



Dove va la spesa sanitaria (potenziale)?*

di Fabio Pammolli e Nicola C. Salerno

Dove va la spesa sanitaria? È una domanda che da qualche anno sta tenendo impegnate le riflessioni di economisti e *policy maker*, mettendo alla prova le caratteristiche di diverse metodologie di proiezione. Dalla risposta dipende anche la valutazione su dove vanno le finanze pubbliche, data la rilevanza che questo capitolo di spesa mediamente ha sui bilanci degli Stati, oltre alla sua importanza sociale e politica.

Le dinamiche storiche di lungo periodo

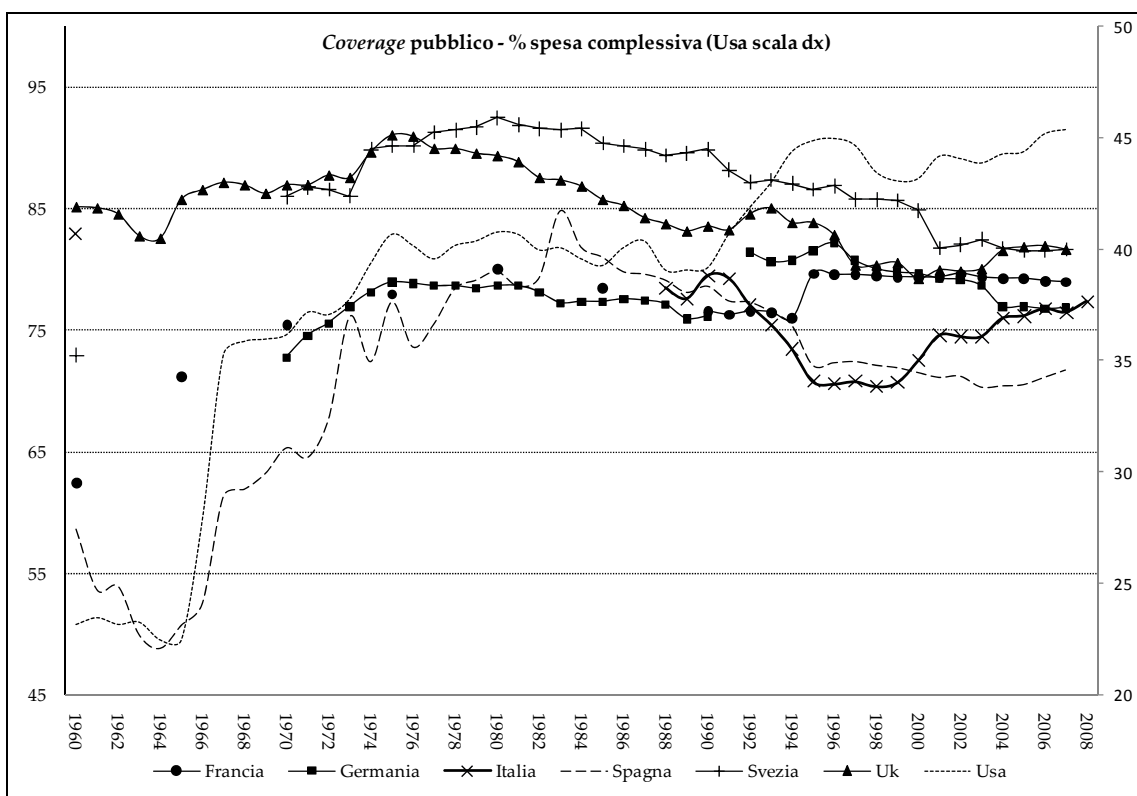
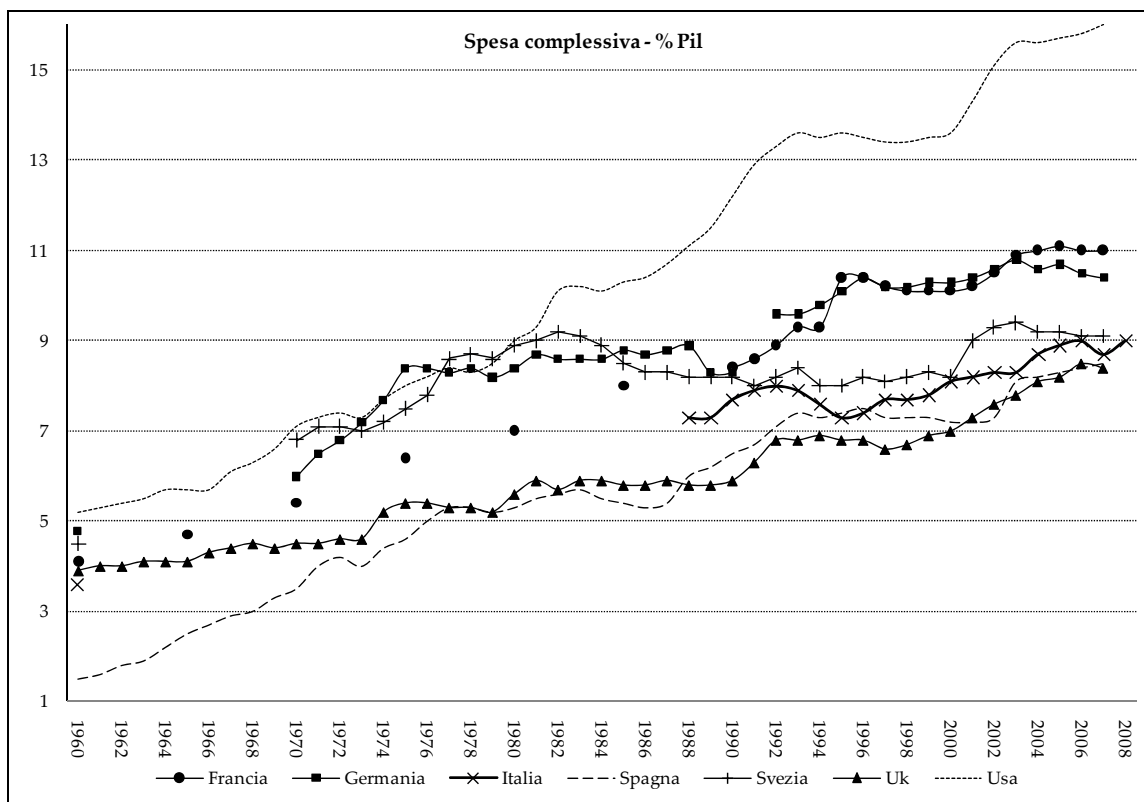
Prima di interrogarci sul futuro, è sempre bene guardare a quel che è già accaduto. I dati Ocse "Health data - 2009" forniscono serie dell'incidenza della spesa sanitaria complessiva (pubblica e privata) sul Pil, complete dal 1960 ad oggi, per Spagna, Uk e Usa. Per la Francia la serie parte dal 1964, per la Germania e per la Svezia dal 1970, per l'Italia soltanto dal 1988. In questi ultimi casi, l'incompletezza delle serie è dovuta alla discontinuità creata da revisioni di contabilità nazionale. Tuttavia, i dati di lungo periodo son troppo rari e preziosi per essere dimenticati del tutto, persino nel loro ordine di grandezza. Le serie più recenti possono essere integrate dalla versioni precedenti di "Health data", per completare i dati al 1960, e coprire per il maggior numero dei Paesi un periodo di cinquant'anni.

