



QUADERNO 4-2010



Finanza Pubblica e Riforme

**I TREND DELLA SPESA SANITARIA
E LE RAGIONI DEL FINANZIAMENTO *MULTIPILLAR*
CON COMPONENTE AD ACCUMULAZIONE REALE**

F. PAMMOLLI, N. C. SALERNO

Il CeRM è da tempo impegnato sulle tematiche delle proiezioni della spesa sanitaria e del disegno dell'assetto regolatorio e del *mix* di strumenti di finanziamento più adatto ad affrontare quel *trade-off*, tra sostenibilità economica e adeguatezza delle terapie, che nei prossimi anni diverrà sempre più stretto.

Questo Quaderno è fa parte di una serie di contributi, nel contempo di approfondimento e di divulgazione, in cui rientrano anche:

Spesa sanitaria: quali ipotesi per quali proiezioni? Ecofin e Ocse a confronto,

Il pilastro complementare a capitalizzazione in sanità: Tredici buone ragioni per i fondi aperti per il welfare,

La sanità in Italia - Tra federalismo, regolazione dei mercati e sostenibilità delle finanze pubbliche,

L'integrazione pubblico-privato nel finanziamento della sanità e dell'assistenza alla persona,

La spesa sanitaria pubblica in Italia: dentro la scatola nera delle differenze regionali - Il modello SaniRegio.

Tutti i lavori sono liberamente accessibili dal sito del CeRM, www.cermlab.it.

Fp & Ncs

Sommario

Nei prossimi decenni tutti i Paesi ad economia e *welfare* sviluppati dovranno fronteggiare spese sanitarie fortemente crescenti, con *trend* potenziali che nel lungo periodo, al 2050-2060, potrebbero raddoppiare o più che raddoppiare l'incidenza della componente pubblica sul Pil. Così emerge dalle più recenti proiezioni Ecofin, Ocse e Fmi. Non si tratta di casi scolastici di proiezione, dal momento che la dinamica storicamente osservata dal Dopoguerra ad oggi ci ha già posto di fronte ad aumenti di questa proporzione e anche più forti. 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.

E si deve anche considerare che la dinamica storicamente osservata ingloba gli effetti delle politiche di governo e stabilizzazione della spesa. Corrispondentemente, la dinamica potenziale, sottostante i profili di incidenza della spesa sul Pil dal 1960 ad oggi, è stata addirittura superiore. Se per la spesa effettiva si deve parlare di incidenza sul Pil raddoppiata e, in alcuni casi, aumentata di multipli sino a tre e oltre, per la spesa potenziale l'aumento sarebbe stato molto più intenso.

Di fronte a tali prospettive, le conclusioni di questo scritto non voglio essere negative ed allarmiste. Al contrario, è necessario che il *policy maker* si attivi subito affinché gli strumenti a disposizione siano all'altezza di governare e bilanciare costantemente il *trade-off* tra sostenibilità finanziaria e adeguatezza delle prestazioni. Il solo finanziamento a ripartizione che contraddistingue la maggior parte dei sistemi sanitari non potrà reggere l'impatto; ma non potrà essere risolutivo neppure l'affiancamento di finanziamenti privati che si affidino a soluzioni di pura mutualità o di assicurazione.

Criteri di finanziamento, in ambito pubblico o privato, che si affidino esclusivamente o prevalentemente al concorso, anno per anno, di tutti gli assicurati per finanziare il fabbisogno espresso nello stesso anno, inevitabilmente produrranno effetti distorsivi e depressivi. È impensabile, infatti, che la massa reddituale disponibile anno per anno possa sostenere il confronto con una spesa a finanziare che crescerà a ritmi così forti. E assieme agli effetti distorsivi/depressivi è certo che emergeranno fenomeni di razionamento e restringimento del perimetro delle prestazioni offerte (sia nel pubblico che nel privato), oppure rinunce dei diretti interessati a dotarsi di coperture sanitarie (nel privato).

Tutti i Paesi, con sistemi a prevalenza pubblica o privata, e tra i pubblici sia quelli di tradizione *bismarckiana* che *beveridgiana*, dovranno affrontare ristrutturazioni profonde nei meccanismi di finanziamento del loro *welfare*.

Un punto cardine dovrà essere la diversificazione multipilastro del finanziamento, per affiancare al criterio a ripartizione pubblico (*pay-as-you-go*), a quello privato di mutualità pura, e a quello assicurativo privato (*pooling* supportato da riserve matematiche), il criterio basato su investimenti reali di lungo termine sui mercati (l'accumulazione reale). Un punto che, nei meccanismi fondamentali, si presenta *tel quel* sia per la sanità che per le pensioni. Finché l'affiancamento del privato al pubblico avviene rimanendo nell'ambito della messa in comune di risorse prodotte anno per anno, per far fronte alle esigenze di coloro che esprimono, in quello stesso anno, richieste di prestazioni, la diversificazione *multi-pillar* non è in grado di produrre cambiamenti strutturali. Il fattore, in grado di fare la differenza sul piano strutturale, è l'inserimento di sufficienti quote di accumulazione reale, tramite programmi di investimento reale di lungo termine.

Laddove ripartizione, mutualità e *pooling* utilizzano redditi già prodotti (intervengono *ex-post*, per dedicare alla finalità sanitaria quote di risorse già prodotte), i programmi di investimento entrano nei processi di generazione delle risorse e le preparano per tempo. Tramite una idonea gestione di portafoglio, i risparmi andrebbero a selezionare le migliori opportunità di rendimento a livello internazionale, potendosi rivolgere anche ai Paesi più giovani, quelli dove le fasce di età attive saranno ancora per molto tempo più numerose di quelle anziane (una sorta di riequilibrio internazionale degli

effetti dell'invecchiamento). Diminuirebbe la pressione sugli attivi, si guadagnerebbe base imponibile al Fisco nazionale e, nella misura in cui i risparmi trovassero valide opportunità di investimento all'interno, si darebbe impulso alla capitalizzazione dell'economia e alla crescita.

Ovviamente l'accumulazione non è una panacea; le mancano alcune proprietà positive e salienti sia della ripartizione (è il canale più adatto alla coesione *inter* e *inter* generazionale e alle finalità redistributive) che della mutualità e della copertura assicurativa (contro grandi eventi avversi, l'accumulazione potrebbe non essere capiente o, se lo è, potrebbe venire in larga parte assorbita). Inoltre, un eccesso di accumulazione reale esporrebbe troppo al rischio di mercato e potrebbe essere esso stesso fonte di instabilità dei mercati (in concomitanza con i flussi di investimento e di disinvestimento).

Dal «giusto dosaggio» delle modalità di finanziamento dipenderà anche la possibilità di dare copertura finanziaria agli altri istituti di *welfare* a carattere redistributivo (famiglia, figli/minori, conciliazione vita-lavoro, accesso/mantenimento casa di abitazione, contrasto povertà, etc.), che non possono prescindere dalla copertura pubblica e dal finanziamento a ripartizione. Se la ripartizione sarà «monopolizzata» dalla sanità, in aggiunta al carico di finanziamento assorbito dalle pensioni, poco o nullo spazio rimarrà per dare copertura agli altri istituti di *welfare* e agli altri bisogni.

Sostenibilità finanziaria e sostenibilità sociale sono intrinsecamente connesse e passano per il crocevia di un *multi-pillar* che, nel *mix* di finanziamento, abbia una sufficiente quota di accumulazione reale delle risorse. Il «giusto dosaggio» delle modalità di finanziamento sarà un aspetto cruciale delle politiche sanitarie e *welfariste* dei prossimi anni.

Fp e Ncs, Luglio 2010

INDICE

Sommario	4
Introduzione	7
Le dinamiche storiche di lungo periodo	7
Le proiezioni Awg-Ecofin	13
Le proiezioni Ocse	15
Dinamica potenziale o dinamica già corretta dalla <i>policy</i>?	17
Le proiezioni di spesa e la «critica di Lucas»	18
Dalle proiezioni di spesa sanitaria all'equilibrio socio-economico	22
Le ragioni del <i>multi-pillar</i> in sanità	24
Conclusioni	28

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. Questo scritto vuole portare un contributo di riflessione che, a partire dalla dinamica che storicamente la spesa ha fatto registrare (paragrafo 1.), si interroghi sulla dinamica futura (secondo paragrafi 2., 3., 4. e 5.) e sulle soluzioni più adatte a governarla (paragrafo 6., paragrafo 7. e paragrafo conclusivo).

1. Le dinamiche storiche di lungo periodo

Prima di guardare al futuro prossimo e remoto, è sempre bene guardare a quel che è già accaduto. I dati Ocse, *Health data - 2009*¹ 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 sono troppo rari e preziosi per essere dimenticati del tutto, persino nel loro ordine di grandezza. Le serie più recenti possono essere integrate dalle 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 paesi Ue-15 e negli Usa, l'incidenza della spesa sanitaria complessiva sul Pil si è almeno raddoppiata dal 1960 ad oggi. Per la Spagna il multiplo è addirittura per cinque/sei. 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 oltre il 70%, 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 volte².

