, la colonna (5) i coefficienti di regressione standardizzati in base al modello OLS completo di colonna (4) - e la colonna (6) le stime del modello autoregressivo spaziale.

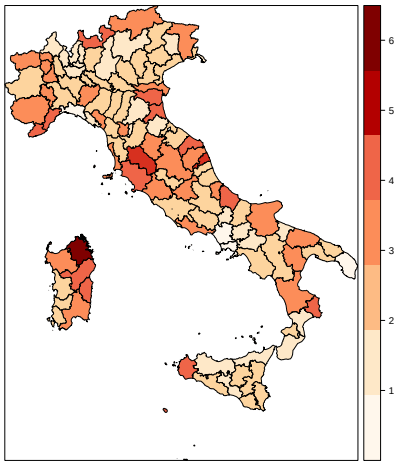
Come evidenziato nel paragrafo 7.2 le caratteristiche della popolazione, le caratteristiche fisiche e sociali della zona (altitudine, reddito, percentuale di laureati e persone in cerca di occupazione) e dei fattori di rischio locali (feriti in incidenti stradali) devono essere disponibili sia a livello provinciale sia a livello comunale per poter stimare successivamente l'indicatore



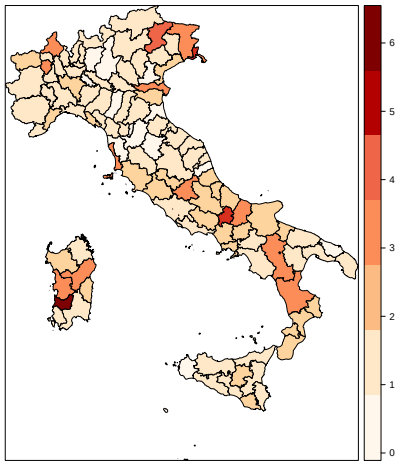
(a) *Fattore #1.*



(b) *Fattore #2.*



(c) *Fattore #3.*



(d) *Fattore #4.*

Figura 23: Indicatore composito di domanda a livello provinciale: analisi di tipo ACP

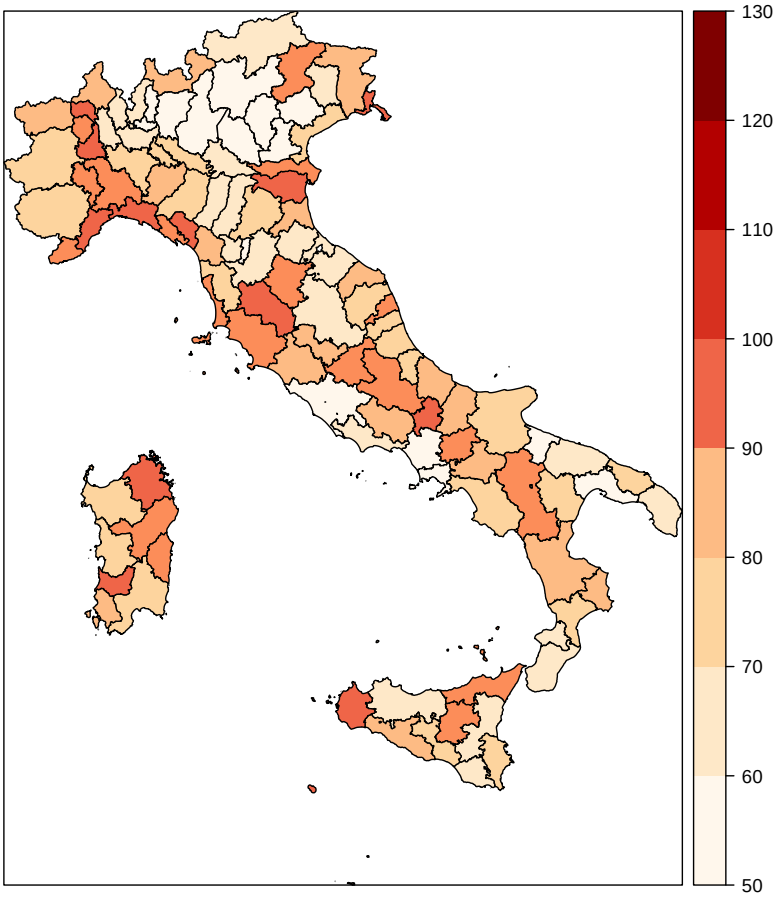


Figura 24: Indicatore composito RBoD di domanda a livello provinciale

composito a livello disaggregato. Tutte le covariate sono - in questo esercizio - espresse come medie provinciali delle corrispondenti variabili a livello comunale.

	Variabile dipendente: CI di tipo BoD robusto					
	(1)	(2)	(3)	(4)	(5)	(6)
Popolazione (55-64 anni)	−0.019	−0.018	−0.076***	−0.085***	-5.606	-0.074***
Popolazione (65-80 anni)	−0.011	−0.011	0.033*	0.037**	3.206	0.036**
Popolazione (>80 anni)	0.059**	0.060**	0.058**	0.057**	2.008	0.042*
Altitudine		0.014*	0.011*	0.011*	0.148	0.003
Reddito medio			−0.001**	−0.001**	-0.275	-0.001
Numero di laureati			−0.009*	−0.014**	-0.871	-0.010**
Persone in cerca di occupazione			0.022*	0.027**	0.869	0.023***
Feriti in incidenti stradali				0.168**	0.523	0.034
Intercetta	80.763***	74.696***	100.110***	99.701***	0.000	72.859***
R^2	0.230	0.260	0.489	0.515	-	-
R^2corretto	0.205	0.227	0.448	0.471	-	-
ρ	-	-	-	-	-	0.2714**
AIC	764.17	762.33	732.8	729.68	-	859.09
Note:						*p<0.1; **p<0.05; ***p<0.01

Tabella 12: Indicatore BoD robusto e caratteristiche della domanda a livello provinciale: modelli a confronto

Si noti, in particolare, le stime del modello completo - colonna (4): *(i)* il coefficiente della popolazione è negativo per la classe di età più giovane, mentre è positivo e crescente al crescere dell'età; *(ii)* il reddito e il numero dei laureati mostrano coefficienti negativi (effetto di sostituzione tra beni pubblici e privati quale effetto di un maggior livello culturale sulla prevenzione), mentre nelle zone più povere del paese (persone in cerca di occupazione) il rischio sanitario aumenta; infine, *(iii)* una maggiore rischiosità delle infrastrutture stradali (feriti in incidenti stradali) aumenta in maniera diretta la domanda di cure sanitarie.
Infine, le stime SAR (colonna 6, ρ positivo e significativo) mettono in evidenza la necessità di includere effetti spaziali nel modello di disaggregazione territoriale.

7.3.3 Disaggregazione spaziale a livello comunale: un approccio di tipo Chow-Lin

Partendo dalla relazione stimata tra l'indicatore composito robusto e le caratteristiche della domanda a livello provinciale, il metodo Chow-Lin può essere utilmente utilizzato per stimare il relativo indicatore a livello comunale.
Secondo i modelli stimati nel precedente paragrafo - nei quali le covariate erano costruite come media provinciale - la matrice C è stata costruita secondo l'equazione (20).

La figura 25 mostra come l'eterogeneità sia molto significativa all'interno di ciascuna Provincia, mettendo in evidenza - ancora una volta - la necessità di scendere ad un maggiore livello di dettaglio per meglio comprendere le regolarità spaziali della domanda sanitaria.

La figura 26 permette di indagare la domanda stimata per le Regioni del Nord-Ovest: è facile osservare come, oltre alle evidenti regolarità cross-regionali, emergano anche fattori di rischio specifici che potrebbero, essere legati a fattori industriali (come per Genova).

Il rapporto tra la stima naive e quella totale (vedi equazione 29) può rappresentare una buona misura relativa alla precisione delle stime disaggregate; più si avvicina ad 1 più l'ipotesi (3) (*Reliable indicators*) è soddisfatta e meno la stima disaggregata è influenzata dall'errore spaziale non spiegato. In questa applicazione il rapporto non è mai inferiore a 0,85 - per ogni Comune - a dimostrazione di come la stima disaggregata sia fortemente caratterizzata dalla componente deterministica del modello.

7.3.4 Domanda ed offerta a livello comunale tra prossimità e specializzazione

L'indicatore di domanda stimato a livello comunale è particolarmente utile al *policy maker* locale (Regionale, nel caso italiano) a cui consente di studiare in dettaglio le caratteristiche del proprio territorio; tuttavia, tale strumento di policy può essere sicuramente più utile se legato all'analisi dell'offerta sul territorio.
Lo scopo di questa sezione, quindi, è quello di legare domanda e offerta a livello comunale precisando come sia evidente che esistano interdipendenze tra le due parti e non sia possibile - almeno in un cross-sectional setting - analizzare correttamente la causalità tra domanda e l'allineamento della offerta.
Viceversa, pensiamo sia utile capire se ad un tempo dato la localizzazione delle strutture di cura sia appropriata e se la domanda di salute cresca al crescere della distanza dalle stesse.
I dati ufficiali di offerta di assistenza sanitaria[7] permettono di ricostruire l'elenco completo delle strutture sanitarie e ospedaliere locali a livello nazionale comprensivo degli indirizzi. L'indirizzo georeferenziato tramite la Google Maps API Geocoding ha permesso di sovrapporre la mappa dell'offerta sanitaria con la mappa della domanda stimata a livello comunale (figura 27).

Anche se ancora in modo qualitativo, la figura 27 mette in luce un risultato importante di questa ricerca (ancora una volta reso possibile attraverso l'analisi della domanda e dell'offerta ad un livello inedito): il rischio sanitario aumenta con la distanza fisica dalle strutture sanitarie. La figura 28 mette un focus particolare sull'offerta territoriale e la domanda stimata per le Regioni nord-occidentali italiane; il disallineamento tra il rischio sanitario e l'offerta appare evidente dato che le strutture sanitarie (punti neri) si trovano in zone di colore arancione (basso rischio per la salute) mentre il rischio aumenta all'aumentare della distanza dalla struttura (zone rosse). Quale può essere, quindi, la ragione di questa evidenza? Una

7 Ottenuti a partire dall'"*Elenco delle Aziende sanitarie locali e Strutture di ricovero*", fonte: Nuovo Sistema Informativo Sanitario (NSIS), Ministero della Salute, anno 2013.

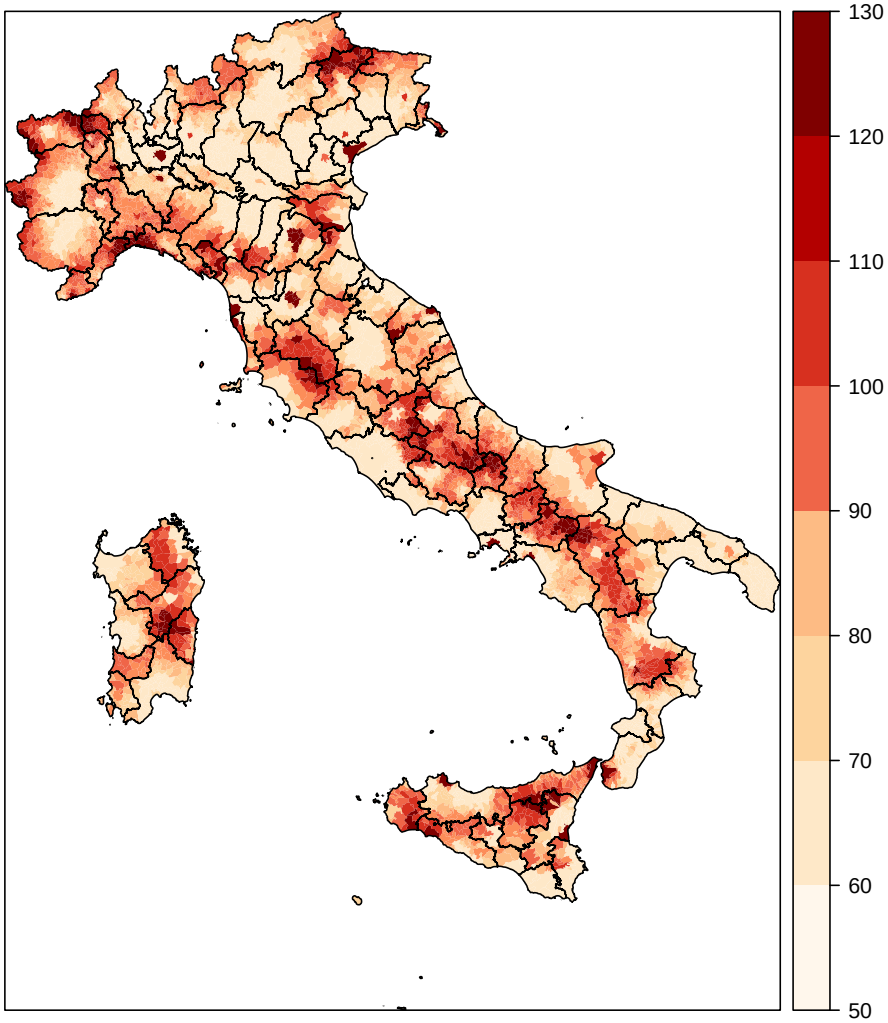


Figura 25: Stima dell'indicatore composito RBoD di domanda a livello comunale: metodo spaziale di tipo Chow-Lin