Nei principali *Partner Ue-15* e negli Usa, l'incidenza della spesa sanitaria complessiva sul Pil si è triplicata o quadruplicata dal 1960 ad oggi. Per la Spagna il multiplo è addirittura per sette/otto. Questo sostenuto *trend* si è realizzato mentre il *coverage* pubblico è variato di poco nei Paesi in cui partiva già elevato, con valori superiori al 75%; mentre, in Francia, in Spagna e soprattutto in Usa, è sensibilmente cresciuto, passando nel primo caso da meno del 65 a più del 75%, nel secondo caso da circa il 55 a poco meno del 76%, e nel terzo da meno del 25 a circa il 45%. Nei Paesi in cui il *coverage* è variato poco, gli incrementi di incidenza della spesa complessiva possono essere riferiti *tout court* anche alla componente di spesa pubblica. Per Paesi come la Francia, la Spagna, gli Usa, se si tiene conto dell'andamento crescente del *coverage*, l'incidenza della spesa sanitaria pubblica è aumentata, dal 1960 ad oggi, di multipli superiori alle sette/otto volte.

Una crescita così intensa si è accompagnata ad un altrettanto sorprendente aumento della vita attesa alla nascita e della vita attesa a 65 anni. La prima ha guadagnato circa un anno ogni quattro/cinque; la seconda circa un anno ogni dieci per gli uomini, e un anno ogni otto/nove. L'allungamento della vita è sicuramente un risultato dell'aumento delle condizioni di salute sostenute dalla maggior spesa sanitaria complessiva e pubblica; ma emerge anche come, negli scorsi cinquant'anni, l'interazione tra spesa e salute non sia giunta a nessuna stabilizzazione, in cui sorgesse anche un controeffetto della più elevata salute sui risparmi di spesa. Le serie storiche si