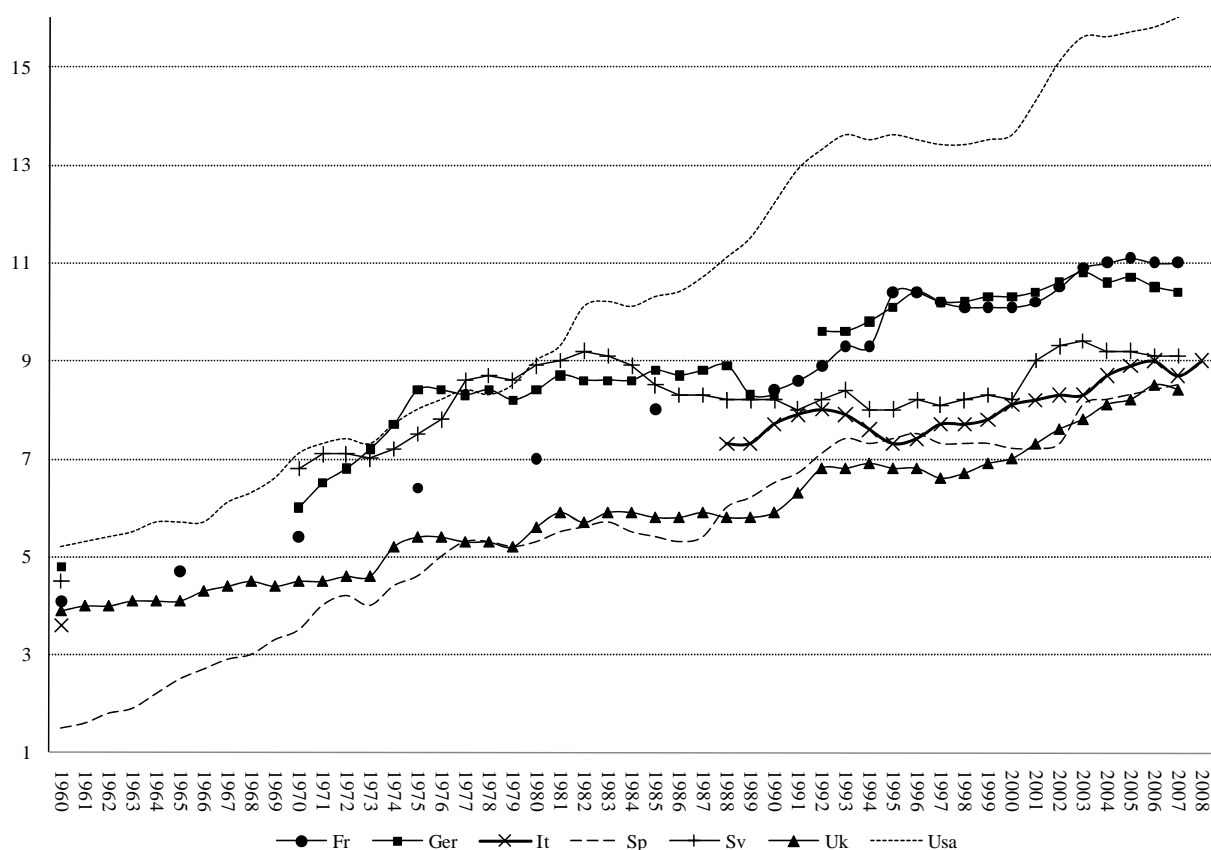
Una crescita così intensa si è accompagnata ad un altrettanto sorprendente aumento della vita attesa alla nascita e della vita attesa a 65 anni. La prima ha guadagnato circa un anno ogni quattro/cinque; la seconda circa un anno ogni dieci per gli uomini, e un anno ogni otto/nove per le donne.

¹ Ocse, *Health data, database online*, varie edizioni sino a quella del 2009.

² Non si riscontra, nei dati, un'interazione endogena inversa tra il *coverage* della spesa pubblica e la dinamica della spesa (pubblica e complessiva). Un esempio significativo arriva dal confronto Usa-Italia. Gli Usa dedicano, in termini di Pil, più risorse pubbliche dell'Italia alla sanità, anche se nel complesso mostrano un *coverage* inferiore perché la componente privata è molto più consistente che in Italia: se fosse vera l'interazione endogena dal *coverage* alla spesa, tale interazione dovrebbe emergere anche tra un'alta percentuale di spesa pubblica in termini di Pil e la dinamica della stessa spesa. Più interessanti, appaiono, invece altre endogenità attivabili nel *welfare system*, e in particolare quella che va dalla diversificazione degli strumenti di *welfare* (per ottenere la quale il finanziamento multi-pillastro delle pensioni e della sanità va visto come strumentale) al benessere psicofisico durante tutte le fasi della vita.

L'allungamento della vita è sicuramente un risultato dell'aumento delle condizioni di salute sostenuto 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 il contro-effetto della più elevata salute sui risparmi di spesa. Le serie storiche si sono mantenute in costante ed ininterrotta co-evoluzione crescente. E questo è avvenuto nonostante la generale riduzione delle probabilità di morte a tutte le età³.

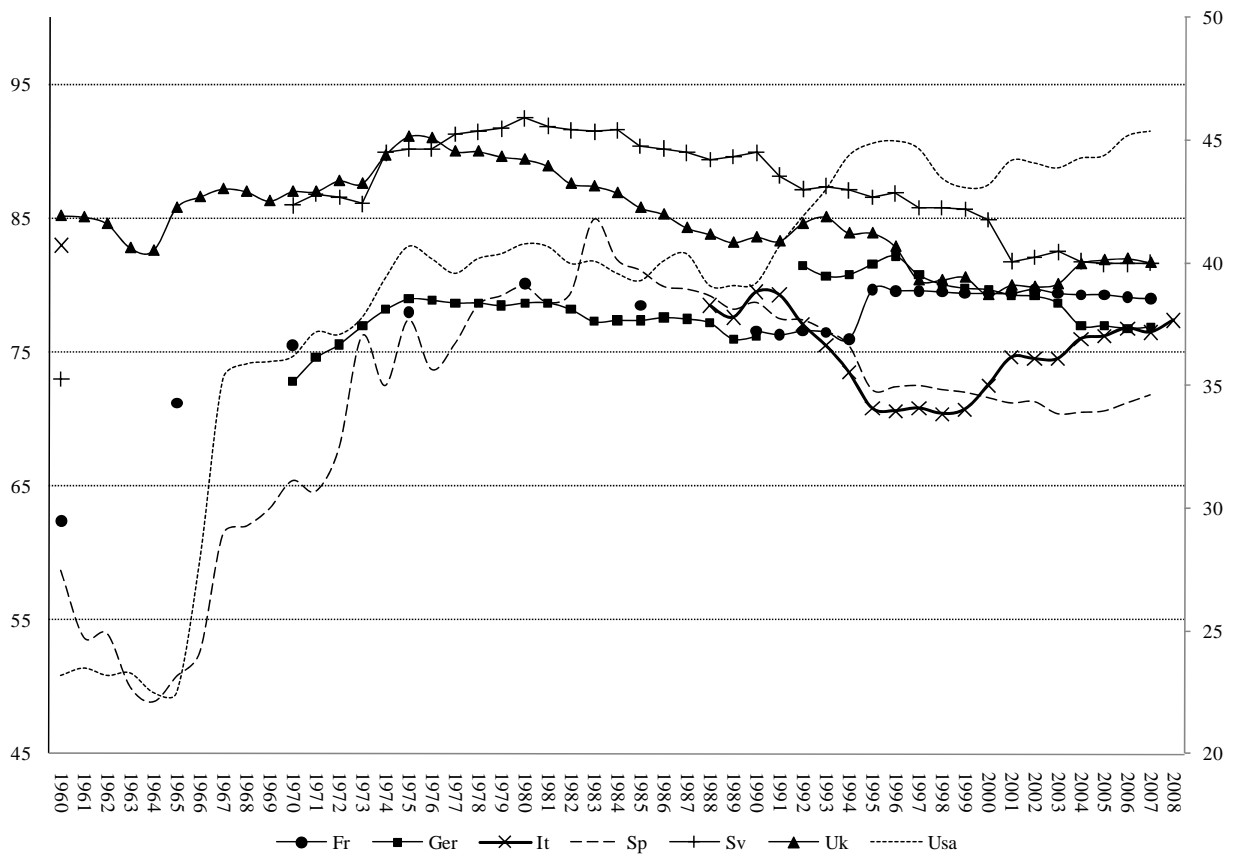
Figura 1 - Andamento dell'incidenza della spesa sanitaria complessiva (pubblica e privata) sul Pil