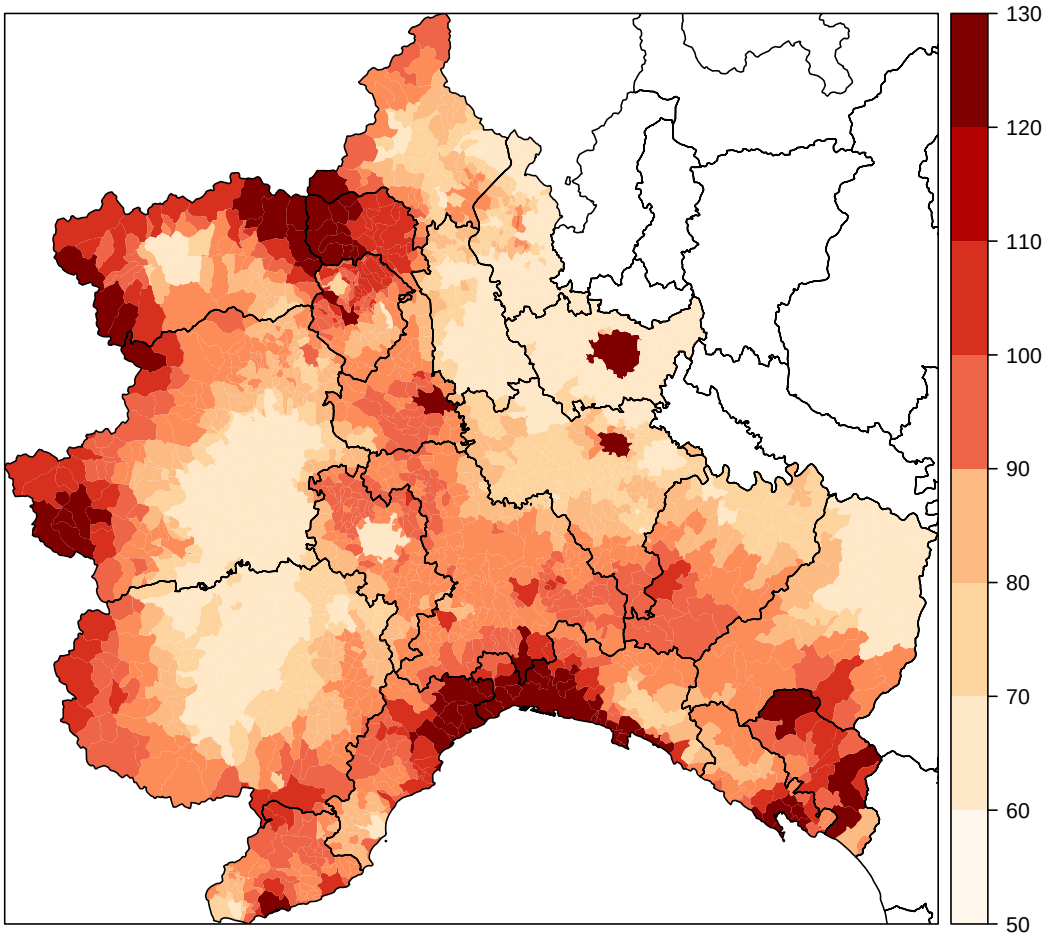


Figura 26: Stima dell'indicatore composito RBoD di domanda a livello comunale: Regioni Nord-Occidentali

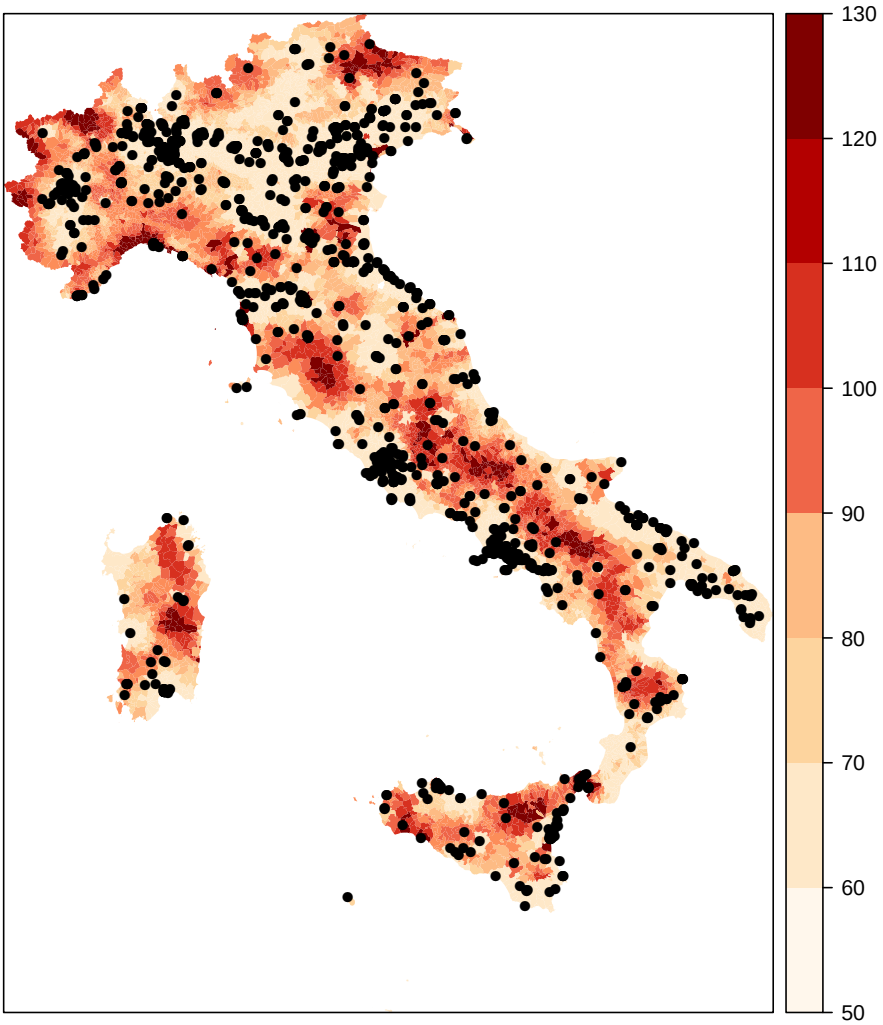
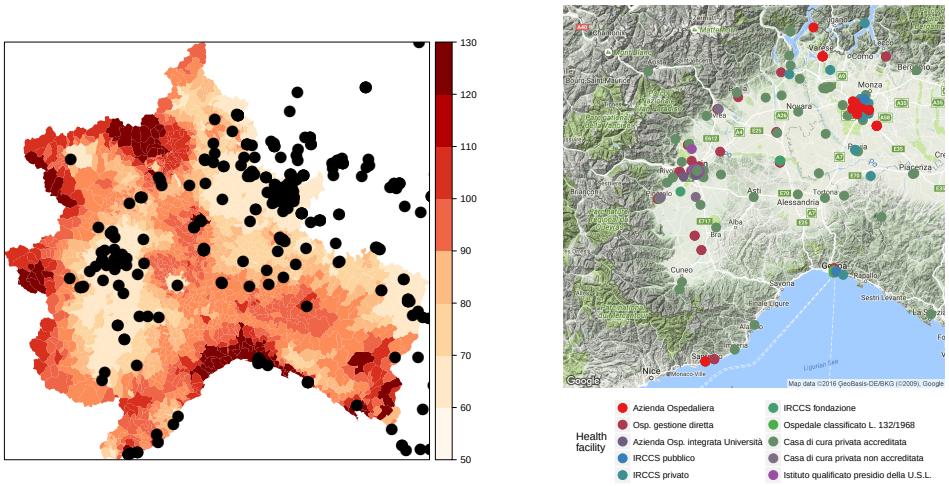


Figura 27: Offerta e domanda stimata a livello comunale

possibile spiegazione - a nostro avviso - potrebbe risiedere nei maggiori costi privati per i residenti delle aree periferiche, montane e/o povere che scoraggiano l'accesso alle strutture sanitarie e diminuiscono l'accesso a forme di medicina preventiva e/o screening.



(a) *Domanda e offerta.*

(b) *Offerta e orografia.*

Figura 28: Offerta e domanda stimata a livello comunale: Regioni Nord-Occidentali

Per fornire stime più precise dell'associazione tra la posizione della struttura e l'indicatore di domanda composito, si procede calcolando la distanza[8] dal centroide di ogni Comune alla posizione di ogni altra struttura sanitaria. Dopo aver calcolato tutte le distanze, quella più breve è scelta come riferimento per ciascun Comune. La figura 29 mostra[9] la correlazione positiva tra la domanda stimata e la distanza minima dalla struttura sanitaria più vicina; questa correlazione positiva, pari a = 0.303, sembra diminuire e non essere più significativa dopo 25 km di distanza.

Una misura più precisa di associazione spaziale tra due variabili territoriali è fornita dal coefficiente di codispersione (Vallejos, 2008) per un dato numero di classi di distanza. Il procedimento di calcolo prevede il calcolo del coefficiente di codispersione per due sequenze spaziali definite su griglie regolari (bin) dopo aver determinato un numero determinato di lag di stanza. Successivamente il coefficiente di codispersione viene calcolato per ogni

8 La distanza più breve tra due punti ("great-circle-distance" o "as the crow flies") è stata calcolata in base al metodo ellisoidale di Vincenty (1975). Tale metodo - anche se computazionalmente più intenso rispetto alla formula di Haversine (half-versed-sine) utilizza un ellissoide e permette di ottenere risultati sono molto precisi.

9 L'asse x nella figura è limitato a 50 km per esigenze grafiche.

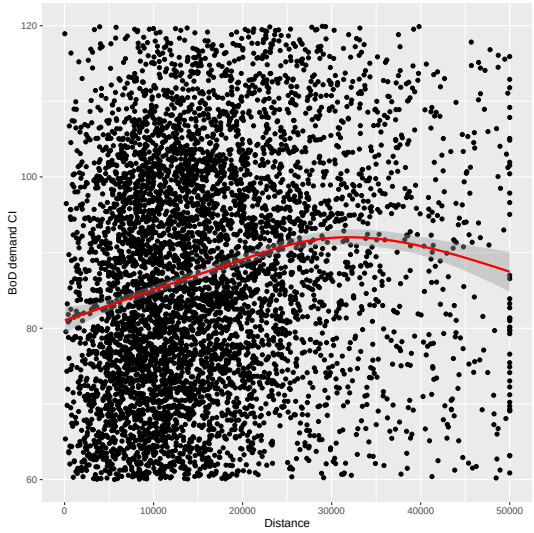


Figura 29: Domanda stimata e distanza minima dalla struttura sanitaria (in metri)

bin. La tabella 13 mostra che - a prescindere dalla classe di distanza - il coefficiente di codispersione è significativamente positivo e che nelle aree più remote (classe 13 - dove maggiore è la maggiore distanza tra Comune e struttura sanitaria) questo coefficiente rimane significativo.

7.4 OSSERVAZIONI CONCLUSIVE

Negli ultimi anni, in Italia, sono in corso importanti cambiamenti nella gestione dei servizi sanitari, determinati principalmente da esigenze di controllo del bilancio: le singole ASL sono cresciute in termini di dimensioni perdendo le caratteristiche di conoscenza e gestione del territorio, mentre le funzioni di programmazione dell'offerta tendono a tornare sotto il controllo dell'Amministrazione Centrale.
Questo cambiamento non è necessariamente negativo, anzi, ma necessita - a nostro parere - di riprendere in considerazione l'esigenza di conoscere la domanda di salute al livello più disaggregato possibile al fine di riallineare dinamicamente la relativa offerta, soprattutto in un contesto di "quasi-mercato" in cui l'offerta non tende a riallinearsi naturalmente nel breve termine.

