

POLITIQUES DE SANTÉ DANS UN MODÈLE MACROÉCONOMIQUE : *UN EXERCICE DE SIMULATION*

Nicolas Couderc

Stagiaire au Département des études de l'OFCE

Vincent Touzé, Bruno Ventelou *

Département des études de l'OFCE

Cet article propose une étude de la dynamique de la dépense de santé dans un modèle de croissance avec accumulation de capital, et analyse l'incidence de quelques variantes de politiques de santé (augmentation du ticket modérateur, contingentement, contrôle qualitatif) sur les grandes variables macroéconomiques, l'état de santé de la population et la satisfaction des agents. Les offreurs de soins réalisent un choix d'allocation du temps de travail entre deux activités, l'une utile, l'autre inutile (manière de rendre compte d'un comportement de « demande induite par l'offre »). La demande de soins émane des ménages, qui procèdent à des arbitrages entre la consommation de bien agrégé et des biens de santé dont ils ne payent qu'une partie des coûts (afin de rendre compte d'une myopie du patient). Ces choix sont sensibles aux progrès techniques de gamme et de productivité qui affectent les différents secteurs de production.

Le niveau français des dépenses de santé inquiète : la part des dépenses de santé en France représente 9,6 % du PIB en 2000, la situant dans la fourchette haute des pays de l'OCDE (8 % en moyenne), et en tête des pays de l'Union européenne. Le caractère socialisé de la consommation du bien santé est, tout d'abord, accusé de pousser les ménages à surconsommer, car, du fait de la participation de l'assurance maladie, ces derniers ne supportent qu'une fraction du coût des soins. Par ailleurs, on subodore des comportements stratégiques de la part des offreurs du secteur, visant à susciter une « demande induite »¹ de soins.

* Les auteurs remercient Jean-Paul Fitoussi et Jacques Le Cacheux pour leurs commentaires sur une première version de l'article. Cette recherche a reçu le soutien financier de la DREES, ministère de l'Emploi et de la Solidarité. Les idées défendues ici n'engagent que leurs auteurs.

1. Cf. Rochaix et Jacobzone (1997) pour une définition générale de la « demande induite », basée sur une analyse de la relation offreurs et demandeurs de soins médicaux. La distorsion de consommation est la conséquence d'une asymétrie informationnelle sur la qualité réelle du produit vendu : l'offreur de soins dispose d'une meilleure information que le patient sur l'efficacité des soins prescrits. Il en tire parti, pour susciter une demande de soins, qui — en information parfaite — n'aurait pas lieu d'être.

Juillet 2001

Revue de l'OFCE 78

Face à ces interrogations, diverses voies sont étudiées pour tenter de freiner les dépenses de santé, en réduisant la part qualifiée « d'inutile » des dépenses de santé (*i.e.* qui n'améliorent pas l'état de santé des patients), sans toucher aux dépenses « utiles ». Parmi les voies proposées, les politiques de contingentement des soins de santé sont, de manière explicite ou implicite, sollicitées dans ce but. Le bien-fondé de ce type de mesures est néanmoins dépendant de l'élasticité des consommations dites « inutiles » aux mesures de contingentement. Lorsque la marge de manœuvre des offreurs de soins est importante, rien ne dit qu'elles conduisent réellement à l'objectif visé.

On a donc cherché à construire un modèle macroéconomique d'équilibre général à générations imbriquées, pertinent pour analyser l'incidence des politiques de la santé sur l'économie, le taux d'épargne macroéconomique, l'état de santé général de la population. La section 2 de cet article a pour but d'exposer les soubassements théoriques du modèle. Dans une section 3, on cherchera à analyser les effets du progrès technique de gamme et de productivité² sur les consommations de santé des agents ; on analysera aussi les logiques à l'œuvre face à une modification éventuelle de la qualité des soins (le rapport « utile/inutile »). Enfin, quelques variantes pertinentes des politiques de la santé seront étudiées (contingentement quantitatif ou combiné avec un contrôle des prix, enveloppe globale avec reversements...), en portant une attention particulière à leurs effets sur le poids des dépenses de santé dans le PIB et sur l'état de santé de la population. Cette version de l'article présente les résultats sous forme de simulation. L'intégralité du modèle et les résultats analytiques sont présentés dans Touzé et Ventelou (2000).

Présentation du modèle

Le cadre général du modèle est proche de celui proposé par Diamond (1965) : on considère une économie composée d'agents vivant deux périodes et travaillant uniquement lors de la première. Nous posons l'existence de deux secteurs économiques, santé et bien de consommation courante.