fonte: elaborazione degli autori su Ocse

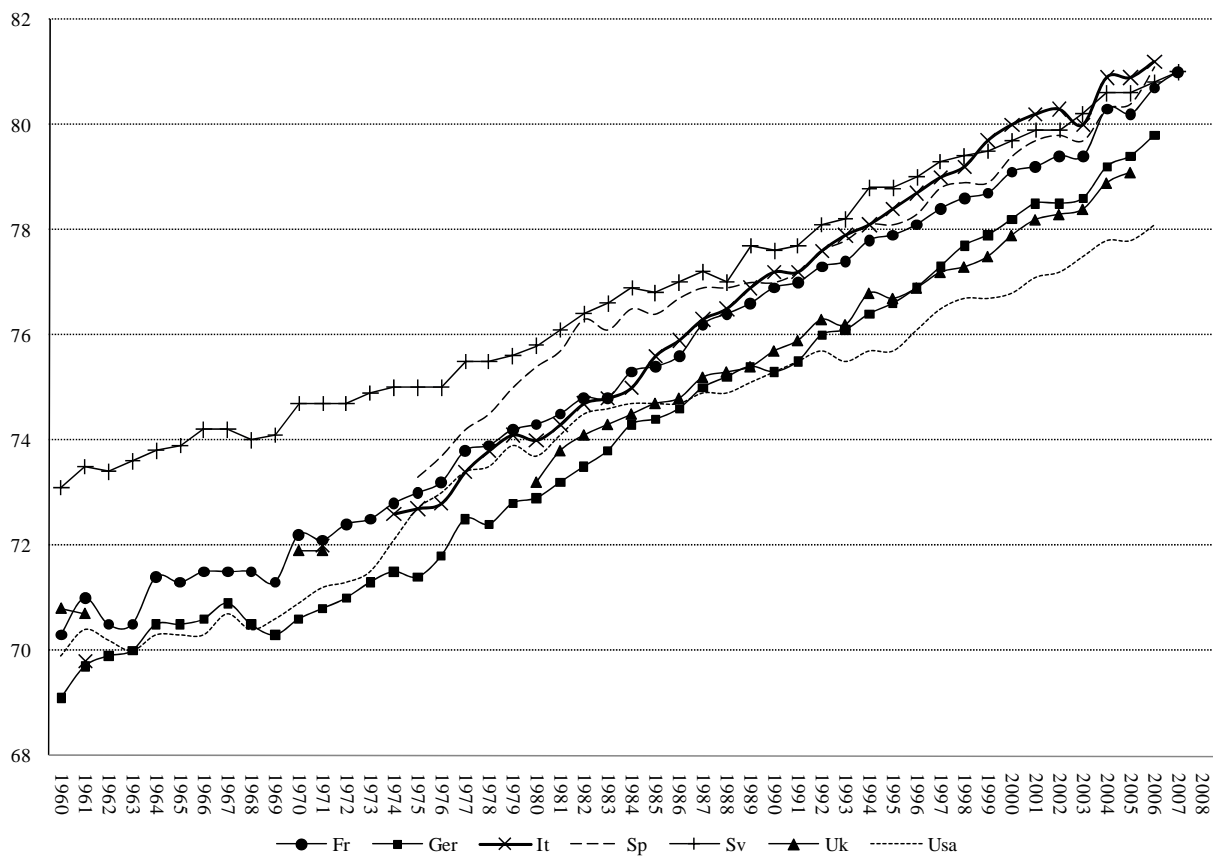
³ Una quota significativa della spesa sanitaria si colloca nell'anno del decesso (i cosiddetti *death related costs*). In Italia, per fare un esempio, il numero totale di anni di vita persi (altro indicatore di stato di salute di una comunità utilizzato dall'Ocse) si è ridotto del 74% dal 1960, *cfr.* <http://www.ecosante.org/index2.php?base=OCDE&langh=ENG&lang=ENG&sessionid=> (ultima visita: Giungo 2010).

Figura 2 - Andamento del coverage pubblico in percentuale della spesa sanitaria complessiva



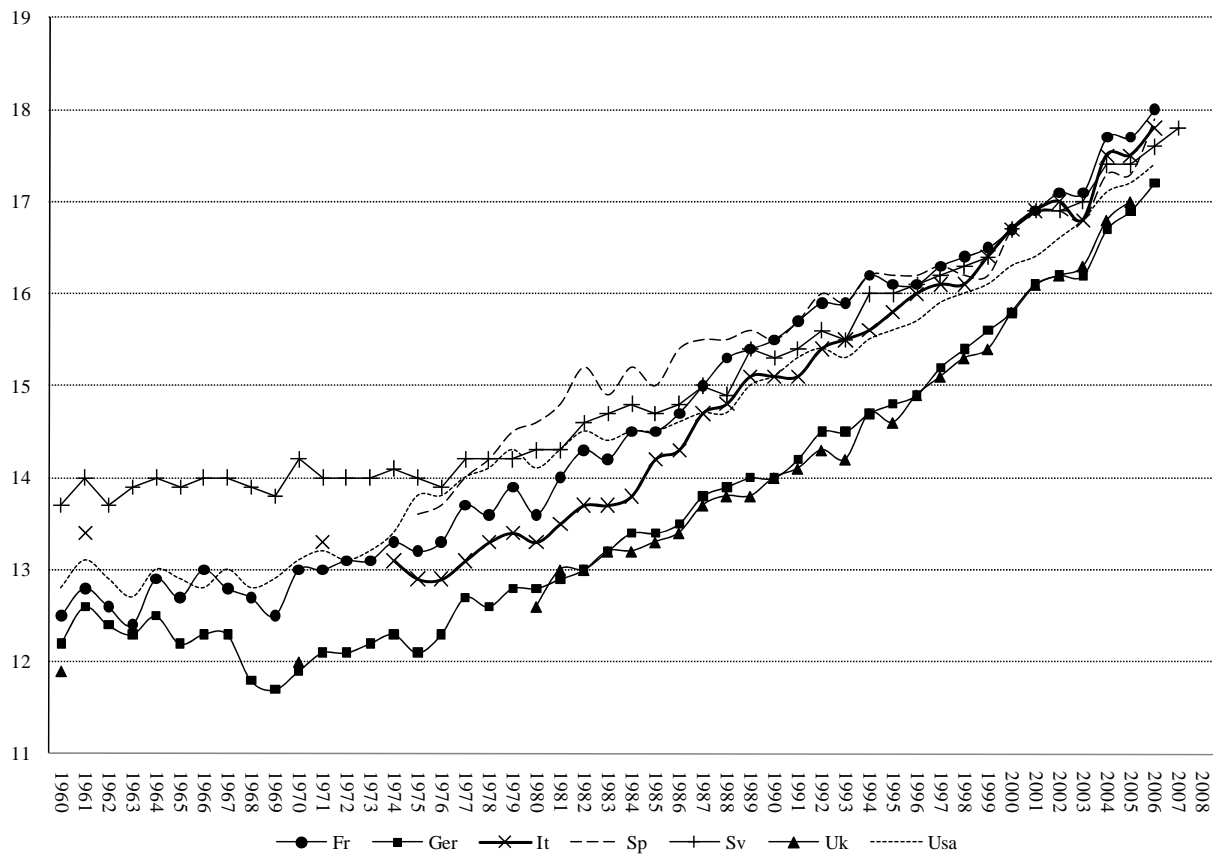
fonte: elaborazione degli autori su Ocse

Figura 3 - Progressione della vita attesa alla nascita



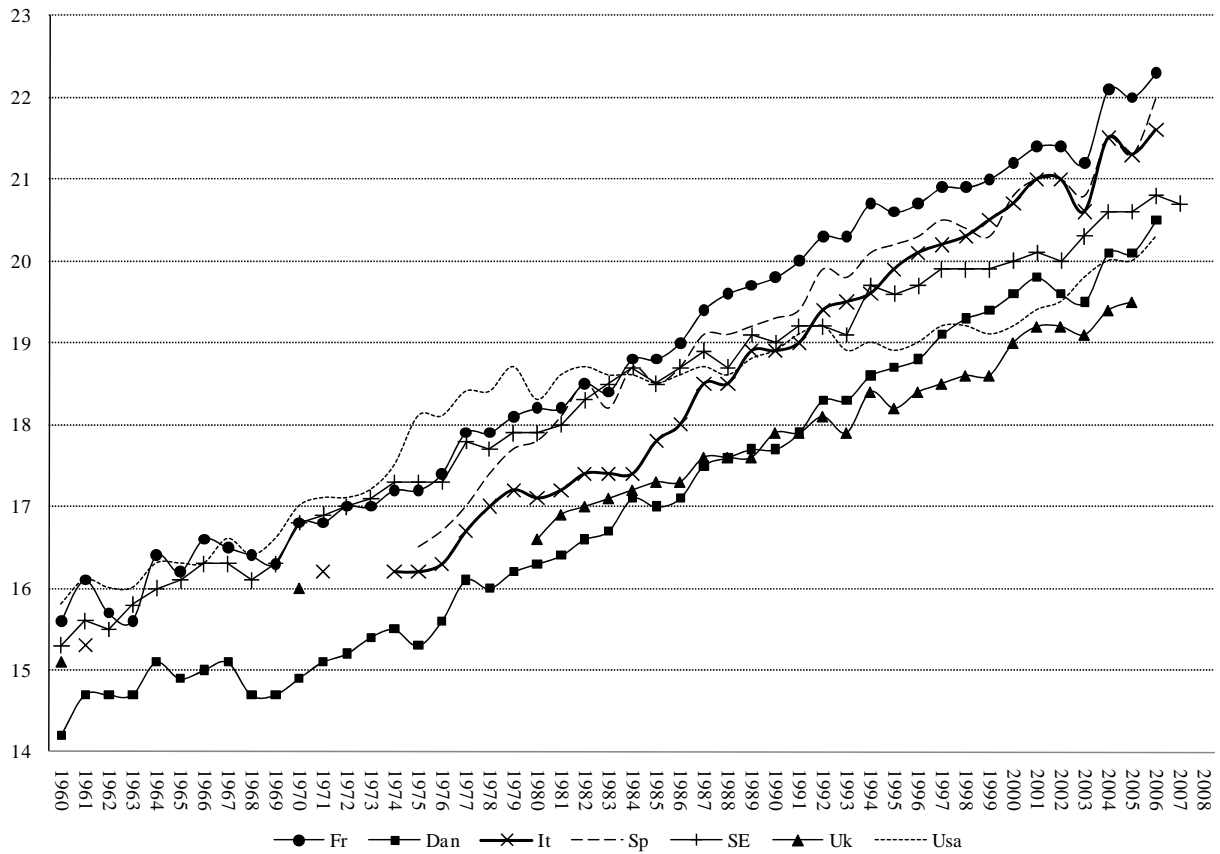
fonte: elaborazione degli autori su Ocse

Figura 4 - Progressione della vita attesa a 65 anni per gli uomini



fonte: elaborazione degli autori su Ocse

Figura 5 - Progressione della vita attesa a 65 anni per le donne



fonte: elaborazione degli autori su Ocse

2. Le proiezioni Awg-Ecofin⁴

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 paesi, 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 p.p. in Italia. Sul fronte della spesa di lungodegenza (*long-term* o *ltc*), gli incrementi di incidenza sono di 1,2 p.p. per l'Ue-15, di 1,1 per l'Ue-27, e di 1,3 per l'Italia.