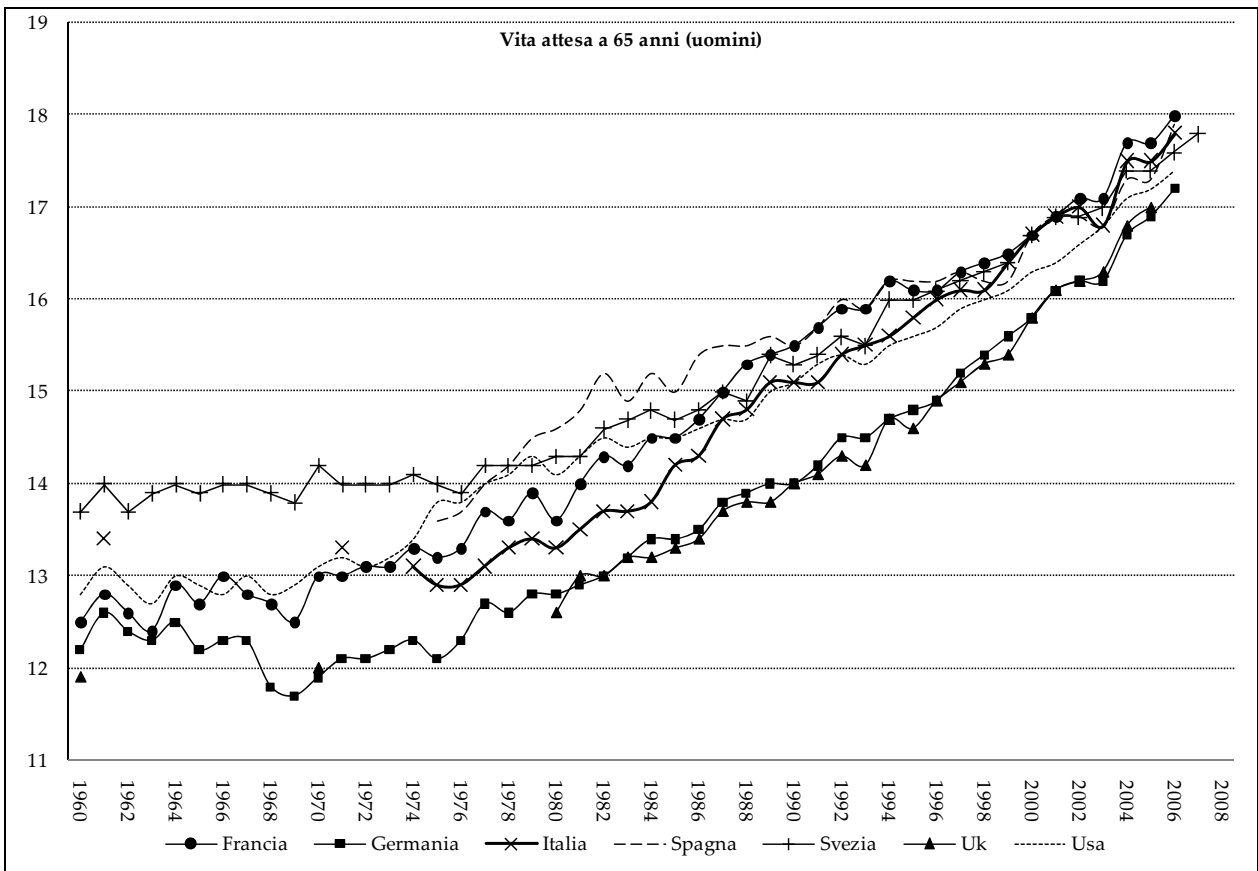
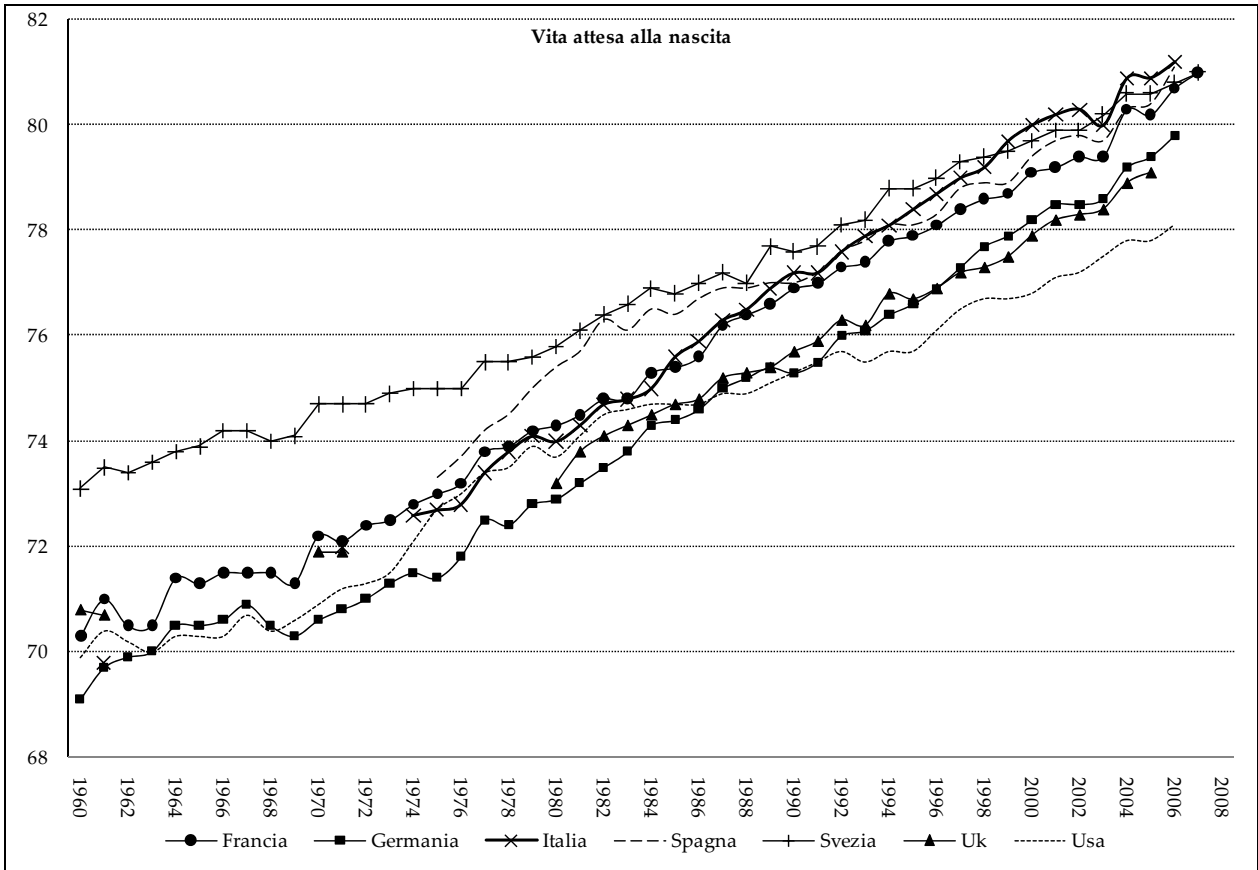
* Paper per il 1° sottogruppo "Andamenti della spesa, sostenibilità e modelli di finanziamento" del gruppo di studio di Astrid su "Le politiche sanitarie", coordinato da Claudio De Vincenti, Renato Finocchi Ghersi e Andrea Tardiola