Date queste premesse, per procedere con un'adeguata programmazione è necessario disporre di dati a livello micro; i dati provinciali, lì dove sono disponibili, non sono sufficienti

Classe di distanza	Limite superiore	Cardinalità	Coefficiente
1	0,4012	470651	0,2084
2	0,8024	1068371	0,2166
3	1,2036	1353979	0,2135
4	1,6048	1396710	0,1989
5	2,006	1319460	0,1911
6	2,4072	1236447	0,1751
7	2,8084	1154419	0,1506
8	3,2096	1050112	0,1331
9	3,6108	978088	0,1282
10	4,012	904622	0,1269
11	4,4132	827775	0,1182
12	4,8144	735889	0,1348
13	5,2156	4089397	0,1817

Tabella 13: Coefficiente di codispersione tra indicatore di rischio e distanza al variare della classe di distanza

in quanto la presenza dei grandi centri urbani spesso maschera le differenze intra-provinciali. La procedura Chow-Lin proposta ha permesso di stimare un indicatore della domanda a livello comunale disaggregato. Le variabili rilevanti, utilizzate per disaggregare spazialmente l'indice provinciale, sono state raccolte a livello comunale ed il loro contributo alla stima è stato valutato su un campo spaziale continuo. La stima disaggregata, infine, è costituita da una parte deterministica legata a fattori municipali e da una parte stocastica legata a fattori provinciali.

L'indicatore di domanda disaggregato a livello comunale fornisce due importanti risultati: *(i)* regolarità cross-regionali esistono e sono significative in Italia sia in termini di domanda che di offerta; tali regolarità dovrebbero essere prese in considerazione per evitare un'offerta diseguale collegata ai singoli confini amministrativi regionali; *(i)* il rischio sanitario aumenta con la distanza fisica dalle strutture sanitarie stesse: elevati costi privati per i residenti delle aree periferiche, montane e/o povere, infatti, possono scoraggiare l'accesso alle strutture sanitarie e diminuire il ricorso alla medicina preventiva e agli screening.

Da un punto di vista statistico, la ricerca ha mostrato che, in assenza di un dati ufficiali più affidabili, le tecniche di disaggregazione spaziali possono essere molto utili nell'analisi del mercato anche se gli errori di misura che presentano non possono essere eliminati.
Questo lavoro costituisce un primo tentativo di fornire prove empiriche, tratte da tecni-

che di disaggregazione spaziale, utili per migliorare la ripartizione dei fondi pubblici e la localizzazione delle unità sanitarie locali in una prospettiva sub-regionale.

Molto si può fare ancora: l'incredibile aumento della disponibilità di dati sul comportamento individuale e sociale può sicuramente aiutare a migliorare le decisioni e le politiche pubbliche. I dati provenienti da più fonti eterogenee (istruzione, reddito, patrimonio, stato epidemiologico personale) possono essere integrati e combinati per studiare l'impatto delle decisioni politiche ad una risoluzione il più possibile dettagliata al fine di ricostruire modelli omogenei di individui anche nel tempo (Bohning *et al.*, 2000).

8 CONCLUSIONI

Il modello predisposto per SaniRegio2016, nonostante presenti gli stessi obiettivi e utilizzi, in gran parte, la stessa base dati delle versioni precedenti, risulta notevolmente ampliato nel ventaglio dei risultati e delle tecniche econometriche impiegate. In SaniRegio2016, in particolare, accanto all'analisi della spesa, si considerano in modo esplicito e dettagliato i livelli delle prestazioni erogate (standard e storico), gli input impiegati e quindi il livello di spesa efficiente di ogni sistema regionale. Si è introdotto, inoltre, l'output-gap, ovvero la differenza tra l'output storico e l'output standard misurato con riferimento alla struttura della popolazione residente. Le regioni con output-gap positivo sono quelle che, diversamente dalle regioni con output-gap negativo, riescono a soddisfare meglio la domanda del proprio territorio. Accanto all'output-gap, poi, è stato possibile misurare il livello di spesa attribuibile all'offerta di servizi erogati sopra o sotto lo standard.

I risultati finali mostrano, tra il 1998 il 2002, un incremento dell'inefficienza tecnica che spiega, da sola, parte del forte incremento della spesa registrato in quegli anni. Dal 2002 in poi, invece, si registra una costante riduzione dell'inefficienza tecnica grazie a un trend costante nel livello dell'input lavoro per abitante, un trend decrescente nel livello delle dotazioni strumentali per abitante ed un incremento dell'output per abitante, soprattutto a partire dal 2008, anno a partire dal quale si consolidano sia i maggiori recuperi di efficienza sia le prime riduzioni nel livello della spesa corrente in termini reali. Conseguentemente, gli incrementi di spesa osservati tra il 2002 e il 2006 possono essere attribuiti in modo prevalente al livello dei salari e dei prezzi d'acquisto dei beni e servizi intermedi a causa dalla dinamica dei rinnovi contrattuali superiore all'inflazione programmata e dalla mancanza di controllo sugli ordinatori di spesa.

Dalla distribuzione territoriale dell'inefficienza tecnica emerge un acuirsi nel tempo del divario tra le regioni del Centro-Nord e quelle del Sud confermando i risultati prodotti nelle precedenti versioni di SaniRegio. Nel 2012 le regioni del Sud, incluse le Isole, presentano valori di inefficienza quasi il doppio più alti rispetto alle regioni del Centro e del Nord (incluse le regioni a statuto speciale). Le cause del mancato recupero di efficienza osservato nelle regioni meridionali vanno ricercate non solo nei più bassi livelli qualitativi e quantitativi dei servizi offerti, ma anche nella non corretta proporzione tra gli input. Da ultimo, tra le cause del ritardo in termini di efficienza delle regioni del Sud un ruolo non secondario va attribuito all'impatto generato dall'adozione dei piani di rientro che hanno interessato

principalmente le regioni meridionali (Lazio, Abruzzo, Molise, Campania, Calabria e Sicilia hanno avviato il piano di rientro nel periodo 2006-2012 e sono ancora in vigenza di piano). Da qui la necessità di riformare parzialmente questo strumento di governance sviluppando sistemi di monitoraggio che pongano maggiore enfasi sul livello dei servizi erogati a parità di input al fine di garantire che, a fronte di un miglioramento dei saldi finanziari, non si registri un deterioramento delle prestazioni.

Con riferimento all'anno 2012, l'ultimo anno per i quale si dispone di dati effettivi sulle misure di output, si registra un output-gap positivo o prossimo allo zero nel nord del Paese (la principale eccezione, in questo caso, è rappresentata dal Trentino Alto Adige); le regioni del Centro-Sud, al contrario, mostrano un output-gap negativo (le uniche eccezioni sono rappresentate dal Molise e dalla Basilicata).

Come ulteriore esercizio, inoltre, si è proceduto al calcolo del riparto della spesa sanitaria utilizzando come regioni benchmark la Lombardia, il Veneto e l'Umbira essendo quelle che, nel 2012, presentano inefficienza tecnica e output-gap prossimi allo zero. Dal punto di vista operativo, la spesa media standard per fascia di popolazione valorizzata per queste due regioni è stata utilizzata per calcolare la spesa standard delle altre regioni in base alla distribuzione regionale della popolazione nelle varie fasce di età. I risultati finali mostrano che, sia utilizzando le due regioni benchmark, sia adoperando direttamente i fitted values della funzione di spesa, si ottiene una variazione nella distribuzione delle risorse rispetto alla spesa storica non molto difforme. Con l'utilizzo della funzione di spesa, però, è possibile isolare la componente di spesa attribuibile all'inefficienza tecnica generata dalla presenza di input in eccesso oltre che dall'output-gap, modificando in questo modo il riparto delle risorse a favore delle regioni più virtuose. Se si considera l'impatto sulla spesa generato dall'inefficienza tecnica, infatti, tutte le regioni del Sud subiscono una significativa penalizzazione nel riparto delle risorse, nonostante caratterizzate da un output-gap prevalentemente negativo.

Dall'analisi della distribuzione territoriale dei differenziali di spesa, inefficienza ed output-gap registrati tra il 2014 e il 2012 emerge, inoltre, come nelle regioni meridionali (dove si registrano i picchi più alti di riduzione della spesa), diversamente dalle regioni del Nord, la contrazione della spesa corrente sia avvenuta principalmente a seguito della riduzione delle livello delle prestazioni piuttosto che per la riduzione dei livelli di inefficienza.

Tali conclusioni relative ai macro aggregati di spesa pubblica sono fondamentali per una corretta programmazione dei trasferimenti statali, ma rischiano di rimanere sterili se non calati in modo corretto ed equo sui singoli territori o a livello più micro.
Da questa considerazione è nato l'interesse, quindi, per stimare la domanda di salute ad un livello il più possibile disaggregato al fine da un lato di fornire strumenti statistici per riallineare dinamicamente la relativa offerta e dall'altro di sottolineare sempre più l'importanza di produrre/diffondere dati sul comportamento individuale e sociale che possano aiutare a migliorare le decisioni e le politiche pubbliche.

L'indicatore di domanda disaggregato a livello comunale ha fornito due importanti risultati sulle regolarità cross-regionali in termini di domanda e di offerta e sul rischio sanitario che aumenta all'aumentare della distanza fisica dalle strutture sanitarie; tali risultati, ancora

in forma provvisoria, dovrebbero essere prese in considerazione per evitare un'offerta diseguale collegata ai singoli confini amministrativi regionali e in ultima analisi per costruire sistemi che nell'interazione tra centro e periferia possano collaborare per una migliore e più puntuale offerta locale.

9 APPENDICE

VARIABILI ELEMENTARI UTILIZZATE NELL'ANALISI DI EFFICIENZA

Variabile	Etichetta	Ruolo nell'analisi
ANNESTESIA_EXTRA	Apparecchi per anestesia extraospedalieri	Input
CONTAGLOBULI	Contaglobuli automatici differenziali pubblico priv. accred.	Input
INFERMIERI_DIP_PRIV	Personale inferm. ist. cura pubblico priv. accred.	Input
INFERMIERI_DIP_PUB	Personale inferm. ist. cura pubblico	Input
IPERBARICHE	Camere iperbariche pubblici,privati accreditati	Input
MEDICI_DIP_PRIV	Medici odont. ist. cura pubblico priv. accred.	Input
MEDICI_DIP_PUB	Medici odont. ist. cura pubblico	Input
MEDICI_DIP_SSN	Medici odont. dip. SSN	Input
POST_LET_DAY_PRC	Posti letto ospedalieri day hospital per 10.000 abit.	Input
POSTI_LETTO_PRC3	Posti letto ospedalieri ordinari per 10.000 abit.	Input
RIABILITAZ_DIP_PRIV	Pers. funzioni riabilitazione ist. cura pubblico priv. accred.	Input
RIABILITAZ_DIP_PUB	Pers. funzioni riabilitazione ist. cura pubblico	Input
TAC	T.A.C. pubblici, privati accreditati	Input
TAVOLI_EXTRA	Tavoli operatori extraospedalieri	Input
TAVOLI_RADIO	Tavoli radiocomandati per apparecchi radiologici pubb. privato accreditato	Input
TECNICI_DIP_PRIV	Pers. tecnico-sanitario ist. cura pubblico priv. accred.	Input
TECNICI_DIP_PUB	Pers. tecnico-sanitario ist. cura pubblico	Input
ASSIST_PEDIATRA	Numero medio di assistiti 0-14 per pediatra di base	Output
ASSIST_PER_MEDICO	Numero medio di assistiti per medico generico	Output
DEGENZA_MEDIA_ACUTI_PRIV	Degenza media privato accreditato acuti	Output
DEGENZA_MEDIA_PRIV	Degenza media privato accreditato	Output
DEGENZA_MEDIA_RIABILITAZ	Degenza media lungodegenza e riabilitazione	Output
DEGENZA_MEDIA_RIABILITAZ_PRIV	Degenza media privato accreditato lungodegenza e riabilitazione	Output
DEGENZA_MEDIA_RIABILITAZ_PUB	Degenza media pubblici lungodegenza e riabilitazione	Output
DEGENZE_ACUTI	Degenze acuti	Output
DEGENZE_PRIV	Degenze privati accreditati	Output
DEGENZE_PRIV_ACUTE	Degenze privato accreditato acuti	Output
DEGENZE_PUB	Degenze pubblico	Output
DEGENZE_PUB_ACUTE	Degenze pubblico acuti	Output
DEGENZE_RIABILITZ	Degenze lungodegenze e riabilitazione	Output
DEGENZE_RIABILITZ_PRIV	Degenze privato accreditato lungodegenze e riabilitazione	Output
DEGENZE_RIABILITZ_PUB	Degenze pubblici lungodegenza e riabilitazione	Output
DEGENZE_TOT	Degenze totali	Output
GIORNATE_DEG	Giornate di degenza	Output
GIORNATE_DEG_ACUTI	Giornate di degenza acuti	Output
GIORNATE_DEG_ACUTI_PRIV	Giornate di degenza priv. accr. acuti	Output
GIORNATE_DEG_ACUTI_PUB	Giornate di degenza pubblici acuti	Output
GIORNATE_DEG_PRIV	Giornate di degenza privato accreditato	Output
GIORNATE_DEG_PUB	Giornate di degenza pubblico	Output
GIORNATE_DEG_RIABILITAZ	Giornate di degenza lungodegenza e riabilitazione	Output
GIORNATE_DEG_RIABILITAZ_PRIV	Giornate di degenza priv. accr. lungodegenza e riabilitazione	Output
GIORNATE_DEG_RIABILITAZ_PUB	Giornate di degenza pubblici lungodegenza e riabilitazione	Output
MOBILITA_PRC	Saldi mobilità ospedaliera interregionale (attivi - passivi) - per 10.000 abit.	Output
MOBILITA2_PRC	Saldi mobilità ospedaliera interregionale (attivi - passivi) / ricoveri residenti	Output
TASSO_POSTI_LETTO_OSP	Tasso utilizzo posti letto ospedalieri	Output