Tavola 1 - Proiezioni di spesa Awg-Ecofin: scenario centrale e scenario «tecnologico»

Spesa sanitaria pubblica - % Pil

	scenario centrale (o Awg-reference scenario)				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 degli autori su 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 ad 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*.

⁴ Ci si riferisce ai due lavori Awg-Ecofin, *The impact of ageing populations on public expenditure*, 2006, e Awg-Ecofin, *Ageing Report - 2009 edition*, 2009. Per una loro dettagliata disamina, cfr. F. Pammolli e N.C. Salerno, *Spesa sanitaria: quali ipotesi per quali proiezioni? Ecofin e Ocse a confronto*, Nota Cerm, n. 1-2009, liberamente disponibile sul sito www.cermlab.it. 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.

Quanto alla 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»⁵), tranne che per le prestazioni *cash* dove il riferimento è il Pil *pro-capite*. L'elasticità delle prestazioni è, in 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 Paese⁶.

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'unità⁷. 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, a 6,3 p.p. in Ue-25, e a 5,3 p.p. in Italia. Più o meno dovunque, la spesa quasi raddoppia in termini di Pil.

⁵ Questo effetto consiste nel fatto che, se i guadagni di produttività nello svolgimento di una attività sono lenti mentre le remunerazioni di coloro che la svolgono seguono la dinamica reale media della massa retributiva nel sistema economico, il costo unitario di quella stessa attività cresce nel tempo. Sinora, l'«effetto Baumol» ha coinvolto numerose prestazioni sanitarie e socio-sanitarie, in primo luogo quelle di assistenza continuata alla persona inabile/disabile.

⁶ Cfr. F. Pammolli e N.C. Salerno, *La sanità in Italia*, 2008, ed. Arel – il Mulino; in particolare il capitolo dedicato all'inquadramento internazionale.

⁷ In particolare, l'elasticità al Pil è pari a 0,7, ma poi al tasso di crescita così ottenuto si aggiungono 2 punti percentuali per tener conto degli effetti, sia lato domanda che lato offerta, degli avanzamenti della scienza e della tecnologia (di qui il nome di «tecnologico»).

3. Le proiezioni Ocse⁸

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.

Tavola 2 - Proiezioni di spesa Ocse: scenario a più intensa crescita

	Spesa sanitaria pubblica - % Pil	
	<i>scenario a più intensa crescita</i>	
	spesa acuta e per lungodegenza	
	2005	Δ 2005-2050
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 degli autori su Ecofin

Per la voce *acute*, il tasso di crescita della spesa *pro-capite* è pari a quello del Pil *pro-capite* moltiplicato 1,2 (elasticità maggiore dell'unità) 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 *pro-capite* 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 è ipotizzato convergere al 70 per cento in tutti i Paesi⁹; 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'altra in aumento *year-by-year* come conseguenza di dinamiche lato domanda e di aumenti dei costi connessi al progresso scientifico-tecnologico.

⁸ Ocse, *Projecting OECD health care and long term care expenditures: what are the main drivers?*, 2006. Per una disamina dettagliata, cfr. F. Pammolli e N.C. Salerno, *Spesa sanitaria: quali ipotesi per quali proiezioni? Ecofin e Ocse a confronto*, cit..

⁹ 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 per cento, con quello delle donne poco al di sopra del 50 per cento.

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 raddoppio¹⁰.

¹⁰ 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, non doppio ma addirittura multiplo.

4. Dinamica potenziale o dinamica già corretta dalla *policy*?

Più recentemente, è sopraggiunto anche un lavoro del Fondo Monetario Internazionale che, nel *Fiscal Monitor* di Maggio 2010, presenta proiezioni in cui la spesa sanitaria pubblica raddoppia la sua incidenza sul Pil a cinquant'anni.

Come comporre assieme l'evidenza storica con esercizi di proiezione che presentano scenari anche molto diversi tra loro? Cruciale diviene l'esame delle ipotesi alla base delle proiezioni.

Se la differenza tra Ecofin e Ocse risiede soprattutto nel modo in cui le proiezioni raffigurano il legame tra la dinamica della spesa e lo sviluppo economico-sociale¹¹, 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 *coverage* e qualità/innovatività delle prestazioni da un lato e sostenibilità finanziaria dall'altro, 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 Ocse¹², 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 singoli Paesi.

Quando applicati al tema specifico delle proiezioni di lungo termine della spesa sanitaria, la modellistica econometrica si scontra spesso con critiche «à la Lucas»¹³.

¹¹ 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.

¹² Basata su estrapolazione di profili di spesa *pro-capite* per fasce di età-sesso.

¹³ 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. Su questo punto, ulteriori argomentazioni sono rinvenibili in N.C. Salerno, *La critica di Lucas e le proiezioni della spesa sanitaria*, in www.crusoe.it, 7 Aprile 2010.

5. Le proiezioni di spesa e la «critica di Lucas»

Negli anni Settanta, l'economista Robert Emerson Lucas sviluppò il concetto che poi da lui prese il nome divenendo la «critica di Lucas». In *Econometric policy evaluation: a critique* (1975), Lucas mise per la prima volta in evidenza la rilevanza delle reazioni degli agenti economici ai comportamenti e alle scelte del *policy maker*. Continuare ad ignorare queste reazioni avrebbe, inevitabilmente, causato errori sistematici nelle valutazioni di impatto delle opzioni di politica economica.

L'intuizione è alla base del filone delle aspettative razionali e ultrarazionali, che hanno trovato applicazione in tanta modellistica di politica fiscale e monetaria. Questa stessa intuizione ha anche "mandato in soffitta" tutta una econometria di prima maniera che, trascurando completamente l'endogenità dei parametri di reazione alla *policy*, aveva immaginato di poterli stimare una volta per tutte, per poi applicarli a vari scenari in esercizi di statica comparata.

Ma che cosa ha a che fare la critica di Lucas con la sanità? Resta un punto di riferimento da non perdere nel dibattito, in corso, su quali siano le metodologie più corrette per proiettare gli andamenti di spesa a medio-lungo. Vediamo perché.

Come si è visto, dai modelli di proiezione ad oggi disponibili (nazionali ed internazionali) deriva un *range* per la variazione di incidenza sul Pil della spesa sanitaria pubblica tanto ampio da rimanere sostanzialmente indeterminato: nei prossimi cinquant'anni, alcune fonti come Ocse ed Ecofin (lo scenario «tecnologico») segnalano il raddoppio o più che raddoppio, per giunta senza inversione di tendenza, per tutti i Paesi ad economie e *welfare* sviluppati; altre fonti, come il capitolo di sostenibilità di lungo termine del Programma di Stabilità (che mutua con alcune modifiche lo scenario *Awg-reference* di Ecofin), riportano una crescita meno sostenuta ma pur sempre significativa e senza inversione di tendenza (per l'Italia +2 p.p. circa). In particolare per l'Italia, esistono altre fonti ancora (centri studi pubblici e privati) che vedono l'incidenza arrivare, dopo una fase di forte crescita, a stabilizzarsi tra il 2030 e il 2035, per poi cominciare a ridursi (secondo un profilo a gobba che richiama, a livello qualitativo, quello dell'incidenza della spesa pensionistica trainato dagli andamenti demografici).

Come orientarsi in questo labirinto di proiezioni? Il punto di partenza è la domanda sollevata al paragrafo precedente: di che cosa si è alla ricerca? della dinamica potenziale della spesa, oppure di quella che già sconta interventi di contenimento e stabilizzazione? Se le proiezioni già scontano non meglio precisate azioni di *policy*, da un lato l'urgenza dell'*agenda* ne risulta falsata e, dall'altro, si dà per scontata la praticabilità di quelle stesse azioni, cosa che, invece, potrebbe essere tutta da verificare, sia sul piano tecnico che su quello politico.

In particolare, la condizione di «a parità di normativa corrente» alla base di tutte le proiezioni appare ineludibile (al più si possono includere gli effetti di riforme già approvate e in corso di attuazione); ma, all'interno di questa cornice, non si può non considerare il potenziale di spesa che potrebbe sorgere per quegli stessi ambiti di prestazione oggi coperti (in Italia, i livelli essenziali sanitari e socio-sanitari), per i quali potrebbe aumentare la platea di beneficiari e, soprattutto, potrebbero sopraggiungere nuove tecnologie/tecniche di intervento.

Esempi significativi sono:

- l'ampliarsi dei destinatari delle prestazioni di assistenza continuata per non autosufficienza, di pari passo con l'invecchiamento della popolazione e con

- l'evolversi di grandezze economiche come il tasso di attività e di occupazione femminile (in Italia è un livello essenziale, ancorché lungi dalla realizzazione, *cfr.* Legge n. 328 del 2000, Dpcm del 14 Febbraio 2001, Dpcm del 29 Novembre 2001);
- l'elasticità della domanda al Pil che, come si è già detto, dal Dopoguerra si è attestata a livelli significativamente superiori all'unità su orizzonti temporali medio-lunghi e in tutti i Paesi ad economia e *welfare* sviluppati;
 - l'ampliarsi del *basket* delle prestazioni rientranti negli ambiti di intervento oggi coperti, come conseguenza dell'avanzamento scientifico-tecnologico (si pensi alle terapie biotecnologie e personalizzate, alle cure contro il cancro, la sclerosi multipla, etc.).