sono mantenute in costante ed ininterrotta coevoluzione crescente. E questo è avvenuto nonostante la generale riduzione delle probabilità di morte a tutte le età¹.

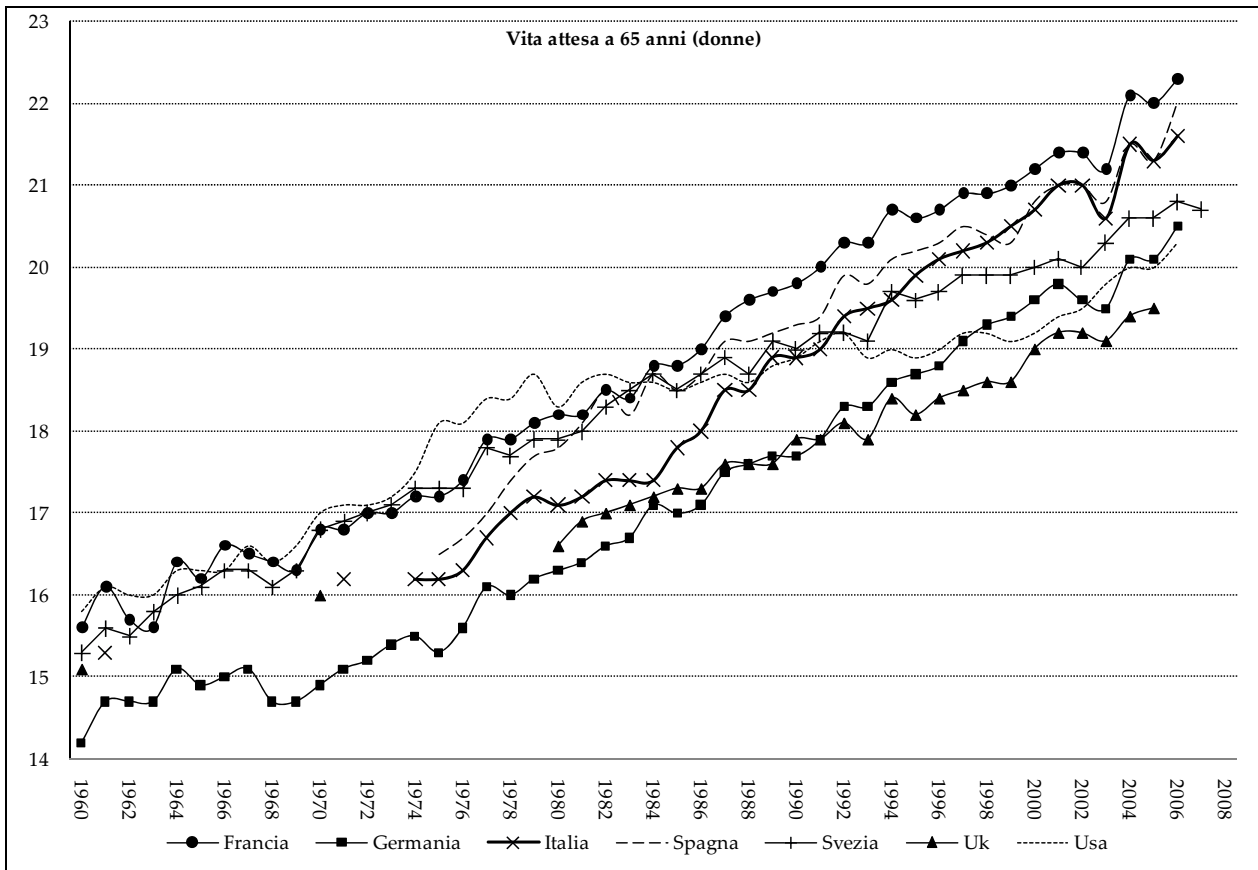


fonte: elaborazione Cerm su Ocse

¹ Una quota significativa della spesa sanitaria si colloca nell'anno del decesso (i cosiddetti *death related costs*). In Italia, per fare un esempio, il numero totale di anni di vita persi (altro indicatore di stato di salute di una comunità utilizzato dall'Ocse, *cfr.* <http://www.ecosante.org/index2.php?base=OCDE&langh=ENG&langs=ENG&sessionid=>) si è ridotto del 74% dal 1960.