Tabella 14: Variabili elementari utilizzate nell'analisi di efficienza

PUNTEGGI FATTORIALI E INDICATORE BOD PER PROVINCIA

Si riportano di seguito i punteggi fattoriali ed il relativo indicatore composito RBoD per Provincia relativo all'indicatore di rischio sanitario proposto nel paragrafo 7.3.2.

Codice	Provincia	Fattore #1	Fattore #2	Fattore #3	Fattore #4	RBoD CI
1	Torino	3.23	1.70	2.10	1.66	0.75
2	Vercelli	4.14	2.97	3.16	1.26	1.00
3	Novara	2.73	1.62	2.11	1.06	0.66
4	Cuneo	2.97	2.30	3.06	1.27	0.78
5	Asti	3.71	3.14	3.70	1.43	0.97
6	Alessandria	3.77	3.61	2.46	0.93	0.96
7	Aosta	2.98	0.47	2.96	2.69	0.82
8	Imperia	3.20	3.37	3.71	1.41	0.96
9	Savona	3.58	3.41	4.41	2.07	1.00
10	Genova	4.35	3.89	0.80	1.31	1.00
11	La Spezia	3.86	3.58	1.65	2.25	0.93
12	Varese	3.00	1.07	1.23	1.12	0.64
13	Como	2.73	0.92	1.42	1.30	0.61
14	Sondrio	2.74	0.53	4.10	1.05	0.87
15	Milano	3.00	0.90	1.22	0.96	0.64
16	Bergamo	2.27	0.41	2.26	0.75	0.59
17	Brescia	2.43	0.67	1.73	0.59	0.58
18	Pavia	3.41	2.23	2.18	2.17	0.79
19	Cremona	3.07	1.42	2.25	1.60	0.74
20	Mantova	2.10	1.98	2.78	1.07	0.62
21	Bolzano/Bozen	1.57	0.15	2.82	1.47	0.61
22	Trento	2.32	0.74	1.82	1.46	0.57
23	Verona	2.20	1.13	2.41	0.88	0.59
24	Vicenza	2.22	0.45	2.11	1.36	0.57
25	Belluno	3.68	1.50	2.16	3.78	0.91
26	Treviso	2.01	0.26	2.43	1.06	0.56
27	Venezia	2.74	1.18	2.35	2.13	0.71
28	Padova	2.31	1.09	2.27	0.74	0.60
29	Rovigo	2.59	1.98	3.13	3.34	0.90
30	Udine	3.23	1.00	3.12	2.98	0.89
31	Gorizia	3.67	1.11	2.38	5.07	1.00
32	Trieste	4.82	3.85	1.57	3.36	1.00
33	Piacenza	3.41	2.09	2.99	1.31	0.86
34	Parma	3.12	1.81	2.14	1.23	0.74
35	Reggio nell'Emilia	2.67	1.52	2.37	0.41	0.67
36	Modena	2.68	1.21	2.53	0.90	0.69
37	Bologna	3.40	2.13	1.85	1.42	0.76
38	Ferrara	3.05	2.84	4.55	2.20	1.00
39	Ravenna	2.93	2.24	3.95	0.82	0.87
40	Forli-Cesena	2.91	1.80	1.52	1.00	0.65
41	Pesaro e Urbino	2.43	2.30	2.54	1.26	0.66
42	Ancona	2.78	2.22	3.77	1.15	0.83
43	Macerata	2.75	2.81	3.13	1.06	0.80
44	Ascoli Pice	2.56	2.59	2.17	0.67	0.71
45	Massa-Carrara	3.33	3.74	2.95	1.35	1.00
46	Lucca	2.67	3.23	2.47	1.50	0.87
47	Pistoia	2.30	2.38	2.81	1.53	0.69
48	Firenze	2.77	2.22	2.49	0.65	0.70
49	Livorno	2.24	2.70	3.47	2.84	0.94
50	Pisa	2.66	2.70	2.53	0.84	0.74
51	Arezzo	2.44	3.34	3.19	0.36	0.92
52	Siena	2.22	2.60	4.97	1.73	1.00
53	Grosseto	2.42	3.33	4.15	1.81	0.97
54	Perugia	2.35	2.33	2.97	0.88	0.68
55	Terni	3.21	2.83	2.51	2.51	0.84

Tabella 15: Punteggi fattoriali e indicatore composito RBoD per Provincia (1/2)

Codice	Provincia	Fattore #1	Fattore #2	Fattore #3	Fattore #4	RBoD CI
56	Viterbo	2.09	2.62	3.64	2.01	0.85
57	Rieti	1.87	3.13	2.80	2.83	0.94
58	Roma	2.15	1.67	2.01	1.09	0.55
59	Latina	1.18	1.32	3.21	0.97	0.62
60	Frosinone	1.26	2.90	2.50	2.45	0.84
61	Caserta	0.88	1.81	0.40	1.91	0.52
62	Benevento	1.10	3.79	0.98	2.32	0.98
63	Napoli	0.87	1.82	0.47	2.01	0.53
64	Avellino	1.09	3.45	1.16	1.61	0.89
65	Salerno	0.85	2.63	1.92	1.50	0.71
66	L'Aquila	2.02	3.03	2.78	2.65	0.90
67	Teramo	1.77	1.94	2.53	2.55	0.74
68	Pescara	1.99	1.81	3.48	2.01	0.79
69	Chieti	1.65	2.93	3.71	1.94	0.87
70	Campobasso	1.40	2.81	2.65	2.88	0.89
71	Foggia	0.93	1.55	3.22	2.21	0.76
72	Bari	0.91	1.32	3.02	1.55	0.65
73	Taranto	0.99	1.81	2.27	1.36	0.56
74	Brindisi	1.25	2.77	2.70	0.53	0.76
75	Lecce	1.87	2.38	0.92	1.23	0.62
76	Potenza	1.17	2.98	2.38	2.96	0.90
77	Matera	0.83	2.51	2.83	1.41	0.72
78	Cosenza	0.60	2.02	3.05	3.30	0.89
79	Catanzaro	0.68	2.52	1.64	2.35	0.72
80	Reggio di Calabria	0.85	2.50	1.74	1.90	0.68
81	Trapani	0.97	3.61	3.73	0.74	1.00
82	Palermo	1.54	2.30	1.25	1.17	0.60
83	Messina	1.10	3.60	1.43	1.12	0.93
84	Agrigento	0.66	3.23	1.91	1.50	0.85
85	Caltanissetta	1.07	3.01	1.70	1.58	0.79
86	Enna	0.73	3.25	2.70	2.65	0.92
87	Catania	0.71	2.49	2.56	0.98	0.69
88	Ragusa	0.89	1.89	2.57	2.04	0.68
89	Siracusa	0.49	2.75	2.77	1.35	0.77
90	Sassari	1.91	0.99	3.30	1.91	0.73
91	Nuoro	1.72	0.53	4.31	3.16	0.98
92	Cagliari	2.34	0.37	3.43	1.04	0.74
93	Pordenone	3.12	0.94	1.49	1.46	0.68
94	Isernia	1.39	2.25	1.50	5.03	1.00
95	Oristano	2.26	1.81	2.49	2.75	0.75
96	Biella	3.43	2.27	3.40	3.45	0.98
97	Lecco	2.42	0.64	1.71	1.42	0.57
98	Lodi	3.17	0.89	2.40	0.30	0.76
99	Rimini	2.64	1.09	2.15	0.87	0.65
100	Prato	2.23	1.77	1.81	0.35	0.55
101	Crotone	0.40	1.43	4.03	2.23	0.87
102	Vibo Valentia	1.08	1.99	1.98	1.89	0.62
103	Verbano-Cusio Ossola	3.65	1.47	1.32	3.09	0.81
104	Olbia-Tempio	0.96	0.18	6.10	1.29	1.00
105	Ogliastra	2.00	0.61	4.17	2.70	0.93
106	Medio Campidano	2.65	0.37	2.07	6.14	1.00
107	Carbonia-Iglesias	3.53	0.95	2.77	1.52	0.86
108	Monza e della Brianza	2.55	0.24	1.21	1.49	0.56
109	Fermo	1.82	2.10	4.83	1.11	0.94
110	Barletta-Andia-Trani	0.91	1.05	1.60	2.47	0.56

Tabella 16: Punteggi fattoriali e indicatore composito RBoD per Provincia (2/2)

EFFICIENZA TECNICA PRODUTTIVA

La determinazione di una misura di efficienza tecnica si basa sulla conoscenza della cosiddetta "funzione di produzione" che rende possibile il confronto, all'interno di un ben definito processo di produzione, tra le singole unità con la frontiera efficiente di produzione. La misura della distanza di ciascuna unità dalla frontiera rappresenta il modo più immediato per valutare l'efficienza (Farrell, 1957).

La misura dell'efficienza consiste nell'attribuire ad ogni unità decisionale Decision Making Unit (DMU) un punteggio compreso tra 0 (completa inefficienza) ed 1 (massima efficienza). Tale misura può essere riferita agli input oppure agli output ed individua sempre un benchmark (cioè un punto di riferimento che l'unità produttiva potrebbe raggiungere migliorando la propria efficienza, ovvero contraendo gli input o massimizzando gli output).

La vera funzione di produzione, tuttavia, così come la frontiera efficiente non sono generalmente note, ma lo sono solamente un insieme di informazioni su ciascuna unità di produzione, che sono quindi essenziali per stimare la frontiera stessa.

Nella letteratura econometrica, la specificazione e la successiva stima della funzione di frontiera sono state sviluppate essenzialmente secondo due diversi filoni principali, almeno fino alla fine degli anni '90:

- Le analisi di tipo parametrico quali la Stochastic Frontier Analysis (Stochastic Frontier Analysis (SFA))(Meeusen & van den Broeck, 1977) o la Deterministic Frontier Analysis (Deterministic Frontier Analysis (DFA)) (Aigner & Chu, 1968);

- Le analisi di tipo non parametrico quali la Data Envelopment Analysis (DEA) (Farrell, 1957; Charnes *et al.*, 1978) o la Free Disposal Hull (Free Disposal Hull (FDH)) (Deprins *et al.*, 1984; Grosskopf, 1996).

In un'analisi parametrica è indispensabile precisare *a priori* una forma funzionale esplicita della frontiera di produzione, mentre le analisi di tipo non parametrico sono caratterizzate dalla capacità di determinare l'efficienza relativa delle unità attraverso la programmazione lineare, senza specificare alcuna forma funzionale.

In altre parole, a differenza delle tecniche parametriche, l'approccio DEA-FDH consente la determinazione dell'efficienza relativa delle unità decisionali in assenza di una descrizione dettagliata del processo di produzione. Se quest'ultimo, a prima vista, sembra rendere questo approccio particolarmente flessibile e generalizzabile, il principale svantaggio delle tecniche DEA (o FDH) risiede proprio nella sua natura deterministica.