Nei modelli di proiezione basati sui profili di spesa per sesso-età, come quello Ocse e quello Ecofin, il potenziale di crescita è affidato al ricorso, nella scelta dei parametri con cui fare evolvere quegli stessi profili, di un tasso di crescita della spesa con un significativo *spread* rispetto al tasso di crescita del Pil, e di tassi di *formal care* per l'assistenza ai non autosufficienti coerenti con gli scenari evolutivi dell'attività e dell'occupazione femminile. A queste ipotesi si aggiunge quella che i tassi di inabilità/invalidità abbiano una duplice dinamica: in diminuzione su orizzonti di medio breve, per tener conto del continuo miglioramento della salute; ma in aumento su orizzonti di medio-lungo, per tener conto dell'emersione di domanda inespresa, di pari passo sia con le possibilità aperte dalle nuove tecniche/tecnologie, sia dell'evoluzione stessa del concetto di *activities of daily living*.

In questo senso, appare riduttivo definire meramente contabile questa famiglia di modelli, perché essi necessitano di ipotesi sulla dinamica, e la calibrazione dei loro parametri avviene con la finalità di soddisfare delle relazioni di coerenza interna macro e microeconomica. Ovviamente, il lungo orizzonte temporale cui sono applicati e la loro struttura semplice ci obbligano ad utilizzarli consapevoli che possono soltanto fornire indicazioni sull'ordine di grandezza delle pressioni che giungeranno dalla spesa sanitaria sui conti pubblici, e non certo previsioni puntuali. E a questo proposito, non si può non sottolineare come, se i più forti *driver* di spesa che Ocse ed Ecofin distribuiscono in scenari diversi, fossero compresenti e interagissero sotto un unico scenario (situazione che non è possibile escludere *a priori*), l'ordine di grandezza delle proiezioni di spesa diverrebbe di gran lunga superiore al raddoppio dell'incidenza sul Pil.

Ma torniamo alla critica di Lucas. La modellistica che più facilmente riporta la stabilizzazione di lungo periodo della spesa e, addirittura, la sua riduzione in termini di Pil dopo una fase di picco, è modellistica econometrica. Si tratta, a grandi linee, di modelli che stimano i coefficienti di impatto di *set* di variabili esplicative sull'arco degli ultimi 30-40 anni, per poi utilizzare questi coefficienti per proiettare l'impatto delle stesse variabili esplicative nel futuro remoto. È sicuramente una modellistica più sofisticata e tecnicamente più evoluta rispetto alla famiglia dei modelli calibrati ma, nello specifico delle proiezioni di lungo termine, questa maggior raffinatezza può essere un'arma a doppio taglio. La stima dei coefficienti di impatto, infatti, rimane inevitabilmente «sporcata» dalla *policy*, soprattutto nei casi di Paesi, come l'Italia, che hanno attraversato stagioni di esplicita forte compressione di spesa sanitaria per ragioni di bilancio. Anche l'inserimento di *dummy* che segnalino, anno per anno, l'adozione di interventi normativi settoriali difficilmente può essere una soluzione, perché non si riesce a tener conto del controvalore delle misure e della domanda troncata (oltretutto da considerarsi sia nella dimensione della quantità che della qualità).

Se Lucas ha messo in guardia dall'utilizzare coefficienti di impatto delle *policy* che ignorassero le reazioni degli agenti economici, per la stessa ragione è necessario prestare

attenzione quando si utilizzano coefficienti di impatto di variabili esplicative della spesa sanitaria stimate sui dati storici. Questi coefficienti, infatti, danno per scontato che le azioni di *policy* condotte nel passato siano replicabili *tout court* nel lungo periodo, senza tener conto dei mutamenti nella demografia, nella società, nel mercato del lavoro, nella disponibilità di terapie e trattamenti, nella posizione relativa tra Paesi per quanto riguarda l'offerta di prestazioni sanitarie e socio-sanitarie, etc.. Paradossalmente, proprio questa ipotesi, che almeno a prima vista sembrerebbe di neutralità («senza modifiche normative» o condizioni simili), realizza, invece, uno scenario di proiezione niente affatto neutrale. Ben lungi dall'essere una condizione di *ceteris paribus*, si impone così una prospettiva in cui le scelte più o meno recenti di *policy* potranno essere replicate nel lungo periodo, continuando a produrre gli effetti già riscontrati, senza scontrarsi con ostacoli e vincoli nuovi.

Su questo punto specifico, l'errore sembrerebbe risiedere nella interpretazione della condizione di «a parità di ...» sotto la quale svolgere le proiezioni. Un conto è mantenere i riferimenti correnti al perimetro delle prestazioni e alle regole di accesso alle stesse; questi riferimenti sono indispensabili per la stessa caratterizzazione del sistema sanitario di cui si vuol proiettare l'andamento. Altro conto, invece, è introdurre ipotesi che vanno ad influenzare direttamente le *performance* finanziarie (*in primis*, l'ipotesi sull'elasticità al Pil/reddito). Esempio ne sia il fatto che in Italia la forte compressione di spesa sanitaria durante la prima metà degli anni Novanta si è svolta senza nessuna modifica né del perimetro delle prestazioni Ssn né delle regole di accesso: l'offerta Ssn è rimasta formalmente onnicomprensiva e con prestazioni erogate secondo universalismo assoluto, ma nella realtà le file d'attesa sono aumentate, la qualità è caduta, le Asl e le Ao hanno cominciato ad accumulare posizioni debitorie esplicite e nascoste, etc.. Utilizzare coefficienti stimati su dati storici appartenenti anche a stagioni in cui le scelte finanziarie hanno condizionato in maniera pesante il sistema sanitario senza che emergesse il disegno di nuovo equilibrio strutturale significa, di fatto, ipotizzare che in futuro possano essere prolungate *sine die* le tensioni, economiche e sociali, che si accumulano quando tra offerta teorica di prestazioni e risorse disponibili per realizzarle la forbice è ampia e in divaricazione.

Se davvero si vuole che le proiezioni di medio-lungo termine della spesa sanitaria (come di altri capitoli) possano passare indicazioni utili al *policy maker*, è necessario interrogarsi su come queste effettivamente riescano a cogliere l'evoluzione potenziale e a descrivere le tensioni attese per i vincoli di bilancio e per le compatibilità e le proporzioni economiche e sociali (a livello micro e macro). In questa prospettiva, la modellistica econometrica è esposta al rischio di un sistematico *backward-looking*, la cui correzione è difficile perché, soprattutto in sanità e su orizzonti pluridecennali, è difficile l'inserimento di variabili di controllo per tener conto della domanda inespressa/troncata e della qualità, o delle tensioni che possono nascere dai differenziali di assistenza sanitaria tra Regioni/Paesi, o della spesa che può nascondersi nell'accumulazione di posizioni debitorie sommerse (di cui spesso è persino difficile trovare traccia nella contabilità).

Un modello econometrico con un ottimo *fitting* della spesa storica italiana dagli inizi degli anni Novanta ad oggi potrebbe, paradossalmente, proprio per questa «qualità» essere inadatto in sede di proiezione, inadatto a dare indicazioni sulla dinamica prospettica di lungo periodo. A meno che non sia corredato da variabili di controllo che diano sufficiente garanzia di isolare i coefficienti di impatto delle variabili esplicative dalle interferenze di *policy*, un modello di questo genere implicitamente ribalterebbe nel futuro remoto le scelte di *policy* effettuate nel passato, dandone per scontato sia la praticabilità che la replicabilità degli effetti (quella critica *à la* Lucas citata all'inizio). Ma il punto è proprio questo: in che misura siamo in grado di controllare per fenomeni complessi e multidimensionali come quelli prima citati?

Senza voler eccedere nel discutere di difetti e pregi di questa e quella modellistica alle quali si sta chiedendo, tra l'altro, qualcosa di estremamente complesso in sé (descrivere gli andamenti a cinquant'anni di voci di spesa importanti e politicamente e socialmente sensibili come quella sanitaria), da queste considerazioni deriva una chiara sollecitazione a non sottovalutare l'approccio calibrato e, soprattutto, a non nutrire sospetti nei suoi confronti per il fatto che sono i modelli calibrati a riportare gli ordini di grandezza più macroscopici sulla dinamica della spesa sanitaria rispetto al Pil: come si è visto (Ecofin, Ocse, Fmi), per tutti i Paesi ad economia e *welfare* sviluppati il raddoppio o il più che raddoppio al 2050-2060 senza accenni di inversione di tendenza. Se si guarda al panorama internazionale (paragrafo 1.), negli ultimi cinquant'anni l'incidenza della spesa sanitaria pubblica sul Pil è aumentata di multipli sino a sette/otto volte. Non può apparire così strano che nei prossimi 50 anni la dinamica di spesa vada almeno verso un raddoppio dell'incidenza sul Pil.

6. Dalle proiezioni di spesa sanitaria all'equilibrio socio-economico

Ricapitolando, le proiezioni Ecofin-Ocse-Fmi 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 né più né meno 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 per cento circa al 50 per cento se si fa riferimento allo scenario centrale di Awg-Ecofin, oppure significativamente al di sotto del 50 per cento 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 per cento del Pil nel 2050, sarebbe necessario un profilo di avanzi di bilancio continuamente crescente sino a superare il 10 per cento del Pil, con conseguente totale «ingessamento» della politica annuale di bilancio¹⁴.