Les fonctions de production du bien de consommation agrégé sont à rendements factoriels décroissants, le progrès technique est neutre au sens de Hicks, et les facteurs de production sont rémunérés à leur productivité marginale. On est donc dans un cadre classique de modèle macroéconomique à générations imbriquées avec accumulation de capital et croissance exogène (le taux d'accumulation des connaissances qui permet un progrès technique de productivité est stable et exogène).

Le modèle comporte trois types d'agents. Les agents sont : i) les offreurs de soins (plus exactement les personnels travaillant dans le secteur de la santé) et les salariés des entreprises productrices du bien de consommation courante agrégé : ceux-ci se répartissent en deux sous-populations, ii) les malades et iii) les non-malades. La part de ménages malades dans l'économie est constante

2. Cf. plus loin pour une distinction des progrès techniques de gamme et de productivité dans le secteur médical, inspirée de Weisbrod (1991).

au cours du temps. Une vue synthétique des principes retenus dans la construction du modèle est permise par le tableau économique d'ensemble (tableau 1, *infra*).

Le secteur médical

Les offreurs de soins tirent leurs revenus de la vente des biens santé, payés pour une part par les malades (*ticket modérateur*), pour le reste par l'assurance maladie³. Ce revenu est l'objet d'un arbitrage intertemporel (consommation immédiate *versus* consommation future) fonction du rendement de l'épargne.

De plus, ils réalisent un arbitrage dans l'allocation de leurs efforts entre deux types d'activités (utiles et inutiles), en fonction des coûts subjectifs associés à chaque type d'effort⁴. La part inutile de la production est produite et vendue, en information asymétrique, du seul fait de la capacité des offreurs de soins *d'induire la demande* (Evans, 1974).

Les rythmes de progrès technique de productivité sont différents dans la production de bien médical utile et inutile. Le taux de progrès technique de gamme est commun, stable et exogène. Néanmoins, nous verrons que l'apparition de nouveaux biens médicaux tend à déformer les quantités relatives de biens médicaux utiles et inutiles.

Les ménages « malades » en première période

La population des malades⁵ est la seule à consommer du bien santé en première période de vie. Comme on vient de le voir, ce bien santé est par hypothèse produit en deux qualités : la première est utile (il améliore l'état de santé du patient), la seconde ne l'est pas ; les malades sont par nature incapables de différencier à l'avance les soins utiles des soins inutiles, du fait de l'asymétrie d'information inhérente à la relation médecin/patient⁶. En ce qui concerne les ménages inactifs (i.e. vivant leur deuxième période de vie), nous supposons que la dépense de santé, et la morbidité, sont systématiques (elles surviennent avec une probabilité égale à 1).

3. Du fait des hypothèses retenues (voir annexe 2, fonctions d'utilité logarithmiques), l'offre de bien santé est inélastique au prix. L'offre de bien dépend des coûts subjectifs et des conditions physiques de production (progrès technique de productivité et de gamme). La proportion de biens médicaux utiles sur l'ensemble des biens médicaux et la quantité totale de bien santé produite sont donc déterminées par les offreurs de soins, avant leur confrontation au prix d'équilibre du bien santé.

4. On peut par cette voie envisager d'introduire la déontologie médicale dans le modèle : un coût subjectif de l'activité de production de bien santé inutile élevé correspond, toutes choses égales par ailleurs, à une forte déontologie médicale.

5. Pour simplifier la présentation, on appelle « malades » les ménages actifs victimes de l'aléa maladie, et « non-malades » les ménages actifs non victimes de cet aléa (bien que l'ensemble des ménages, une fois inactifs, consomment tous des soins de santé).

6. Voir annexe 2.

L'existence d'une assurance maladie augmente encore la surconsommation de bien médical. Le fonctionnement de l'assurance est simple : l'ensemble des salariés paie une cotisation basée sur leur salaire brut, et les caisses d'assurance maladie prennent à leur charge une fraction du coût de la santé, les patients ne payant alors plus qu'un « *ticket modérateur* » (part du prix de la consommation santé non pris en charge par l'assurance maladie). Dans une variante du modèle, on a aussi envisagé l'hypothèse dans laquelle l'assurance maladie reverse au malade un paiement forfaitaire, laissant peser sur lui le coût explicite de la santé : l'agent ne peut alors connecter sa dépense santé à la valeur du paiement forfaitaire ; dans cette variante, l'arbitrage de l'agent entre consommation de bien agrégé et bien de santé n'est pas faussé par le co-paiement. On pose comme hypothèse de travail que les comptes de l'assurance maladie sont équilibrés : le taux de cotisations obligatoires varie pour assurer l'équilibre des comptes.