Quando si utilizzano tali procedure non è possibile riconoscere se la differenza in termini di efficienza, vale a dire la distanza tra l'output osservato e il massimo possibile, sia causato dalla sola inefficienza tecnica o da effetti di disturbo di tipo accidentale (Greene, 2008).

Pertanto, non è possibile determinare se, ad esempio, l'inefficienza è causata da una condizione sfavorevole delle variabili di contesto esterne ed è quindi indipendente delle azioni dell'imprenditore o dell'amministratore pubblico (si pensi ad esempio ai vincoli sull'offerta

come ad esempio i vincoli geomorfologici di una Regione) oppure può essere spiegata da altri fattori (come ad esempio la differente qualità della gestione del personale all'interno di un ospedale) o ancor di più da fattori casuali o da errori di misura degli input/output.

Il modello parametrico di frontiera stocastica, quindi, oltre a fornire informazioni utili sull'attività produttiva, supera i limiti connessi con i modelli deterministici, contribuendo a fornire una dettagliata analisi anche delle fonti di inefficienza non direttamente attribuibili alla politica aziendale e/o ai disturbi di tipo casuale.

Per contro lo svantaggio più rilevante associato con l'approccio SFA è la mancanza di flessibilità associato con la specificazione di una data forma funzionale. Per superare questo problema, nell'ultimo decennio sono state considerate forme più flessibili della funzione di produzione rilassando alcune proprietà della stessa (si pensi ad esempio all'introduzione dei modelli GAM nell'SFA (Vidoli *et al.*, 2012), cercando di eliminare le differenze tra i due filoni principali di stima con la proposta di metodi semi-parametrici o semi-nonparametrici (si veda, in particolare, il contributo di Kuosmanen & Kortelainen (2012) che appare come il più attraente tra le nuove procedure proposte - metodi Stochastic semi-Nonparametric Envelopment of Data (StoNED) e Stochastic semi-Nonparametric Envelopment of Z variables Data (StoNEZD)) o controllando gli effetti casuali sulle stime (Order-m, Daraio & Simar, 2007).

Oltre ai classici metodi di stima, dunque, negli ultimi dieci anni sono stati sviluppati una moltitudine di metodi che possono essere definiti come "ponte" tra i due filoni principali: la tabella 17 suggerisce una tassonomia.

Tecnologia di produzione	Tecniche di stima
Non parametrica	DEA, Farrell (1957); Charnes *et al.* (1978) FDH, Deprins *et al.* (1984); Grosskopf (1996) Order-m, Daraio & Simar (2007)
Semi - nonparametrica	StoNED, Kuosmanen & Kortelainen (2012) StoNEZD, Johnson & Kuosmanen (2011)
Semiparametrica	Park & Simar (1994), Park *et al.* (2007)
Parametrica	SFA, Aigner *et al.* (1977); Meeusen & van den Broeck (1977) DFA, Aigner & Chu (1968)

Tabella 17: Tecniche di stima dell'efficienza per tecnologia di produzione

Se quindi i modelli di tipo parametrico permettono una buona descrizione sia delle variabili di contesto che influenzano il processo produttivo sia dell'effetto del tempo sulla frontiera, rimane il problema di poter includere nell'analisi solo y univariate in un contesto applicativo, come quello sanitario, dove invece sono molteplici gli output da considerare.

Tuttavia, la funzione di produzione, come la frontiera efficiente, non è generalmente nota, ma lo sono solamente le informazioni, in termini di input ed output, per ciascuna unità

di produzione. Si è preferito quindi utilizzare tecniche non parametriche che permettano la stima d'efficienza a partire dai dati stessi, ovvero che permettano di ricavare misure di efficienza in modo non parametrico.

Più formalmente, si consideri una tecnologia nella quale l'attività di una generica DMU è caratterizzata da un insieme di input $x \in \mathbb{R}_+^p$ utilizzati per produrre un insieme di output $y \in \mathbb{R}_+^q$. L'insieme di produzione Ψ è il set di tutte le combinazioni tecnicamente fattibili di (x, y):

$$\Psi = \{(x, y) \in \mathbb{R}^{p+q} | \ x \ \mathrm{puo'} \ \mathrm{produrre} \ y\} \tag{30}$$

ed il DGP (Data Generating Process) può essere scritto come:

$$H(x, y) = \mathrm{Prob}(X \leqslant x, Y \geqslant y) \tag{31}$$

Poste queste premesse, Ψ rappresenta il "supporto" di $H(x, y)$.

I punteggi di efficienza di tipo Farrell-Debreu per un determinato set produttivo $(x, y) \in \Psi$ possono essere definiti sia in termini di output massimo producibile a parità di input (*output oriented*) sia in termini di input minimi potenzialmente impiegabili a parità di output prodotto (*input oriented*). Naturalmente la scelta dell'orientamento deve essere fatta ex ante dal ricercatore in base alla natura del problema analizzato e in accordo ai vincoli sottostanti al processo decisionale (si pensi, ad esempio, ai servizi pubblici nei quali è più semplice per l'Amministrazione produrre più output piuttosto che contrarre l'input impiegato).

Da un punto di vista algebrico il problema di ottimizzazione può essere espresso come:

$$\text{Input oriented: } \theta(x, y) = \inf\{\theta | (\theta x, y) \in \Psi\} \tag{32}$$

$$\text{Output oriented: } \lambda(x, y) = \sup\{\lambda | (x, \lambda y) \in \Psi\} \tag{33}$$

Dato che Ψ non è conosciuto e deve essere stimato a partire da un campione casuale di unità produttive $\chi = \{(X_i, Y_i) | i = 1, \ldots, n\}$ in un framework di frontiera deterministica, assumiamo che $\mathrm{Prob}((X_i, Y_i) \in \Psi)$.

In altre parole, in accordo con le precedenti definizioni, il supporto della variabile casuale (X, Y) dove Ψ si suppone sia compatta, deve essere stimato.

Gli stimatori non parametrici più popolari e maggiormente utilizzati in letteratura sono l'FDH e la DEA, entrambi basati sulle idee di inviluppo dei dati proposte da Farrel e Debreu. Le differenze tra i due stimatori risiedono nel fatto che mentre il metodo FDH non richiede l'ipotesi di convessità per la funzione di produzione per cui i benchmark stimati risiedono esattamente sulla frontiera di Ψ, la DEA, al contrario, ipotizza la possibilità che esistano unità ottimali combinazione lineare di unità efficienti. La figura 30 mostra tali differenze di base.

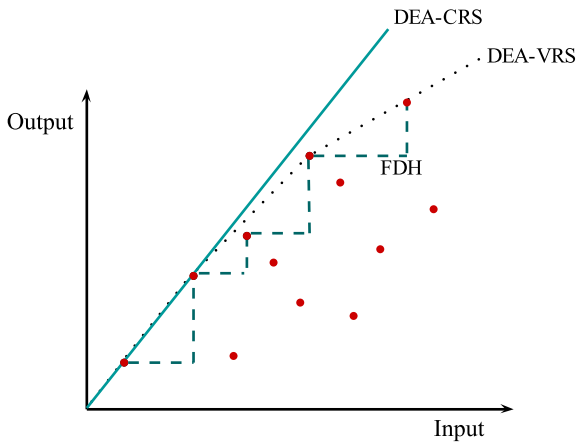


Figura 30: FDH, CRS-DEA e VRS-DEA

Lo stimatore FDH può essere quindi calcolato tramite il free disposal hull - si veda la figura 31 - dei punti in esame X (qui nella versione input oriented):

$$\widehat{\Psi}_{FDH} = \{(x, y) \in \mathbb{R}^{p+q} | y < Y_i, x \geqslant X_i, \quad i = 1, \ldots, n\} \tag{34}$$

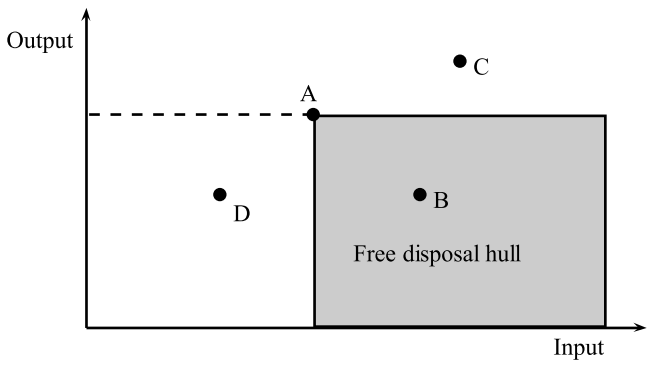


Figura 31: Free disposal hull per la generica unità A

I punteggi di efficienza FDH si ottengono quindi introducendo $\widehat{\Psi}_{FDH}$ stimato nell'equazione 32 e 33 al posto dello Ψ non conosciuto.

Introducendo l'ipotesi di convessità di Ψ è possibile introdurre il concetto di "convex hull" di $\widehat{\Psi}_{FDH}$ al fine di ricavare:

$$\widehat{\Psi}_{DEA} = \{(x, y) \in \mathbb{R}^{p+q} | y < \sum_{i=1}^{n} \gamma_i yi \text{ for } (\gamma_1, \ldots, \gamma_n)$$

sotto i vincoli che:

$$\sum_{i=1}^{n} \gamma_i = 1; \gamma_i \geqslant 0, i = 1, \ldots, n \tag{35}$$

L'equazione (35) evidenzia uno dei maggiori vantaggi delle tecniche DEA-FDH rispetto alle tecniche parametriche: il fatto che tali modelli permettano di valutare l'efficienza in scenari multi-input e multi-output senza alcuna assunzione circa la forma funzionale e che consentano di confrontare la singola DMU con una combinazione di pari (peers).

Oltre a tali vantaggi, alcune caratteristiche sono state recentemente profondamente riviste da diversi autori (si veda, ad esempio Daraio & Simar, 2007) con lo scopo di superare i limiti della formulazione classica e di generalizzare l'applicabilità.
In particolare, negli ultimi anni, si è cercato di ovviare ad alcuni svantaggi delle tecniche classiche, inconvenienti che sono ancora più evidenti nelle applicazioni al settore pubblico[1].

La prima caratteristica negativa dei metodi classici è certamente l'estrema sensitività alla presenza di dati fuori scala o outliers. Per superare questo limite Cazals *et al.* (2002) ha proposto uno stimatore non parametrico della frontiera più robusto ai casi anomali o estremi; tale stimatore, basato sul concetto di "*expected minimum input function*" di ordine m (Order-m), è stato successivamente esteso da Daraio & Simar (2005) al caso multivariato.

Tale metodo mitiga l'effetto degli outliers introducendo una distribuzione di probabilità per le unità nel set Ψ. Per una introduzione più completa di questo metodo, si veda il paragrafo successivo nel quale tale idea è stata estesa all'ambito degli indicatori compositi.

La seconda caratteristica negativa di metodi classici è la non possibilità di eseguire test statistici per la stima; tale caratteristica è stata parzialmente attenuata da Simar e Wilson negli ultimi dieci anni (si veda ad esempio Simar & Wilson, 2007) tramite l'utilizzo di metodi bootstrap, al fine di ottenere stime non distorte. Il presupposto fondamentale, alla base di tale approccio, è che la distribuzione bootstrap approssimi correttamente la distribuzione non nota se il processo di generazione dei dati (DGP) risulti essere uno stimatore consistente del DGP non conosciuto. Anche per il metodo di tipo Order-m, Badin *et al.* (2010) hanno esteso le tecniche di bootstrap al fine di ottenere stime non distorte.

1 Per brevità espositiva, focalizzeremo l'attenzione solamente sue due aspetti: la robustezza e la non possibilità di eseguire test statistici per la stima.

INDICATORI COMPOSITI

Gli indicatori compositi *DEA-based* sono stati utilizzati per valutare il mercato europeo del lavoro (Storrie & Bjurek, 2000), i sistemi sanitari nazionali (Smith, 2002b; Kwon, 2003), le politiche europee di inclusione sociale (Cherchye *et al.*, 2004), la politica del mercato interno (Cherchye *et al.*, 2005) e lo Human Development Index (Mahlberg & Obersteiner, 2001; Despotis, 2005).