fonte: elaborazione Cerm su Ocse



fonte: elaborazione Cerm su Ocse

Le proiezioni Awg-Ecofin²

Se questa è la storia degli ultimi cinquant'anni, quali proiezioni si possono avanzare per i prossimi cinquanta? Ecofin ed Ocse vi hanno dedicato due recenti lavori; ed è proprio dal confronto ragionato fra i due che si possono ottenere informazioni preziose per i *policy maker*.

Il Gruppo di Lavoro sull'Invecchiamento della Popolazione di Ecofin (Awg-Ecofin) è giunto ormai alla sua terza *release* di proiezioni per la spesa sanitaria pubblica *acute* e di lungodegenza. Nello scenario centrale o *Awg-reference*, quello incorporato nell'analisi annuale di sostenibilità di lungo termine delle finanze pubbliche dei *Partner*, la *acute* vede aumentare la sua incidenza sul Pil, da oggi al 2060, di 1,5 p.p. in Ue-15 e Ue-27, e di 1,1 in Italia. Sul fronte della spesa di lungodegenza, gli incrementi di incidenza sono di 1,2 per l'Ue-15, 1,1 per l'Ue-27, e 1,3 per l'Italia.

² Per una analisi dettagliata, cfr. Pammolli F. e N. C. Salerno (2009), "Spesa sanitaria: quali ipotesi per quali proiezioni? Ecofin e Ocse a confronto", Nota Cerm n. 1-2009, liberamente disponibile su <http://www.cermlab.it/argomenti.php?group=sanita&item=36>. Si tenga presente che Ecofin considera un perimetro di spesa sanitaria pubblica che, nel caso dell'Italia, è più ampio di quello relativo alla sola gestione del Ssn, includendo voci appartenenti ai bilanci degli Enti Locali e, in particolare, dei Comuni.

	Spesa sanitaria pubblica - % Pil				
	scenario centrale (o Awg reference)				scenario tecnologico
	acuta		per lungodegenza		acuta
	2007	Δ 2007-2060	2007	Δ 2007-2060	Δ 2007-2060
Francia	8,1	1,2	1,4	0,6	6,8
Germania	7,4	1,8	0,9	1,4	7,2
Italia	5,9	1,1	1,7	1,3	5,3
Spagna	5,5	1,6	0,5	0,9	5,6
Svezia	7,2	0,8	3,5	2,3	5,4
Uk	7,5	1,9	0,8	0,5	7,4
Ue-15	6,9	1,5	1,3	1,2	6,4
Ue-27	6,7	1,5	1,2	1,1	6,3

fonte: elaborazione Cerm su Awg-Ecofin

Per apprezzare a fondo le proiezioni è necessario inquadrarle all'interno delle ipotesi più importanti adottate da Awg-Ecofin, oltre quella di invarianza della *policy*. Per la voce *acute*, a partire della spesa *pro-capite* per fasce di età-sesso nell'anno base (il 2007), questa stessa spesa si evolve nel tempo allo stesso tasso di crescita del Pil *pro-capite*. A questa dinamica si aggiunge una elasticità della domanda di prestazioni al Pil *pro-capite* pari a 1,1 nel 2007, e poi convergente con continuità all'unità al 2060. Infine, tra un anno e l'altro i profili di spesa età-sesso si traslano in avanti della metà dell'incremento di vita attesa, il cosiddetto *dynamic profiling*. Per la voce *long-term*, la spesa *pro-capite* per fasce di età-sesso cresce allo stesso tasso del Pil per occupato (per dar conto dell'effetto "Baumol"), tranne che per le prestazioni *cash* dove il riferimento è il Pil *pro-capite*. L'elasticità delle prestazioni è, questo caso, sempre unitaria. Tra un anno e l'altro le probabilità di evento inabilitante/invalidante si riducono della metà della riduzione della probabilità di morte (*dynamic profiling* applicato alla lungodegenza). Infine, la probabilità di ricevere prestazioni istituzionalizzate e formalizzate (al di fuori della rete familiare e amicale) rimane stabile ai valori del 2007.

Le ipotesi dello scenario centrale, che pure conducono ad un'incidenza continuamente crescente sul Pil senza inversione o stabilizzazione, sono ottimistiche, nella misura in cui considerano i fattori che possono concorrere al contenimento della spesa (il *dynamic profiling* su tutti), mentre lasciano in secondo piano quelli che potrebbero esercitare l'effetto opposto. In particolare ci si riferisce alla dinamica della spesa adagiata a quella del Pil, e all'elasticità della domanda solo inizialmente e per la sola voce *acute* superiore all'unità.