Enfin, il convient de préciser que l'existence d'un progrès technique de gamme modifie la demande en bien santé des agents malades, et en définitive leur goût pour la santé. L'apparition de nouveaux biens santé permet la satisfaction d'une demande préexistante qui ne pouvait être comblée du fait de la carence de l'offre. Des maladies jusque là incurables, des pathologies traitées avec des techniques peu perfectionnées, ou peu confortables pour le patient, deviennent traitables ; la demande de santé des malades croît parallèlement (on peut assimiler cette hausse à l'existence d'un goût des agents pour *la variété* de bien santé). Le progrès technique de gamme aura des effets complexes sur l'équilibre du secteur santé, car il agit sur l'offre médicale aussi bien que sur la demande de bien santé. Cela ne remet pas en cause l'existence de deux qualités de travail offert par le secteur médical⁷.

L'intégration du progrès technique de gamme dans le modèle

On choisit de distinguer dans le modèle deux types de progrès technique agissant sur la production du bien médical. On suit en cela la distinction faite par Weisbrod (1991) :

— L'extension de la gamme (i.e. l'augmentation du nombre de soins, traitements ou médicaments à la disposition des malades) permet d'envisager de traiter des maladies ou pathologies précédemment incurables, ou simplement de proposer des soins plus confortables au patient (on peut, dans une vision extrême du progrès technique de gamme, supposer qu'une part de ce progrès crée une complexification « cosmétique » des soins). Il doit *a priori* augmenter le coefficient des dépenses de santé dans le budget des ménages.

— Au contraire, le progrès technique de productivité, à gamme de soins constante, permet aux offreurs de soins de produire les soins de manière plus efficace, à moindre coût. Il permet *a priori* des « économies » de dépenses de santé.

Du côté de la demande, le progrès technique de gamme, I_p , est intégré dans la fonction de préférence des agents de la manière suivante (en notant Z_t la

7. Pour la traduction mathématique de ces hypothèses, on se reportera à l'annexe 2, « choix de modélisation ». Une présentation des résultats du modèle sur l'influence du progrès technique de gamme sur la demande de bien de santé est faite *infra*.

consommation des agents lors de leur deuxième période de vie, h_t la consommation de bien santé des malades et γ leur goût pour la santé) :

$$U(C_t, Z^+, h_y, h_o^+) = \ln C_t + \gamma \sum_{c=1}^{I(t)} \ln(h_{y,c}) + \beta \ln Z^+ + \gamma^\# \sum_{c=1}^{I(t+\#)} \ln(h_{o,c}^+) \quad (1)$$

Donc, en réduisant l'hétérogénéité des soins de santé et en définissant \bar{h}_t comme le bien médical moyen, on obtient :

$$U(C_t, Z^+, h_y, h_o^+) = \ln C_t + \gamma I_t \ln \bar{h}_y + \beta (\ln Z^+ + \gamma^\# I^+ \ln \bar{h}_o^+) \quad (2)$$

Les innovations médicales de gamme sont valorisées par les agents sans décroissance de l'utilité marginale (effet variété). Tout se passe finalement comme si le goût pour la santé des agents était modifié de manière à incorporer les nouveaux produits.

Côté offre, le progrès technique joue de deux façons : d'une part, il diminue le temps de travail l disponible pour chaque activité médicale (puisque le nombre $I(t)$ de celles-ci augmente), mais d'autre part, augmente le nombre de biens disponibles et donc l'offre médicale totale.

$$h_t^s = \sum (\bar{h}_{c,t}^s), \text{ avec } \bar{h}_t^s = B \left(\frac{l}{I_t} \right)^b \quad (3)$$

Les ménages « non-malades » en première période

Ces ménages paient leur cotisation à l'assurance maladie, mais ne consomment pas de bien santé : l'aléa maladie de première période de vie ne s'est pas réalisé pour ces ménages. En revanche, une fois inactifs, les « non-malades » sont soumis au même arbitrage que les ménages « malades » : la maladie devient une certitude. L'arbitrage intertemporel des non-malades est donc entre la consommation immédiate et l'épargne, cette dernière étant destinée à consommer du bien agrégé et du bien santé en seconde période de vie, en fonction de leur goût pour la santé et de leur préférence pour le présent. L'épargne des ménages non-malades en première période, comme d'ailleurs celle des autres types d'agents, est placée auprès des entreprises, et rémunérée à un taux d'intérêt égal à la productivité marginale du capital.