Oggi, in Italia, ogni persona attiva finanzia la spesa pubblica per pensioni e sanità dedicata agli ultrasessantacinquenni con risorse equivalenti ad oltre il 31 per cento del Pil *pro-capite*. Al 2050, questo valore sarà del 42,2 per cento se si considerano le proiezioni per pensioni e sanità dello scenario centrale di Awg-Ecofin; e del 48,6 per cento se le proiezioni della sanitaria sono quelle dello scenario a più intensa crescita dell'Ocse¹⁵. L'incremento dei *ratios* al 2050 deriverà soprattutto dalla sanità, come testimonia il dato che emerge se le proiezioni sanitarie Awg-Ecofin (scenario *Awg-reference*) sono sostituite con quelle Ocse. Nella media Ue-15, i tre valori sono rispettivamente 25,2, 38,3 e 45%¹⁶.

¹⁴ Cfr. F. Pammolli e N.C. Salerno, *La sanità in Italia – Federalismo, regolazione dei mercati, sostenibilità delle finanze pubbliche*, Bologna, il Mulino – Arel, 2008. Gli avanzi sono quelli complessivi di bilancio. La previsione di spesa è essenziale per verificare quale avanzo di bilancio è necessario per ottenere un prefissato *target* di debito pubblico (in % Pil). Se una voce di spesa aumenta in percentuale del Pil, di pari passo si dovranno correggere o le entrate (in aumento) o le altre spese (in diminuzione), a parità di *target* che si desidera raggiungere. La dinamica delle entrate e delle uscite definisce la dinamica del saldo di bilancio. Nell'esercizio condotto, le ipotesi macroeconomiche (evoluzione del Pil, tasso di interesse sul debito, evoluzione delle altre voci di spesa collegate e non collegate all'età) sono mutuare dal Programma di Stabilità dell'Italia del Ministero dell'Economia e delle Finanze.

¹⁵ L'esercizio è descritto in dettaglio in F. Pammolli e N.C. Salerno, *Spesa sanitaria: quali ipotesi per quali proiezioni? Ecofin e Ocse a confronto*, cit.. La spesa per ultra 65enni è rapportata agli attivi (o agli occupati in altro scenario pure sviluppato), mentre come grandezza di scala si utilizza il Pil *pro-capite* su tutti i residenti. Con il rapporto tra, da un lato, la spesa sopportata da ogni attivo a titolo di ripartizione (*pay-as-you-go*) e, dall'altro, il Pil per cittadino residente, si intende cogliere quale percentuale del Pil di cui mediamente dispone un cittadino attivo/occupato andrà a servizio della ripartizione. Un ordine di grandezza che aiuta ad interrogarsi sulla sostenibilità di un meccanismo di finanziamento affidato quasi esclusivamente alla ripartizione e dei flussi di redistribuzione che esso formalmente attiva tra individui e tra generazioni. Sono i flussi formali di redistribuzione che rilevano ai fini della valutazione degli effetti distorsivi che possono diffondersi sul mercato del lavoro (offerta e domanda) e sulle scelte di investimento. Ipotesi che riguardino la condivisione interna ai nuclei familiari delle prestazioni sanitarie e assistenziali (i flussi di redistribuzione informali) non possono essere considerate, perché trascurano le modalità con cui le risorse si formano nel sistema economico e sono allocate. Nel citato lavoro di Pammolli e Salerno i calcoli sono descritti integralmente.

¹⁶ Molto interessante, su questo punto, un confronto con la contabilità intergenerazionale descritta per gli Usa in L.J. Kotlikoff, *The Healthcare Fix - Universal Insurance for All Americans*, 2007, 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

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 per cento del Pil *pro-capite*. Al 2050, questo valore sarà del 62,3 per cento se si considerano le proiezioni per pensioni e sanità dello scenario centrale di Awg-Ecofin; e del 71,5 per cento 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 per cento. I sistemi di finanziamento *monopillar pay-as-you-go* pubblici, che contraddistinguono quasi tutti i Paesi europei, reggeranno a cambiamenti di questa entità?¹⁷

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 per cento). 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.

anni. Per inciso, il volume contiene numerosi spunti per valutare i vincoli finanziari con cui dovrà misurarsi la sanità pubblica Usa dopo che la recente riforma «Obama» ne ha esteso il *coverage*.

¹⁷ Sono coinvolti tutti, sia sistemi à la *Bismarck* che sistemi à la *Beveridge*. Variazioni significative (tra i 10 e i 20 punti percentuali) 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, e sicuramente meglio dell'Italia, con elevati cunei contributivi sui redditi da lavoro.

7. Le ragioni del *multi-pillar* in sanità

Dopo la prima domanda, «Dinamica potenziale o dinamica già corretta dalla *policy*?», a cui si sono proposte argomentazioni in risposta nei paragrafi 4. e 5., il paragrafo 6. ne ha aggiunto una seconda: «Reggerà il finanziamento a ripartizione?». Se non si inseriranno quote ad accumulazione reale nella predisposizione delle risorse con cui fronteggiare i fabbisogni futuri, la pressione sui redditi (da lavoro e non) anno per anno prodotti diventerà esorbitante, con probabili effetti distorsivi e depressivi sull'attività economica, che si abbineranno ad effetti di razionamento delle prestazioni dovuti, questi ultimi, alla difficoltà delle risorse disponibili a tenere il passo del fabbisogno potenziale. Ma sarà solo il *pay-as-you-go* dei sistemi pubblici ad essere posto sotto pressione?

La ripartizione nei sistemi pubblici universalistici

I sistemi sanitari pubblici universalistici à la *Beveridge* (come l'italiano Ssn) si basano sul finanziamento a ripartizione o *pay-as-you-go* (tramite la fiscalità generale). Già ora non vale più il cosiddetto «teorema di Aaron» che, nelle fasi demografiche di espansione, identifica la ripartizione come il meccanismo più efficiente sia per spostare risorse nel tempo (le pensioni) sia per finanziare prestazioni sanitarie per l'intera cittadinanza. Negli anni '60 e '70, quando quel teorema trovava riscontro nella demografia, con un contributo relativamente contenuto e poco distorsivo a carico di tutti i redditi si potevano reperire risorse adeguate. In prospettiva, quel contributo dovrà aumentare sino a toccare soglie insostenibili, con effetti depressivi sul mercato del lavoro, sugli investimenti e sulla produttività. Sistemi a prevalente finanziamento a ripartizione dovranno necessariamente ristrutturarsi, e la soluzione non potrà limitarsi a restringere il perimetro delle prestazioni fornite oppure ad aumentare la compartecipazione dei cittadini ai costi (per le proporzioni in cui si manifesteranno le pressioni della spesa, sarebbero due non-soluzioni).

La mutualità nei sistemi pubblici categoriali

La criticità appena descritta abbraccia anche i sistemi pubblici à la *Bismarck*, basati sulla mutualità obbligatoria interna a comparti, settori, categorie. Per tenere il passo della spesa, i contributi a carico degli iscritti e dei loro datori di lavoro dovranno necessariamente aumentare riproducendo quella stessa pressione sugli occupati che, nei sistemi universalistici, passa per la via fiscale. C'è da dire, inoltre, che le differenze tra sistemi *beveridgiani* e *bismarckiani*, nette sino a tutta la metà del Novecento, si sono poi ridotte notevolmente quando i *bismarckiani* hanno mano mano esteso, sempre con logica categoriale, la copertura della platea degli occupati. L'affidarsi alla fiscalità generale, piuttosto che a contribuzioni specifiche levate per comparti/settori/categorie, non crea di per sé diverse capacità di fronteggiare una dinamica di spesa così intensa, soprattutto quando entrambi i sistemi si prefiggono obiettivi di ampia copertura sia in termini di beneficiari che di prestazioni. In entrambi i casi, la necessità di reperire risorse crescenti si trasforma in una pressione sempre più alta sui redditi che il sistema economico genera anno per anno.

La mutualità privata

In effetti, la mutualità categoriale obbligatoria nei sistemi *bismarckiani* altro non fa che applicare il finanziamento a ripartizione, anno per anno, non sul complesso dei redditi, ma sui singoli sottogruppi di redditi individuati a seconda di comparto/settore/categoria. L'insostenibilità della ripartizione, quando utilizzata come criterio esclusivo di

finanziamento, si estende *tout court* anche ai sistemi *bismarckiani*. E per la stessa ragione arriva a coinvolgere anche le forme di mutualità privata (come le nostre società di mutuo soccorso), con cui gruppi di persone scelgono volontariamente di mettere in comune delle risorse per fare *risk sharing*. Di fronte a dinamiche di spesa così intense, le organizzazioni di mutualità privata saranno costrette o ad aumentare significativamente i contributi dei partecipanti (anche a titolo di compartecipazione ai costi), o a restringere il novero delle prestazioni, oppure a selezionare (*screaming*) in maniera più stringente i partecipanti all'ingresso (rimedio in contrasto con la definizione stessa di mutualità). Non è un caso che, scostandosi dallo spirito originario con cui sono sorte nell'Ottocento, molte società di mutuo soccorso oggi si rivolgano esclusivamente alle fasce dei quadri e dei dirigenti. Fintantoché la mutualità privata si affida, per il finanziamento, ad una ripartizione della spesa sui propri iscritti, emergono le stesse criticità evidenziate per i sistemi pubblici *beveridgiani* e *bismarckiani*, con una aggravante: quando la ripartizione si svolge su gruppi piccoli, essa rimane statisticamente più esposta sia all'assottigliarsi del numero degli aderenti in età attiva e contributiva, sia all'aumentare della frequenza e del costo medio delle richieste di prestazioni (la ripartizione delle spese trova le condizioni ideali di funzionamento sui grandi numeri). Se nei prossimi anni il *remix* di finanziamento della sanità dovesse avvenire con trasferimento di quote dal carico dei sistemi pubblici alla mutualità privata, si rimarrebbe, in realtà, sempre all'interno del criterio a ripartizione, senza superarne le difficoltà strutturali.