I limiti di questo *set* risaltano soprattutto a confronto con il dato storico che, nonostante il rapporto tra spesa e Pil sia stato spesso influenzato da politiche di controllo e compressione della stessa spesa, riporta di un tasso di crescita della spesa *pro-capite* che mediamente ha sovrastato quello del Pil *pro-capite* di 1-2 p.p. a seconda del Paese.

È questa la ragione per cui Awg-Ecofin include, per la prima volta nella terza *release* di proiezioni, uno scenario "tecnologico", in cui l'elasticità è sempre superiore all'unità³. In questo scenario, per adesso relativo solo alla voce *acute*, gli incrementi di incidenza sul Pil sono significativamente superiori, pari a 6,4 p.p. in Ue-15, 6,3 in Ue-25, e 5,3 in Italia. Più o meno dovunque, la spesa quasi raddoppia in termini di Pil.

³ In particolare, l'elasticità al Pil è pari a 0,7%, ma poi al tasso di crescita così ottenuto si aggiungono 2 p.p. per tener conto degli effetti, sia lato domanda che lato offerta, degli avanzamenti della scienza e della tecnologia.

Le proiezioni Ocse⁴

Lo scenario “tecnologico” di Awg-Ecofin è stato sollecitato dall’esercizio di proiezione precedentemente svolto dall’Ocse sulla componente pubblica della spesa sanitaria. Pur partendo da una metodologia simile e da ipotesi macroeconomiche e demografiche sufficientemente allineate, e pur considerando il *dynamic profiling*, l’Ocse era già giunto a risultati molto diversi da quelli dello scenario centrale di Awg-Ecofin e simili, invece, a quelli dello scenario “tecnologico”. Se si approfondiscono le ipotesi di base, ci si accorge che la differenza si spiega in base al tasso di crescita della spesa, all’elasticità della domanda e all’andamento della quota dell’assistenza di lungo termine istituzionalizzata e formalizzata.

	Spesa sanitaria pubblica - % Pil	
	<i>scenario a più intensa crescita</i>	
	acuta e per lungodegenza	
	2007	Δ 2007-2060
Francia	8,1	7,1
Germania	8,8	6,0
Italia	6,6	9,4
Spagna	5,5	7,2
Svezia	8,6	4,2
Uk	7,2	5,7
Ue-15	7,2	6,4
Ue-27	6,7	6,9

fonte: elaborazione Cerm su Ocse

Per la voce *acute*, il tasso di crescita della spesa *pro-capite* è pari a quello del Pil *pro-capite* moltiplicato 1,2 e con l’aggiunta di uno *spread* di 1,1 p.p. gradualmente convergente a zero entro il 2050. In alternativa a questa ipotesi, negli scenari in cui si ipotizza elasticità unitaria, si aggiunge sempre uno *spread* di 1 p.p. tra tasso di crescita della spesa e del Pil *pro-capite*. Per la voce *long-term*, se è vero che si ipotizza la stessa dinamica del Pil per occupato, si aggiungono tuttavia due fattori determinanti nel traino: la quota di *formal assistance* è inversamente proporzionale al tasso di partecipazione al mercato del lavoro nella fascia 50-64, quest’ultimo a sua volta ipotizzato convergere al 70% in tutti i Paesi⁵; e per i tassi di dipendenza si individua una doppia dinamica, la prima in riduzione su archi decennali di pari passo con il miglioramento dello stato medio di salute (il *dynamic profiling*), l’altro in aumento *year-by-year* come conseguenza di dinamiche lato domanda e di aumenti dei costi connessi al progresso scientifico-tecnologico.

Nella somma di *acute* e *long-term*, Ocse mostra come, se il *set* delle ipotesi si amplia per dare più spazio alle determinanti di crescita *extra* demografiche, gli incrementi di incidenza sul Pil di lungo periodo (al 2050) configurano un raddoppio e in alcuni casi, come in Italia, un più che raddoppio⁶.

⁴ Cfr. Pammolli-Salerno (2009), *cit.*

⁵ Se aumenta il tasso di occupazione, le possibilità di assistenza continuata familiare o amicale si riducono. In Italia il tasso di occupazione complessivo è del 66%, con quello delle donne poco al di sopra del 50%).

⁶ Se si riunissero in un unico scenario i driver di crescita a più forte impatto, che l’Ocse distribuisce in scenari separati, l’ordine di grandezza degli incrementi di incidenza di lungo periodo diverrebbe ancora più macroscopico.

Dinamica potenziale o dinamica già corretta dalla policy?

Se la differenza tra Ecofin e Ocse risiede soprattutto in come le proiezioni raffigurano il legame tra la dinamica di spesa e lo sviluppo economico-sociale⁷, allora preliminarmente alla domanda “Dove va la spesa?” è un’altra domanda: “Quale spesa?”. Quella potenziale, che si realizzerebbe se la domanda potesse sempre essere soddisfatta al meglio, sia in quantità che in qualità? O quella che già sconta interventi di stabilizzazione e di riconduzione all’interno dei vincoli di bilancio? La differenza è sostanziale, perché il *trade-off* cui andiamo incontro, tra da un lato *coverage* e qualità/innovatività delle prestazioni e, dall’altro, sostenibilità finanziaria, lo si coglie nella sua reale portata solo con riferimento alla spesa potenziale; mentre proiezioni che sottintendono azioni contenitive di *policy* ridimensionano quel *trade-off* senza, nel contempo, dar conto della praticabilità delle azioni di stabilizzazione negli anni a venire.