In questo approccio, l'indicatore composito viene espresso come somma pesata di indicatori semplici relativi ad un benchmark, ovvero definita come rapporto tra le prestazioni effettive di un'unità rispetto al suo benchmark:

$$CI_c = \frac{\sum_{q=1}^{Q} I_{cq} w_{cq}}{I^*_{cq}} \tag{36}$$

Quale benchmark bisogna quindi scegliere per ogni singola unità? Il metodo BoD, similmente alla DEA, cerca l'unità (o le unità) che presenti il massimo indicatore composito (dati i vincoli sui pesi):

$$I^*_{cq} = \max_{I_c \in 1..C} \sum_{q=1}^{Q} I_{cq} w_{cq} \tag{37}$$

Una volta quindi trovato il benchmark di riferimento per ogni unità, quali pesi bisogna assegnare, quindi, alla singola unità? la risposta sta nello scegliere i pesi tali che massimizzino il rapporto tra l'unità e il suo benchmark.

$$CI_c = \max_{w_{cq}} \frac{\sum_{q=1}^{Q} I_{cq} w_{cq}}{\max_{I_c \in 1..C} \sum_{q=1}^{Q} I_{cq} w_{cq}} \tag{38}$$

Si notino due proprietà desiderate dell'indicatore: la prima è che ogni unità si confronta con il suo benchmark, ovvero che non esiste un unico o più benchmark uguali per tutte le unità; la seconda è che l'insieme ottimale di pesi (se esiste) garantisce all'unità corrispondente la migliore posizione rispetto a tutte le altre unità del campione. Lo schema dei pesi scelto è quindi tale che l'indicatore risultante per ogni unità sia il più alto possibile; questa proprietà è particolarmente "*useful in policy arena, since policy-makers could not complain about unfair weighting: any other weighting scheme would have generated lower composite scores*" (Nardo *et al.*, 2005).

L'equazione (38) può essere scritta in maniera più compatta - e più simile alla notazione dei modelli non parametrici di frontiera - tenendo in considerazione che "*the classical BoD model can be regarded as an output oriented Data Envelopment Analysis with Constant Returns to Scale (CRS-DEA) (Charnes* et al.*, 1978) model, with all questionnaire items considered as output* ($Y_q \in \mathbb{R}_+, \forall q = 1...Q$) *and a single input equal to one for all observations*", (Witte & Rogge, 2009).

Pertanto l'insieme di produzione di un modello CRS-DEA, nell'ambito framework BoD visto come caso particolare, può essere scritto come:

$$H(\mathbb{1}, \mathbf{y}) = \mathrm{Prob}(X \equiv \mathbb{1}, \mathbf{Y} \geqslant \mathbf{y}) \tag{39}$$

dove Ψ rappresenta il supporto di $H(\mathbb{1}, \mathbf{y})$:

$$\Psi = \left\{ (\mathbb{1}, \mathbf{y}) \in \mathbb{R}_{+}^{1+Q} | H(\mathbb{1}, \mathbf{y}) > 0 \right\}. \tag{40}$$

Avendo definito il supporto Ψ e il random set $H(\mathbb{1}, \mathbf{y})$ è possibile calcolare la Farrel-Debreu (output) efficienza come :

$$\lambda(\mathbb{1}, \mathbf{y}) = \sup\{\lambda > 0 | H(\mathbb{1}, \lambda \mathbf{y}) > 0\} \tag{41}$$

Il CI calcolato tramite l'equazione (41) soddisfa le seguenti proprietà desiderate: i pesi sono determinati **endogeneamente** attraverso le performance osservate e il benchmark non è basato su vincoli o scelte teoriche, ma su combinazioni lineari di performance osservate; il CI di tipo BoD è **monotono debole**, ovvero rispetta tale proprietà: *Sia* CI *una funzione dei aggregazione di* Q *indicatori semplici* I_q, CI *è monotona positiva debole se per ogni* $c > 0$, $CI(I_1, ..., I_q, ..., I_Q) \leqslant CI(I_1, ..., I_q + c, ..., I_Q)$ ed infine, come affermato in precedenza, lo schema dei pesi risulta per la singola unità il **più alto possibile**.

In figura 32 viene rappresentata la distribuzione del CI di tipo BoD; si può facilmente notare che il punteggio composito dipende esclusivamente dalla distanza della frontiera e non dal rapporto *tra* i semplici indicatori; in tale modello si ha quindi perfetta compensabilità tra gli indicatori/output semplici[2].

Gli svantaggi principali del metodo BoD direttamente collegati alla formulazione DEA sono, dunque, la mancanza di robustezza rispetto a dati anomali, la possibile molteplicità delle soluzioni e la perfetta compensabilità tra gli indicatori/output.

Per quanto riguarda la compensabilità, in particolare, Munda & Nardo (2005) affermano che *"if one wants the weights to be interpreted as "importance coefficients" (or equivalently symmetrical importance of variables) non-compensatory aggregation procedures must be used"*.[3]

Per ovviare ai problemi di perfetta compensabilità, Vidoli & Mazziotta (2013) hanno proposto di integrare l'idea proposta da De Muro *et al.* (2010) denominata Metodo di Penalità per Coefficiente di Variazione (MPCV) in un modello BoD standard assumendo che ogni indicatore non possa essere sostituito da un altro o possa esserlo solamente in parte. Secondo questa idea, il metodo comporta l'introduzione di una penalità per tutte le unità non bilanciate su tutti gli indicatori, ovvero:

$$\mathrm{BoDPCV}_i = \mathrm{BoD}_i(1 - cv_i^2), \forall i = 1, \ldots, N \tag{42}$$

2 Nel rapporto SaniRegio2015 si è scelto di utilizzare metodi compensativi non avendo a disposizione informazioni aggiuntive sui livelli relativi desiderati di output; tale aspetto potrà comunque essere preso in considerazione in un ulteriore approfondimento del modello.

3 Per una più completa survey sugli approcci compensativi e non, si veda Vansnick (1990).

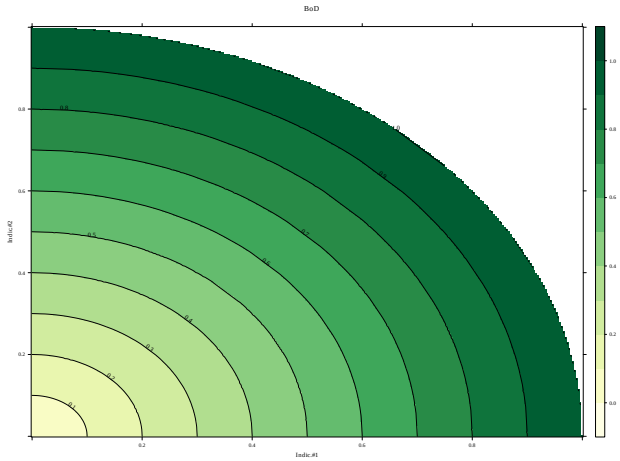


Figura 32: Distribuzione BoD

dove cv_i^2 rappresenta il coefficiente di variazione per l'unità i calcolato orizzontalmente su tutti gli indicatori semplici.

Questo approccio consente la penalizzazione delle unità che, pur avendo un punteggio BoD uguale, presentino un maggiore squilibrio tra gli indicatori stessi (si veda la figura 33).

Rispetto ai singoli approcci BoD ed MPCV, l'approccio Benefit of Doubt Penalized by Coefficient of Variation (BoD-PCV) presenta un duplice vantaggio: (i) per calcolare il CI vengono prese a riferimento sempre le unità sulla frontiera e (ii), allo stesso tempo, viene penalizzata, in caso di una scelta iniziale di tipo non compensativo, la presenza di unità sbilanciate.

L'aver introdotto un criterio di penalizzazione che ci permettesse di tenere in conto della compensabilità tra gli indicatori semplici, non mette al riparo il metodo da evidenti problemi di non robustezza rispetto ai dati anomali. La figura 34, nella cui distribuzione dei dati è stato aggiunta un'unità anomala, mostra come il metodo BoD-PCV sia, infatti, fortemente influenzato da valori fuori scala.

Per ovviare al problema della robustezza nei metodi BoD (e BoD-PCV), Vidoli & Mazziotta (2013) hanno proposto di integrare le idee proposte da Daraio & Simar (2005) riguardanti i metodi robusti non parametrici di tipo Order-m.

Per introdurre più semplicemente i modelli di tipo Order-m verrà esposta primariamente l'idea semplificata alla base del metodo per poi introdurre più formalmente il modello stesso.

Siano ad esempio due indicatori semplici (I_1, I_2) ed un punto C da valutare; la figura 35 evidenzia il problema della traslazione della frontiera, passante dai punti A e B alla presenza di un dato anomalo D.

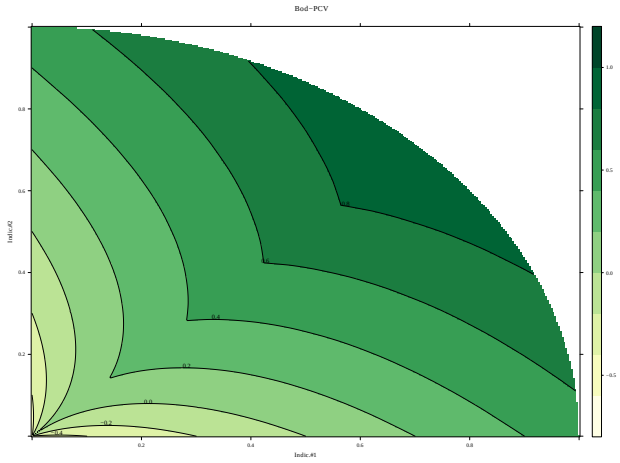


Figura 33: Distribuzione BoD-PCV

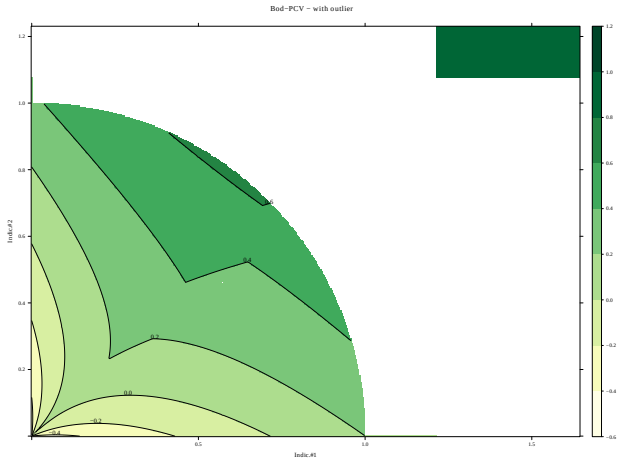


Figura 34: Distribuzione BoD-PCV in presenza di dati anomali

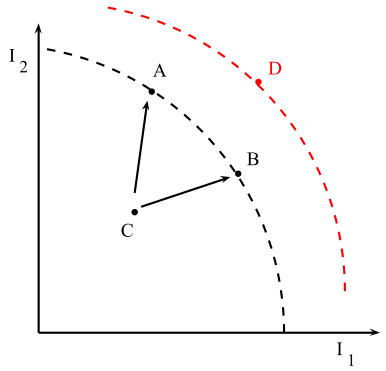


Figura 35: Presenza di outliers in un problema di stima di frontiera

L'idea formalizzata da Daraio & Simar (2005), consiste nel campionare con ripetizione in modo iterativo, per ogni unità - ad esempio C - m osservazioni a partire dall'insieme originario di n osservazioni, scegliendo solamente quelle che presentassero un livello più alto degli indicatori semplici (I_1, I_2) (ovvero a destra della linea rossa di figura 36).

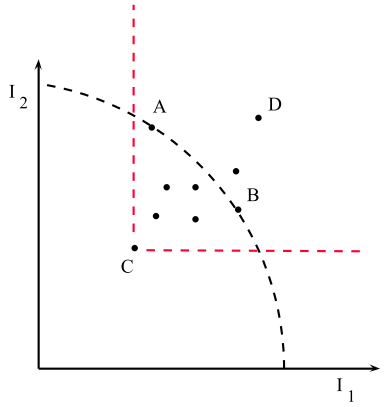


Figura 36: Supporto della generica unità C

In altre parola, con un certo grado di semplificazione, il metodo prevede di:

- campionare con ripetizione m osservazioni che presentino indicatori semplici superiori a quello di osservazione C;
- denominare questo sottoinsieme $\widetilde{\Psi}_{jm}$;

- stimare l'indicatore composito relativamente a questo sottoinsieme $\widetilde{\Psi}_{jm}$ per J volte;
- dopo aver ottenuto J punteggi, calcolare il valore atteso della distribuzione in esame.