L'assicurazione privata

Non è immune da criticità anche la copertura assicurativa privata, cui spesso in Italia le organizzazioni di mutuo soccorso ricorrono svolgendo funzioni "passacarte". La logica assicurativa differisce da quella della mutualità per le modalità con cui contribuzioni e prestazioni attese sono poste in connessione tra di loro: la mutualità non ricerca un proporzionamento specifico tra contributi e prestazioni attese (normalmente, si erogano prestazioni sino ad esaurimento delle risorse anno per anno); al contrario, nella copertura assicurativa questa connessione è centrale, e i premi da corrispondere sono dimensionati a seconda delle caratteristiche del singolo (polizza individuale) o di quelle medie della platea (polizza collettiva), oltre che delle tipologie di prestazioni assicurate. A parte questa distinzione (che è comunque sostanziale dal punto di vista dei rapporti con gli iscritti), le assicurazioni gestiscono la loro esposizione finanziaria facendo *pooling* dei profili di contribuzione/rischio dei loro sottoscrittori (a fronte del pagamento dei premi da parte di tutti, non tutti incorreranno contemporaneamente negli eventi avversi), e mantenendo una base di riserve matematiche investite sui mercati. La proporzione tra *pooling* e risorse investite è cruciale. Quanto più la copertura assume la forma assicurativa pura (tutto o prevalentemente *pooling*), tanto più si ricade nelle stesse criticità del finanziamento a ripartizione dei sistemi pubblici, con premi destinati ad aumentare per tenere il passo della spesa sanitaria, assieme alle compartecipazioni e allo *screaming*. Il *pooling* assicurativo riproduce, in ambito privatistico e su platee più ristrette, il meccanismo della ripartizione nei sistemi pubblici, con l'aggravante, da un punto di vista della sostenibilità sociale, che gli aumenti dei premi possono variare a seconda delle caratteristiche sanitarie soggettive (la differenza rispetto alla mutualità volontaria). È qualcosa cui si sta già assistendo negli Stati Uniti (il Paese in cui la copertura assicurativa è più diffusa), dove i premi delle assicurazioni sanitarie hanno fatto registrare, negli ultimi anni, una dinamica a doppia cifra, inducendo l'Amministrazione "Obama" a parlare apertamente di insostenibilità del sistema (il 22 Febbraio 2010 è stato presentato l'*Health Insurance Industry Fair Competition Act*, H. R. n. 4626).

La convergenza dei modelli: ma verso dove? Verso multi-pillar con quote di accumulazione reale

Se sistemi così diversi, pubblici e privati, prestano il fianco agli scenari di crescita della spesa sanitaria, esiste una soluzione? Al di là degli sforzi di efficientamento (organizzazione dei sistemi sanitari nazionali, concorrenza tra *provider*, concorrenza nel mercato assicurativo, responsabilizzazione dei cittadini fruitori e dei medici prescrittori, etc.), il punto critico appare proprio quello del criterio di finanziamento. Se alla maggior spesa si tenterà di far fronte utilizzando quote crescenti dei redditi prodotti anno per anno dal sistema economico, l'effetto sarà quello di una crescente pressione fiscale e/o contributiva. All'interno dei sistemi pubblici obbligatori tale pressione, evitabile solo nella misura in cui si restringano le prestazioni, porterà con sé effetti distorsivi e depressivi. All'interno dei sistemi privati, è probabile che la pressione crescente si traduca anche in una rinuncia alla copertura, con conseguenze che potrebbero generare maggior spesa in futuro (per la minore attenzione alle cure e l'aggravarsi delle condizioni di salute) e pressioni su altri capitoli del *welfare system*. Fintantoché il finanziamento sarà ricercato esclusivamente/prevalentemente nella ripartizione (i sistemi pubblici), nella mutualità diretta tra gruppi e nel *pooling* assicurativo (i sistemi privati), o in pagamenti diretti dei cittadini, trasferimenti di quote delle prestazioni e della domanda dal pubblico al privato e viceversa non incideranno sulla capacità sistemica di governare le tendenze di spesa. Eppure, apparentemente è proprio questa la tendenza in atto: da una parte, gli Stati Uniti, con il programma "Obama", si stanno muovendo verso una estensione della copertura pubblica finanziata a ripartizione per sopperire all'insostenibilità economica e sociale delle assicurazioni sanitarie private; dall'altra parte, in Europa sta lentamente maturando il processo opposto, con le difficoltà dei sistemi pubblici *beveridgiani* e *bismarckiani* che richiamano l'esigenza di rafforzare la copertura complementare privata (in Italia, i recenti decreti "Turco" e "Sacconi" hanno riproposto il dibattito sullo sviluppo dei fondi sanitari integrativi del Ssn¹⁸). In questa convergenza transatlantica verso un sistema sanitario multi-pillastro maggiormente equilibrato nelle componenti pubbliche e private manca, per adesso, una caratteristica essenziale per ovviare alle criticità del finanziamento basato su ripartizione / mutualità / *pooling* assicurativo e sovraccarico sui redditi prodotti anno per anno: il finanziamento tramite i frutti di programmi di investimento di lungo termine sui mercati. Laddove i criteri di finanziamento oggi prevalenti utilizzano redditi già prodotti (intervengono *ex-post*, per dedicare alla finalità sanitaria quote di risorse già prodotte), i programmi di investimento entrano nei processi di generazione delle risorse e le preparano per tempo. Tramite una idonea gestione di portafoglio, i risparmi andrebbero a selezionare le migliori opportunità di rendimento a livello internazionale, potendosi rivolgere anche ai Paesi più giovani, quelli dove le fasce di età attive saranno ancora per molto tempo più numerose di quelle anziane (una sorta di riequilibrio internazionale degli effetti dell'invecchiamento). Diminuirebbe la pressione sugli attivi, si guadagnerebbe base imponibile al Fisco nazionale e, nella misura in cui i risparmi trovassero valide opportunità di investimento all'interno, si darebbe impulso alla capitalizzazione dell'economia e alla crescita. Ovviamente l'accumulazione non è una panacea; le mancano alcune proprietà positive e salienti sia della ripartizione (è il canale più adatto alla coesione *inter* e *inter* generazionale e alle finalità redistributive¹⁹) che della mutualità e della copertura assicurativa (contro grandi eventi avversi, l'accumulazione potrebbe non essere capiente o, se lo è, potrebbe venire in larga parte assorbita). Inoltre, un eccesso di accumulazione reale

¹⁸ Cfr. F. Pammolli e N.C. Salerno, *L'integrazione pubblico-privato nel finanziamento della sanità e della cura alla persona*, Quaderno Cerm, n. 3-2010, liberamente disponibile sul sito www.cermlab.it

¹⁹ Quantunque obiettivi redistributivi possono esser veicolati anche dal pilastro privato, previa un opportuno disegno delle agevolazioni fiscali (detraibilità *vs.* deducibilità), un coordinamento completo e organico con il pilastro di base universale e selettivo, e regole di funzionamento che favoriscano la copertura assicurativa collettiva dei rischi maggiori (come la *Ltc*).

esporrebbe troppo al rischio di mercato e potrebbe essere esso stesso fonte di instabilità dei mercati (in concomitanza con i flussi di investimento e di disinvestimento²⁰). Queste considerazioni, qui ripercorse in sintesi, lasciano intendere come uno degli aspetti più importanti di quel “territorio di mezzo”, verso cui cominciano ad avvicinarsi i sistemi sanitari dell’Europa e degli Stati Uniti, è proprio il riproporzionamento dei diversi criteri di finanziamento, in un *mix* che permetta di coglierne gli aspetti positivi e di controbilanciarne quelli negativi. Lungo questa strada, che sia per l’Europa che per gli Stati Uniti implica il rafforzamento del canale di finanziamento ad accumulazione reale, è necessario vagliare a fondo le sinergie e le complementarità con i fondi pensione, anche gettando da subito le basi di una cornice regolamentare condivisa e, progressivamente nel tempo, coordinata a livello internazionale²¹.

²⁰ Esiste un rischio di *overreliance* sulla accumulazione reale che è speculare e simmetrico a quello di *overreliance* sulla ripartizione.

²¹ Cfr. F. Pammolli e N.C. Salerno, *L’integrazione pubblico-privato nel finanziamento della sanità e della cura alla persona*, cit..

8. Conclusioni

Nei prossimi decenni tutti i Paesi ad economia e *welfare* sviluppati dovranno fronteggiare spese sanitarie fortemente crescenti, con *trend* potenziali che nel lungo periodo, al 2050-2060, potrebbero raddoppiare o più che raddoppiare l'incidenza della componente pubblica sul Pil. Così emerge dalle più recenti proiezioni Ecofin, Ocse e Fmi. Non si tratta di casi scolastici di proiezione, dal momento che la dinamica storicamente osservata dal Dopoguerra ad oggi ci ha già posto di fronte ad aumenti di questa proporzione e anche più forti. 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.