Se quel che serve è la proiezione di una grandezza che testimoni quanto la crescita fisiologica della spesa sanitaria possa entrare in contrasto con la disponibilità di risorse, allora è necessario guardare al potenziale, perché sarà su quello che si dovrà misurare in futuro la capacità di *governance*.

In questa prospettiva, la modellistica come quella utilizzata da Ecofin e da Ocse⁸, spesso criticata per la semplicità/schematicità di struttura e per il ricorso alla calibrazione dei parametri (a cominciare da quello dell’elasticità della domanda), ha un aspetto positivo rispetto ai modelli econometrici con parametri stimati endogenamente: mentre questi ultimi rischiano di scaricare sulla stima dei parametri l’azione storicamente svolta dalla *policy*, creando inevitabilmente confusione tra dinamica potenziale e dinamica (già) governata della spesa futura, soluzioni *à la* Ecofin-Ocse permettono di utilizzare per i parametri dei valori di lungo periodo frutto di analisi *cross-country*, in grado di gettare luce sulla parte più strutturale dei legami, al di là di quanto possa emergere all’interno delle singole “stagioni” di *policy* che caratterizzano i Paesi⁹.

Dalle proiezioni di spesa sanitaria all’equilibrio socio-economico

Le proiezioni Ecofin-Ocse suggeriscono che nel lungo periodo, senza interventi correttivi di *policy*, l’incidenza della spesa sanitaria pubblica sul Pil possa arrivare a raddoppiarsi o più che raddoppiarsi. È probabile che in futuro si ripeta quanto già accaduto dal 1960 ad oggi, con aumenti straordinari (multipli) di quell’incidenza. È rispetto a questi ordini di grandezza che la *governance* è chiamata, oggi, a progettare per tempo delle soluzioni.

Una simile dinamica potenziale non è sostenibile. Bastano pochi esempi per rendersene conto. In Italia, per stabilizzare l’incidenza della spesa sanitaria pubblica sul Pil al livello attuale, nel 2050 il *coverage* dovrebbe ridursi dall’attuale 75% circa al 50% se si fa riferimento allo scenario centrale di Awg-Ecofin, oppure significativamente al di sotto del 50% se si fa riferimento allo scenario a più intensa crescita tracciato dall’Ocse. Si manifesterebbero problemi sia per la copertura dei livelli essenziali di assistenza sia, a maggior ragione, per l’inclusione dell’innovazione nei livelli essenziali.

Se le proiezioni sono quelle Ocse o dello scenario “tecnologico” di Awg-Ecofin, per ricondurre il debito pubblico italiano al 60% del Pil nel 2050, sarebbe necessario un profilo di

⁷ Ricadono in questo legame l’effetto diretto della crescita del Pil, quello che matura sul lato domanda come aumento della richiesta di prestazioni e tra queste di prestazioni innovative, e quello che matura sul lato dell’offerta come messa a disposizione di tecnologie e tecniche innovative e di prestazioni professionalizzate.

⁸ Basata su estrapolazione di profili di spesa *pro-capite* per fasce di età-sesso.

⁹ Nel caso dell’Italia, per esempio, la forte compressione di spesa condotta durante tutti gli anni Novanta ha sicuramente introdotto una divaricazione tra dinamica potenziale e dinamica contabilizzata nei documenti economici ufficiali, che si è spesso tradotta in allungamento delle file d’attesa, cadute della qualità, stasi dei progetti di adeguamento infrastrutturale, indebitamento sommerso delle Asl e delle Regioni.

avanzi di bilancio continuamente crescente sino superare il 10% del Pil, con conseguente totale “ingessamento” della politica annuale di bilancio¹⁰.

Oggi, in Italia, ogni persona attiva finanzia la spesa pubblica per pensioni e sanità dedicata agli ultrasessantacinquenni con risorse equivalenti ad oltre il 31% del Pil *pro-capite*. Al 2050, questo valore sarà del 42,2% se si considerano le proiezioni per pensioni e sanità dello scenario centrale di Awg-Ecofin; e del 48,6% se le proiezioni della sanitaria sono quelle dello scenario a più intensa crescita dell’Ocse¹¹. Nella media Ue-15, i tre valori sono rispettivamente 25,2, 38,3 e 45%¹². Oggi, in Italia, ogni persona occupata finanzia la spesa pubblica per pensioni e sanità dedicata agli ultrasessantacinquenni con risorse equivalenti ad oltre il 52,6% del Pil *pro-capite*. Al 2050, questo valore sarà del 62,3% se si considerano le proiezioni per pensioni e sanità dello scenario centrale di Awg-Ecofin; e del 71,5% se le proiezioni della sanitaria sono quelle dello scenario a più intensa crescita dell’Ocse. Nella media Ue-15, i tre valori sono rispettivamente 36,2, 51,5 e 60,4%. I sistemi di finanziamento *monopillar* pubblici, che contraddistinguono quasi tutti i Paesi europei, reggeranno a cambiamenti di questa entità?¹³