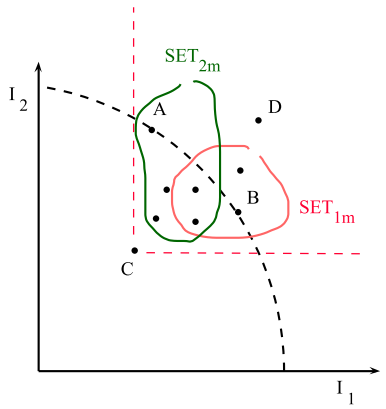


Figura 37: Criteri di selezione del sottoinsieme in un modello di tipo Order-m

Il punteggio che otterremo sarà quindi un "*less extreme*" benchmark per l'unità C rispetto al massimo assoluto livello di output effettivamente raggiungibile (punto D); detto in altri termini l'unità C sarà paragonata a una serie di m peers (potenziali competitors o unità di confronto) aventi indicatori semplici maggiori del suo livello prendendo come benchmark non il massimo CI assoluto, ma un valore atteso del massimo.
Più formalmente l'idea di Daraio & Simar (2005) può essere traslata nel campo degli indicatori compositi, riscrivendo il set di produzione Ψ (si veda l'equazione 39) in termini probabilistici come $\widetilde{\Psi}_m$:

$$\widetilde{\Psi}_m = \bigcup_{j=1}^{m} \{(\mathbb{1}, \mathbf{y}) \in \mathbb{R}_+^{1+Q} | X \equiv \mathbb{1}, \mathbf{Y_j} \geqslant \mathbf{y}\}. \tag{43}$$

dove m è il numero di variabili casuali con ripetizione appartenenti al sottoinsieme $S_m = \{\mathbf{Y_i}\}_{i=1}^{m}$.
Il punteggio dell'indicatore composito di tipo Order-m (chiamato RBoD) può essere quindi definito come:

$$\theta(\mathbb{1}, y_0) = \sup\{\theta | (\theta, y_0) \in \widetilde{\Psi}_m\} = \sup\{\theta | H(\theta, y) > 0\} \tag{44}$$

Le figure seguenti mostrano come, sia nel caso compensativo (figura 38) sia nel caso non compensativo (denominato Robust Benefit of Doubt Penalized by Coefficient of Variation (RBoD-PCV), figura 39) l'approccio di tipo Order-m contribuisca a rendere il framework

di stima dell'indicatore composito robusto e come gli ordinamenti relativi delle unità non vengano affatto distorti dalla presenza di dati anomali.

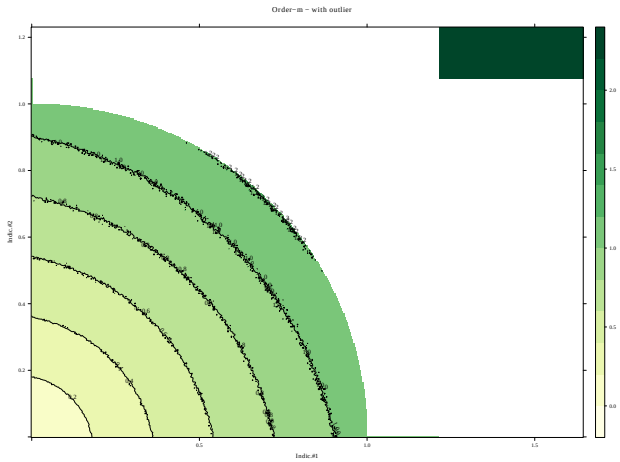


Figura 38: Distribuzione RBoD in presenza di outlier

In ultima analisi l'indicatore composito così ottenuto soddisfa le seguenti proprietà:

- i pesi sono determinati in modo non parametrico senza l'imposizione di vincoli o di scelte teoriche preliminari;
- il CI di tipo BoD è monotono crescente debole;
- lo schema dei pesi risulta per la singola unità il più alto possibile;
- le stime ottenute non sono influenzate da valori anomali.

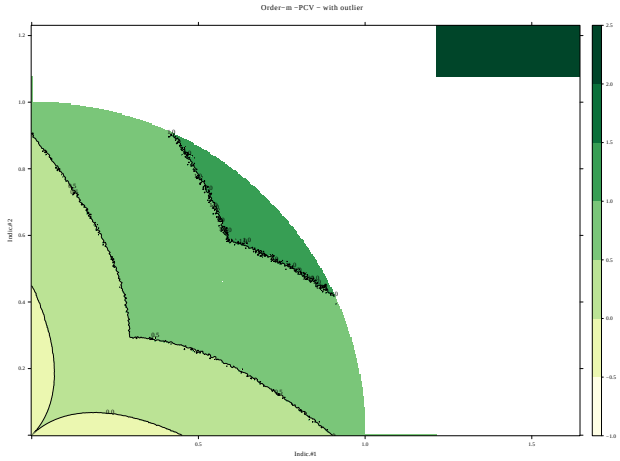


Figura 39: Distribuzione RBoD-PCV in presenza di outlier

BIBLIOGRAFIA

Aigner, D., Lovell, C. A. K., & Schmidt, P. 1977. Formulation and estimation of stochastic frontier production function models. *Journal of econometrics*, **6**, 21–37. (Not cited.)

Aigner, D. J., & Chu, S. F. 1968. On estimating the industry production function. *American economic review*, **58**, 826–839. (Not cited.)

Alexander, Jeffrey A, Lee, Shoou-Yih D, Griffith, John R, Mick, Stephen S, Lin, Xihong, & Banaszak-Holl, Jane. 1999. Do market-level hospital and physician resources affect small area variation in hospital use? *Medical care research and review*, **56**(1), 94–117. (Not cited.)

Andersen, Ronald, & Newman, John F. 1973. Societal and individual determinants of medical care utilization in the united states. *The milbank memorial fund quarterly. health and society*, 95–124. (Not cited.)

Andersen, Ronald M. 1995. Revisiting the behavioral model and access to medical care: does it matter? *Journal of health and social behavior*, 1–10. (Not cited.)

Atella, Vincenzo, Belotti, Federico, Depalo, Domenico, & Mortari, Andrea Piano. 2014. Measuring spatial effects in the presence of institutional constraints: The case of italian local health authority expenditure. *Regional science and urban economics*, **49**, 232 – 241. (Not cited.)

Auteri*, Monica, & Costantini, Mauro. 2004. Is social protection a necessity or a luxury good? new multivariate cointegration panel data results. *Applied economics*, **36**(17), 1887–1898. (Not cited.)

Badin, Luiza, Daraio, Cinzia, & Simar, Leopold. 2010. Optimal bandwidth selection for conditional efficiency measures: A data-driven approach. *European journal of operational research*, **201**(2), 633–640. (Not cited.)

Bankauskaite, Vaida, & Dargent, G. 2007. Health systems performance indicators: methodological issues. *Presupuesto y gasto publico*, **49**, 125 – 137. (Not cited.)

Barros, PP. 1998. The black box of health care expenditure growth determinants. *Health economics*, **7**(6), 533–544. (Not cited.)

Blomqvist, Åke G, & Carter, Richard AL. 1997. Is health care really a luxury? *Journal of health economics*, **16**(2), 207–229. (Not cited.)

Bloomfield, J. 2006. *Controlling, cajoling or co-operating?* Tech. rept. Council of European Municipalities and Regions, Brussels. (Not cited.)

Bohning, D., Dietz, E., & Schlattmann, P. 2000. Space-time mixture modelling of public health data. *Statistics in medicine*, **19(17-18)**, 2333–2344. (Not cited.)

Bollino, C.A., & Polinori, P. 2007. Ricostruzione del valore aggiunto su scala comunale e percorsi di crescita a livello micro-territoriale: il caso dell'umbria. *Rivista di scienze regionali*, **2**, 35–73. (Not cited.)

Breyer, Friedrich, & Felder, Stefan. 2006. Life expectancy and health care expenditures: a new calculation for germany using the costs of dying. *Health policy*, **75**(2), 178–186. (Not cited.)

Carinci, F., Gool, K. Van, Mainz, J., Veillard, J., Pichora, E.C., & Januel, J.M. 2015. Towards actionable international comparisons of health system performance: expert revision of the oecd framework and quality indicators. *International journal for quality in health care*, **27**, 137–146. (Not cited.)

Carr-Hill, Roy A, Sheldon, Trevor A, Smith, Peter, Martin, Stephen, Peacock, Stuart, & Hardman, Geoff. 1994. Allocating resources to health authorities: development of method for small area analysis of use of inpatient services. *Bmj: British medical journal*, **309**(6961), 1046. (Not cited.)

Cazals, C., Florens, J., & Simar, L. 2002. Nonparametric frontier estimation: A robust approach. *Journal of econometrics*, **106 (1)**, 1–25. (Not cited.)

Charnes, A., Cooper, W., & Rhodes, W. 1978. Measuring the efficiency of decision making units. *European journal of operational research*, **2(4)**, 429 – 444. (Not cited.)

Cherchye, L., Moesen, W., & Puyenbroeck, T. 2004. Legitimately diverse, yet comparable: On synthesizing social inclusion performance in the eu. *Journal of common market studies*, **42(5)**, 919 – 955. (Not cited.)

Cherchye, L., Lovell, K., Moesen, W., & Puyenbroeck, T. V. 2005. *One market, one number? a composite indicator assessment of eu internal market dynamics*. Tech. rept. Working Paper Series ces0513, Katholieke Universiteit Leuven, Centrum voor Economische Studien. (Not cited.)

Cho, Cheol-Joo. 1998. An equity-efficiency trade-off model for the optimum location of medical care facilities. *Socio-economic planning sciences*, **32**(2), 99–112. (Not cited.)

Chow, G.C., & Lin, A. 1971. Best linear unbiased interpolation, distribution, and extrapolation of time series by related series. *The rev. of economics and statistics*, **53(4)**, 372–375. (Not cited.)

Crivelli, Luca, Filippini, Massimo, & Mosca, Ilaria. 2006. Federalism and regional health care expenditures: an empirical analysis for the swiss cantons. *Health economics*, **15**(5), 535–541. (Not cited.)

Culyer, AJ. 1990. Cost containment in europe in oecd. *Health care systems in transition*, 29–40. (Not cited.)

Daraio, C., & Simar, L. 2005. Introducing environmental variables in nonparametric frontier models: a probabilistic approach. *Journal of productivity analysis*, **24(1)**, 93–121. (Not cited.)

Daraio, C., & Simar, L. 2007. *Advanced robust and nonparametric methods in efficiency analysis*. Springer. (Not cited.)

De Muro, P., Mazziotta, M., & Pareto, A. 2010. Composite indices of development and poverty: An application to mdgs. *Social indicators research*, 1–18. (Not cited.)

Deprins, D., Simar, L. L., & Tulkens, H. 1984. *The performance of public enterprizes: Concepts and measurements*. North-Holland. Chap. Measuring Labor Inefficiency in Post Offices, pages 243 – 267. (Not cited.)

Despotis, D. K. 2005. Measuring human development via data envelopment analysis: the case of asia and the pacific. *Omega*, **33**(5), 385 – 390. OR and its Applications. (Not cited.)

Di Matteo, Livio. 2005. The macro determinants of health expenditure in the united states and canada: assessing the impact of income, age distribution and time. *Health policy*, **71**(1), 23–42. (Not cited.)

Di Matteo, Livio, & Di Matteo, Rosanna. 1998. Evidence on the determinants of canadian provincial government health expenditures: 1965–1991. *Journal of health economics*, **17**(2), 211–228. (Not cited.)

Dormont, Brigitte, Grignon, Michel, & Huber, Hélène. 2006. Health expenditure growth: reassessing the threat of ageing. *Health economics*, **15**(9), 947–963. (Not cited.)

EU Directorate-General for Economic and Financial Affairs and Economic Policy Committee (Ageing Working Group). 2016. *Joint report on health care and long-term care systems & fiscal sustainability*. Institutional Papers 37. (Not cited.)

Evans, Robert G, & Stoddart, Gregory L. 1990. Producing health, consuming health care. *Social science & medicine*, **31**(12), 1347–1363. (Not cited.)

Evans, Robert G, McGrail, Kimberlyn M, Morgan, Steven G, Barer, Morris L, & Hertzman, Clyde. 2001. Apocalypse no: population aging and the future of health care systems. *Canadian journal on aging/la revue canadienne du vieillissement*, **20**(S1), 160–191. (Not cited.)

Farrell, M. J. 1957. The Measurement of Productive Efficiency. *Journal of the royal statistic society*, **120**(III), 253–281. (Not cited.)

Felder, Stefan, Meier, Markus, & Schmitt, Horst. 2000. Health care expenditure in the last months of life. *Journal of health economics*, **19**(5), 679–695. (Not cited.)