E si deve anche considerare che la dinamica storicamente osservata ingloba gli effetti delle politiche di governo e stabilizzazione della spesa. Corrispondentemente, la dinamica potenziale, sottostante i profili di incidenza della spesa sul Pil dal 1960 ad oggi, è stata addirittura superiore. Se per la spesa effettiva si deve parlare di incidenza sul Pil raddoppiata e, in alcuni casi, aumentata di multipli sino a tre e oltre, per la spesa potenziale l'aumento sarebbe stato molto più intenso. Questo implica, tra l'altro, che si debba prestare molta attenzione nel ricorrere a stime econometriche sui dati storici (spesa pubblica/privata/complessiva e relative variabili esplicative) per ottenere coefficienti utilizzabili in funzione di proiezioni di medio-lungo termine. Esiste una sorta «critica di Lucas» applicabile a questa modellistica, che suggerisce che i coefficienti stimati sullo storico sono sporcati dalla *policy* e dalla *governance* e, se utilizzati per proiezioni, darebbero per scontato la replicabilità *tout court* delle stesse scelte già effettuate in passato, ignorando totalmente che i vincoli e le pressioni di natura economica, politica e sociale saranno diversi e probabilmente con aspetti e proporzioni non ancora sperimentati. In ambito sanitario le politiche di stabilizzazione sono state sempre presenti dal Dopoguerra ad oggi ed anche in maniera più pervasiva che in altri capitoli di spesa, e non è detto che il problema della distorsione della stima possa essere risolto con l'inserimento di *dummy* che tentino di cogliere le discontinuità istituzionali e legislative-normative.

Di fronte a tali prospettive, le conclusioni di questo scritto non voglio essere negative ed allarmiste. Al contrario, è necessario che il *policy maker* si attivi subito affinché gli strumenti a disposizione siano all'altezza di governare e bilanciare costantemente il *trade-off* tra sostenibilità finanziaria e adeguatezza delle prestazioni. Il solo finanziamento a ripartizione non potrà reggere l'impatto; ma non potrà essere risolutivo neppure l'affiancamento di finanziamenti privati che si affidino a soluzioni di pura mutualità o di assicurazione.

Criteri di finanziamento, in ambito pubblico o privato, che si affidino esclusivamente o prevalentemente al concorso, anno per anno, di tutti gli assicurati per finanziare il fabbisogno espresso nello stesso anno, inevitabilmente produrranno effetti distorsivi e depressivi. È impensabile, infatti, che la massa reddituale disponibile anno per anno possa sostenere il confronto con una spesa a finanziare che crescerà a ritmi così forti. E assieme agli effetti distorsivi/depressivi è certo che emergeranno fenomeni di razionamento e restringimento del perimetro delle prestazioni offerte (sia nel pubblico che nel privato), oppure rinunce dei diretti interessati a dotarsi di coperture sanitarie (nel privato)²².

²² Si veda il sintetico ma incisivo, se non addirittura caustico, il commento di Kenneth Rogoff (2007), *Better red than dead?*, su Project Syndicate (<http://www.project-syndicate.org/commentary/rogoff33/English>).

Tutti i Paesi, con sistemi a prevalenza pubblica o privata, e tra i pubblici sia quelli di tradizione *bismarckiana* che *beveridgiana*, dovranno affrontare ristrutturazioni profonde nei meccanismi di finanziamento del loro *welfare*.

Un punto cardine dovrà essere la diversificazione multipilastro del finanziamento, per affiancare al criterio a ripartizione pubblico (*pay-as-you-go*), a quello privato di mutualità pura, e a quello assicurativo privato (*pooling* supportato da riserve matematiche), il criterio basato su investimenti reali di lungo termine sui mercati (l'accumulazione reale)²³. Un punto che, nei meccanismi fondamentali, si presenta *tel quel* sia per la sanità che per le pensioni. Finché l'affiancamento del privato al pubblico avviene rimanendo nell'ambito della messa in comune di risorse prodotte anno per anno, per far fronte alle esigenze di coloro che esprimono, in quello stesso anno, richieste di prestazioni, la diversificazione *multipillar* non è in grado di produrre cambiamenti strutturali. Il fattore, in grado di fare la differenza sul piano strutturale, è l'inserimento di sufficienti quote di accumulazione reale, tramite programmi di investimento reale di lungo termine.

Laddove ripartizione, mutualità e *pooling* utilizzano redditi già prodotti (intervengono *ex-post*, per dedicare alla finalità sanitaria quote di risorse già prodotte), i programmi di investimento entrano nei processi di generazione delle risorse e le preparano per tempo. Tramite una idonea gestione di portafoglio, i risparmi andrebbero a selezionare le migliori opportunità di rendimento a livello internazionale, potendosi rivolgere anche ai Paesi più giovani, quelli dove le fasce di età attive saranno ancora per molto tempo più numerose di quelle anziane (una sorta di riequilibrio internazionale degli effetti dell'invecchiamento). Diminuirebbe la pressione sugli attivi, si guadagnerebbe base imponibile al Fisco nazionale e, nella misura in cui i risparmi trovassero valide opportunità di investimento all'interno, si darebbe impulso alla capitalizzazione dell'economia e alla crescita.

Ovviamente l'accumulazione non è una panacea; le mancano alcune proprietà positive e salienti sia della ripartizione (è il canale più adatto alla coesione *inter* e *inter* generazionale e alle finalità redistributive) che della mutualità e della copertura assicurativa (contro grandi eventi avversi, l'accumulazione potrebbe non essere capiente o, se lo è, potrebbe venire in larga parte assorbita). Inoltre, un eccesso di accumulazione reale esporrebbe troppo al rischio di mercato e potrebbe essere esso stesso fonte di instabilità dei mercati (in concomitanza con i flussi di investimento e di disinvestimento).

Queste considerazioni, qui ripercorse in sintesi, lasciano intendere come uno degli aspetti più importanti di quel "territorio di mezzo", verso cui cominciano ad avvicinarsi i sistemi sanitari dell'Europa e degli Stati Uniti, è proprio il riproporzionamento dei diversi criteri di finanziamento, in un *mix* che permetta di coglierne gli aspetti positivi e di controbilanciarne quelli negativi. La ripartizione (soprattutto in ambito pubblico ma anche in ambito privato con la mutualità e il *pooling*) ha delle prerogative insostituibili, ma i suoi effetti positivi tendono ad essere sopraffatti da quelli negativi, quando la scala di funzionamento supera una certa soglia critica. Uno speculare discorso può esser fatto per l'accumulazione reale. Si deve ricercare il «giusto dosaggio». Lungo questa strada, che sia per l'Europa che per gli Stati Uniti implica il rafforzamento del canale di finanziamento ad accumulazione reale, è necessario vagliare a fondo le sinergie e le complementarietà con i fondi pensione, anche gettando da

²³ Come sinteticamente descritto nel paragrafo 7., i due criteri hanno virtù che si complementano e difetti che si controbilanciano, sia a livello microeconomico (per quanto concerne gli incentivi individuali al lavoro e 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).

subito le basi di una cornice regolamentare condivisa e, progressivamente nel tempo, coordinata a livello internazionale.

Dal «giusto dosaggio» delle modalità di finanziamento²⁴ dipenderà anche la possibilità di dare copertura finanziaria agli altri istituti di *welfare* a carattere redistributivo (famiglia, figli/minori, conciliazione vita-lavoro, accesso/mantenimento casa di abitazione, contrasto povertà, etc.), che non possono prescindere dalla copertura pubblica e dal finanziamento a ripartizione. Se la ripartizione sarà «monopolizzata» dalla sanità, in aggiunta al carico di finanziamento assorbito dalle pensioni, poco o nullo spazio rimarrà per dare copertura agli altri istituti di *welfare* e agli altri bisogni.

Sostenibilità finanziaria e sostenibilità sociale sono intrinsecamente connesse e passano per il crocevia di un *multi-pillar* che, nel *mix* di finanziamento, abbia una sufficiente quota di accumulazione reale delle risorse. Il «giusto dosaggio» delle modalità di finanziamento sarà un aspetto cruciale delle politiche sanitarie e *welfariste* dei prossimi anni.

Luglio 2010

CERM - Via G. Poli n. 29 - 00187 ROMA - ITALY
Tel.: 06 - 69.19.09.42 - Fax: 06 - 69.78.87.75
www.cermlab.it

²⁴ Per il caso italiano, *cfr.* la ricostruzione normativa e le proposte in F. Pammolli e N.C. Salerno, *L'integrazione pubblico-privato nel finanziamento della sanità e della cura alla persona*, cit.. A questo contributo si rimanda anche per una trattazione approfondita sul disegno dello strumento, il fondo *welfare*, che può permettere l'introduzione di quote di accumulazione reale nel finanziamento delle prestazioni sanitarie, farmaceutiche e di assistenza alla persona. Le caratteristiche peculiari del fondo *welfare* sono anche sintetizzate in F. Pammolli e N.C. Salerno, *Il pilastro complementare a capitalizzazione: Tredici buona ragioni per i fondi aperti per il welfare*, Nota Cerm n. 3-2008.

CERM - *Competitività, Regolazione, Mercati*
Via G. Poli n. 29
00187 ROMA, Italy
Tel.: 06 - 69.19.09.42
Fax: 06 - 69.78.87.75
www.cermlab.it
cermlab@cermlab.it

Competitività
Regolazione
Mercati

CERM