Se la dinamica della spesa si posizionasse anche solo in una situazione intermedia tra lo scenario centrale di Awg-Ecofin e quello a più intensa crescita dell’Ocse, in Italia nel 2050 l’incidenza della spesa sanitaria pubblica sul Pil arriverebbe ad approssimare quella della spesa pensionistica pubblica (nel lungo periodo stabilizzata al 14%). Quale spazio rimarrebbe per finanziare gli altri istituti di *welfare*? Questa domanda non può rimanere senza risposta, tenuto conto che la diversificazione degli istituti di *welfare* (per famiglia, minori, povertà, accesso alla casa, *education*, mercato del lavoro, etc.) ha effetti positivi sia sullo stato di salute (fisica e mentale) che sullo sviluppo economico, e con ciò rientra a pieno titolo nelle azioni di *policy* su cui puntare per governare il *trade-off* tra esigenze di spesa sanitaria e disponibilità di risorse. E tenuto altresì conto che gli istituti di *welfare* a finalità redistributiva-assicurativa non possono fare a meno del finanziamento a ripartizione, oggi in Italia esaurito dai due *monopillar* pubblici delle pensioni e della sanità.

Conclusioni e suggerimenti di policy

Nei prossimi decenni tutti i Paesi ad economia e *welfare* sviluppati dovranno fronteggiare spese sanitarie fortemente crescenti, con *trend* potenziali che al 2050 potrebbero raddoppiare o più che raddoppiare l’incidenza della componente pubblica sul Pil. Alla domanda “Dove va la spesa?”, la risposta che appare più responsabile è “È questo l’ordine di grandezza del potenziale di crescita sul Pil”, oltre che delle pressioni che potranno giungere alle finanze pubbliche e all’equilibrio socio-economico.

Di fronte a questo scenario, l’efficientamento dei sistemi sanitari nazionali rimarrà un tassello importante ma non potrà da solo risolvere. Tutti i Paesi, di tradizione *bismarckiana* o *beveridgiana* che essi siano, dovranno affrontare ristrutturazioni profonde dei loro sistemi sanitari e di *welfare*, per addivenire ad un assetto in cui il *trade-off* tra *coverage* e sostenibilità venga governato sulla base di scelte positive e non subito passivamente. Punti cardine dei processi di riforma saranno:

¹⁰ Cfr. Pammolli F. e N. C. Salerno (2008), “La sanità in Italia – Federalismo, regolazione dei mercati, sostenibilità delle finanze pubbliche”, ed. il Mulino – Arel (http://www.cermlab.it/la_sanita_in_italia.php).

¹¹ L’esercizio è descritto in dettaglio in Pammolli-Salerno (2009), *cit.*.

¹² Molto interessante, su questo punto, un confronto con la contabilità intergenerazionale descritta per gli Usa in Kotlikoff L. J. (2007), “The Healthcare Fix - Universal Insurance for All Americans”, The MIT Press. Ne emerge come l’Europa, e in particolare l’Italia, si trovino oggi a vivere sproporzioni che gli Usa paventano di raggiungere tra quindici/venti anni. Per inciso, il volume contiene numerosi spunti per valutare i vincoli finanziari con cui dovranno misurarsi i progetti dell’Amministrazione Obama di estendere il *coverage* pubblico.

¹³ Sono coinvolti tutti, sia sistemi *à la Bismarck* che sistemi *à la Beveridge*. Variazioni significative (tra i 10 e i 20 p.p.) del carico sugli attivi e sugli occupati si realizzeranno, a policy invariata, anche per Paesi, come la Francia, la Germania e la Svezia, che sinora hanno “convissuto” bene, sicuramente meglio dell’Italia, con elevati cunei contributivi sui redditi da lavoro.

- l'adozione dell'universalismo selettivo, come la soluzione più flessibile e più adatta a individuare ordini di priorità e a salvaguardare la funzione redistributiva all'interno di vincoli dei risorse;
- la diversificazione multipilastro del finanziamento delle pensioni e della sanità, per affiancare al criterio a ripartizione pubblico (*pay-as-you-go*) quello basato su investimenti reali di lungo termine sui mercati¹⁴;
- la diversificazione degli istituti di *welfare*, per dare risposta alle esigenze che sorgono in tutte le fasi della vita del singolo e della famiglia, promuovere la partecipazione attiva alla società e al mercato del lavoro, e saldare l'adeguatezza delle tutele sociali su un percorso di solido sviluppo economico (il *welfare-to-work*).

La definizione operativa di livelli essenziali di assistenza sostenibili e il disegno di schemi di compartecipazione ai costi sono strumenti dell'universalismo selettivo. Per l'Italia, inoltre, tutti i punti suggeriti intersecano sotto più profili la trasformazione federalista in corso. Una ragione in più per non dilungare ancora la chiusura della transizione federalista.

¹⁴ I due criteri hanno virtù che si complimentano e difetti che si controbilanciano, sia a livello microeconomico (per quanto concerne gli incentivi individuali alla produzione e la diversificazione del rischio) che a livello macroeconomico (in relazione alla composizione degli investimenti tra nozionali e reali e al sostegno del risparmio di lungo termine).