Ferré, Francesca, Giulio, Antonio, Valerio, Luca, Longhi, Silvia, Lazzari, Agnese, Fattore, Giovanni, Ricciardi, Walter, & Maresso, Anna. 2014. Health System Review: Italy. *Health systems in transition*, **16**(4), 1–168. (Not cited.)

France, G, & Taroni, F. 2005. The evolution of health-policy making in {Italy}. *Journal of health politics, policy and law*, **30**(1-2), 169–188. (Not cited.)

France, George. 1997. Cross-border flows of italian patients within the european union: an international trade approach. *The european journal of public health*, **7**(3), 18–25. (Not cited.)

Gerdtham, Ulf-G. 1993. The impact of aging on health care expenditure in sweden. *Health policy*, **24**(1), 1–8. (Not cited.)

Getzen, Thomas E. 1992. Population aging and the growth of health expenditures. *Journal of gerontology*, **47**(3), S98–S104. (Not cited.)

Giannoni, Margherita, & Hitiris, Theodore. 2002. The regional impact of health care expenditure: the case of italy. *Applied economics*, **34**(14), 1829–1836. (Not cited.)

Goddard, M., & Jacobs, R. 2009. *Performance measurement for health system improvement: Experiences, challenges and prospects*. Cambridge University Press, Cambridge. Chap. Using composite indicators to measure performance in health care, pages 339–368. (Not cited.)

Goldberger, A. S. 1962. Best linear unbiased prediction in the generalized linear regression model,. *American statistical association journal*, **57**, 369–375. (Not cited.)

Gravelle, Hugh, Sutton, Matthew, Morris, Stephen, Windmeijer, Frank, Leyland, Alastair, Dibben, Chris, & Muirhead, Mike. 2003. Modelling supply and demand influences on the use of health care: implications for deriving a needs-based capitation formula. *Health economics*, **12**(12), 985–1004. (Not cited.)

Greene, W. 2008. The econometric approach to efficiency analysis. *In: The measurement of productive efficiency and productivity change* H. O. Fried and C. A. K. Lovell and S. S. Schmidt, eds. Oxford University Press. (Not cited.)

Grosskopf, S. 1996. Statistical Inference and Nonparametric Efficiency: A Selective Survey. *Journal of productivity analysis*, **7**, 161–176. (Not cited.)

Herwartz, Helmut, & Theilen, Bernd. 2003. The determinants of health care expenditure: testing pooling restrictions in small samples. *Health economics*, **12**(2), 113–124. (Not cited.)

Hitiris, Theo. 1997. Health care expenditure and integration in the countries of the european union. *Applied economics*, **29**(1), 1–6. (Not cited.)

Hitiris, Theo, & Posnett, John. 1992. The determinants and effects of health expenditure in developed countries. *Journal of health economics*, **11**(2), 173–181. (Not cited.)

Ibrahim, J.E. 2001. Performance indicators from all perspectives. *International journal for quality in health care*, **13**, 431 – 2. (Not cited.)

Jencks, S.F., Cuerdon, T., Burwen, D.R., & B, B. Fleming. 2000. Quality of medical care delivered to medicare beneficiaries. *Journal of the american medical association*, **284**, 1670 – 1676. (Not cited.)

Johnson, Andrew, & Kuosmanen, Timo. 2011. One-stage estimation of the effects of operational conditions and practices on productive performance: asymptotically normal and efficient, root-n consistent stonezd method. *Journal of productivity analysis*, **36**, 219–230. (Not cited.)

Karatzas, George. 2000. On the determination of the us aggregate health care expenditure. *Applied economics*, **32**(9), 1085–1099. (Not cited.)

Koenig, Lane, Siegel, Jonathan M, Dobson, Allen, Hearle, Keith, Ho, Silver, & Rudowitz, Robin. 2003. Drivers of healthcare expenditures associated with physician services. *The american journal of managed care*, **9**(1), 34–42. (Not cited.)

Krantz, D., Luce, R., Suppes, P., & Tversky, A. 1971. *Foundations of measurement, vol. 1, additive and polynomial representations*. Academic Press, New York. (Not cited.)

Kuosmanen, Timo, & Kortelainen, Mika. 2012. Stochastic non-smooth envelopment of data: semi-parametric frontier estimation subject to shape constraints. *Journal of productivity analysis*, 1–18. (Not cited.)

Kwon, Soonman. 2003. Health and health care. *Social indicators research*, **62-63**(1-3), 171–186. (Not cited.)

Lauer, Jeremy, Lovell, CA Knox, Murray, Christopher, & Evans, David. 2004. World health system performance revisited: the impact of varying the relative importance of health system goals. *Bmc health services research*, **4**(1), 19. (Not cited.)

LeSage, J., & Pace, R. K. 2009. *Introduction to spatial econometrics*. CRC Press. (Not cited.)

Mahlberg, B., & Obersteiner, M. 2001. *Remeasuring the hdi by data envelopment analysis*. Tech. rept. IIASA, Interim Report IR-01-069, Laxemburg, Austria. (Not cited.)

Mapelli, V. 2012. *Il sistema sanitario italiano*. il Mulino, Bologna. (Not cited.)

Meeusen, W., & van den Broeck, J. 1977. Efficiency estimation from cobb-douglas production functions with composed error. *International economic review*, **18**, 435 – 444. (Not cited.)

Mosca, Ilaria. 2007. Decentralization as a determinant of health care expenditure: empirical analysis for oecd countries. *Applied economics letters*, **14**(7), 511–515. (Not cited.)

Munda, G., & Nardo, M. 2005. *Constructing consistent composite indicators: the issue of weights*. Tech. rept. EUR 21834 EN, European Commission. (Not cited.)

Nardo, M., Saisana, M., Saltelli, A., Tarantola, S., Hoffman, A., & Giovannini, E. 2005. Handbook on constructing composite indicators: Methodology and user guide. *Oecd statistics working papers 2005/3, oecd, statistics directorate*. (Not cited.)

Newhouse, Joseph P. 1977. Medical-care expenditure: a cross-national survey. *The journal of human resources*, **12**(1), 115–125. (Not cited.)

Newhouse, Joseph P. 1992. Medical care costs: how much welfare loss? *The journal of economic perspectives*, **6**(3), 3–21. (Not cited.)

Nuti, S., Daraio, C., Speroni, C., & Vainieri, M. 2011. Relationships between technical efficiency and the quality and costs of health care in italy. *International journal for quality in health care*, **23 (3)**, 324 – 330. (Not cited.)

Okunade, Albert A, & Murthy, Vasudeva NR. 2002. Technology as a 'major driver'of health care costs: a cointegration analysis of the newhouse conjecture. *Journal of health economics*, **21**(1), 147–159. (Not cited.)

Paakkonen, Jenni, & Seppala, T.T. 2014. Using composite indicators to evaluate the efficiency of health care system. *Applied economics*, **46:19**, 2242 – 2250. (Not cited.)

Pammolli, Fabio, & Salerno, Nicola Carmine. 2011 (June). *Le differenze regionali nella governance della spesa sanitaria - SaniRegio-2011*. Working Papers 02-2011. Competitività Regole Mecati (CERM). (Not cited.)

Park, B. U., & Simar, L. 1994. Efficient semiparametric estimation in a stochastic frontier model. *Journal of the american statistical association*, **89**(427), 929–936. (Not cited.)

Park, Byeong U., Sickles, Robin C., & Simar, Leopold. 2007. Semiparametric efficient estimation of dynamic panel data models. *Journal of econometrics*, **136**(1), 281–301. (Not cited.)

Pauly, Mark V. 1986. Taxation, health insurance, and market failure in the medical economy. *Journal of economic literature*, **24**(2), 629–675. (Not cited.)

Polasek, W., & Sellner, R. 2008. *Spatial chow-lin methods: Bayesian and ml forecast comparisons,*. Tech. rept. Rimini Centre for Economic Analysis (RCEA), working paper 38-08. (Not cited.)

Porell, Frank W, & Adams, E Kathleen. 1995. Hospital choice models: a review and assessment of their utility for policy impact analysis. *Medical care research and review*, **52**(2), 158–195. (Not cited.)

Rice, Nigel, Smith, Peter, *et al.* 1999. *Approaches to capitation and risk adjustment in health care: an international survey*. University of York, Centre for Health Economics York. (Not cited.)

Roberts, Jennifer. 2000. Spurious regression problems in the determinants of health care expenditure: a comment on hitiris (1997). *Applied economics letters*, **7**(5), 279–283. (Not cited.)

Royston, GHD, Hurst, JW, Lister, EG, & Stewart, PA. 1992. Modelling the use of health services by populations of small areas to inform the allocation of central resources to larger regions. *Socio-economic planning sciences*, **26**(3), 169–180. (Not cited.)

Santana, P. 1993. Acessibilidade e utilizaçao dos serviços de saúde. ensaio metodológico em economia da saúde. *Faculdade de letras da universidade de coimbra. universidade de coimbra: Coimbra*. (Not cited.)

Santana, Paula. 1999. Geografia das desigualdades em saúde e no estado de saúde. *Livro de homenagem a augusto mantas*, 179–235. (Not cited.)

Seshamani, Meena, & Gray, Alastair. 2004. Ageing and health-care expenditure: the red herring argument revisited. *Health economics*, **13**(4), 303–314. (Not cited.)

Simar, Leopold, & Wilson, Paul W. 2007. Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, **136**(1), 31 – 64. (Not cited.)

Smith, P. 2002a. *Measuring up: Improving health system performance in oecd countries*. OECD. Chap. Developing composite indicators for assessing health system efficiency, pages 295–318. (Not cited.)

Smith, P.C. 2002b. Measuring health system performance. *European journal of health economics*, **3 (3)**, 145 – 148. (Not cited.)

Sorensen, Peter Birch. 2014 (July). *Reforming public service provision: what have we learned?* University of Copenhagen, EPRU and CESifo, Background paper for keynote lecture presented at the CESifo Venice Summer Institute Workshop on REFORMING THE PUBLIC SECTOR at Venice International University. (Not cited.)

Stearns, Sally C, & Norton, Edward C. 2004. Time to include time to death? the future of health care expenditure predictions. *Health economics*, **13**(4), 315–327. (Not cited.)

Storrie, D., & Bjurek, H. 2000. *Benchmarking european labour market performance with efficiency frontier techniques*. Tech. rept. CELMS Discussion papers, Goteborg University. (Not cited.)

Toth, F. 2014. How health care regionalization in italy is widening the north-south gap. *Health economics policy and law*, **9(3)**, 231–249. (Not cited.)

Tulchinsky, T.H., & Varavikova, E. 2009. *The new public health*. Elsevier Academic Press. (Not cited.)

Vallejos, R. 2008. Assessing the association between two spatial or temporal sequences. *Journal of applied statistics*, **35**, 1323–1343. (Not cited.)

van Doorslaer, E., & Koolman, X. 2004. Explaining the differences in income-related health inequalities across european countries. *Health economics*, **13(7)**, 609–628. (Not cited.)

Vansnick, J. 1990. *Measurement theory and decision aid*. Springer-Verlag, Berlin. Chap. Readings in multiple criteria decision aid, pages 81–100. (Not cited.)

Vidoli, F., & Mazziotta, C. 2013. Robust weighted composite indicators by means of frontier methods with an application to european infrastructure endowment. *Italian journal of applied statistics*, **23(2)**, 259 – 282. (Not cited.)

Vidoli, F., Cardillo, C., Ferrara, G., & Mazzeo, P. 2012. *Sustainable value in italian agricultural field: parametric and semi-parametric frontier efficiency models*. 2 step evaluation in Journal of Productivity Analysis. (Not cited.)

Vincenty, T. 1975. Direct and inverse solutions of geodesics on the ellipsoid with application of nested equations. *Survey review*, **23(176)**, 88–93. (Not cited.)

Werblow, Andreas, Felder, Stefan, & Zweifel, Peter. 2007. Population ageing and health care expenditure: a school of 'red herrings'? *Health economics*, **16**(10), 1109–1126. (Not cited.)

Witte, K. D., & Rogge, N. 2009. *Accounting for exogenous influences in a benevolent performance evaluation of teachers*. Tech. rept. Working Paper Series ces0913, Katholieke Universiteit Leuven, Centrum voor Economische Studien. (Not cited.)

Zweifel, Peter, Felder, Stefan, & Meiers, Markus. 1999. Ageing of population and health care expenditure: a red herring? *Health economics*, **8**(6), 485–496. (Not cited.)