Bouclage macroéconomique et équilibre en situation de laisser-faire

Ce sont les marchés du capital et de la santé, qui permettent tout à la fois d'obtenir l'équilibre général et d'établir la dynamique du modèle.

On admet que le « stock de capital » de l'économie est fourni et utilisé par le secteur traditionnel (i.e. : le bien de consommation agrégé ; autrement dit, il n'y a pas de capital dans la production de santé). Le « capital » est égal à l'épargne totale de la période précédente, ce qui signifie, d'un point de vue économique, que tout investissement productif doit être financé *ex-ante* par une accumulation d'épargne (et que le taux d'amortissement est fixé à zéro).

I. Tableau économique d'ensemble *

	TEE							
	Emplois			Ressources				
	Ménages malades	Ménages non-malades	Entreprise	Offreurs de soins	Ménages malades	Ménages non-malades	Entreprise	Offreurs de soins
Opérations sur le bien de consommation	c, z	c', z'		$c_{im} z_m$			y^1	
Opérations sur le bien santé	$(1-\lambda) ph^y + ph^o$	$ph^{o'}$						$(1-\lambda) ph^y + ph^o$
Assurance santé	π	π			J^2			λph^y
Revenus du travail et de l'épargne			$w_e (1+r) s_{TOT}$		$w_e (1+r) s$	$w_e (1+r) s'$		$(1+r) s_m$
	TOF							
	Flux nets de créances			Flux nets de dettes				
Épargne	s	s'		s_m				s_{TOT}

* On se reportera à l'annexe I en fin d'article pour la définition des notations utilisées.

1. Le prix du bien de consommation agrégé est arbitrairement normé à 1.

2... dans la variante qui comporte un paiement forfaitaire des actes médicaux. On admet alors que les agents ne peuvent connecter ce paiement forfaitaire à leurs dépenses de santé.

Résultats généraux : l'analyse positive

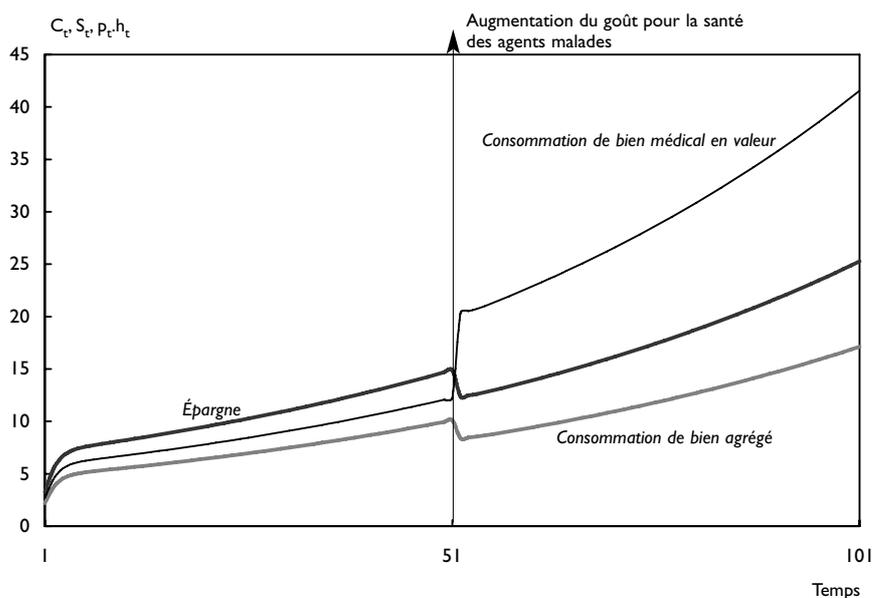
On commence par analyser les propriétés descriptives du modèle : en quoi le modèle permet-il de reproduire les faits stylisés de l'évolution de la dépense de santé dans l'économie ? Après un travail élémentaire de statique comparative (destiné essentiellement à tester la robustesse du modèle), on passe à l'analyse des distorsions et aux effets des progrès techniques.

Rôle des paramètres exogènes : un test du modèle

Goût pour la santé

Une augmentation du goût pour la santé des agents augmente leur demande de bien santé, au détriment de leur demande de bien de consommation agrégé et d'épargne. Cette augmentation de la demande de santé conduit à une élévation du prix du bien santé (elle vient se confronter à une offre médicale rigide). De ce fait, la consommation de bien santé en valeur augmente. On vérifie bien ces intuitions :

I. Effet d'une modification du goût pour la santé des agents

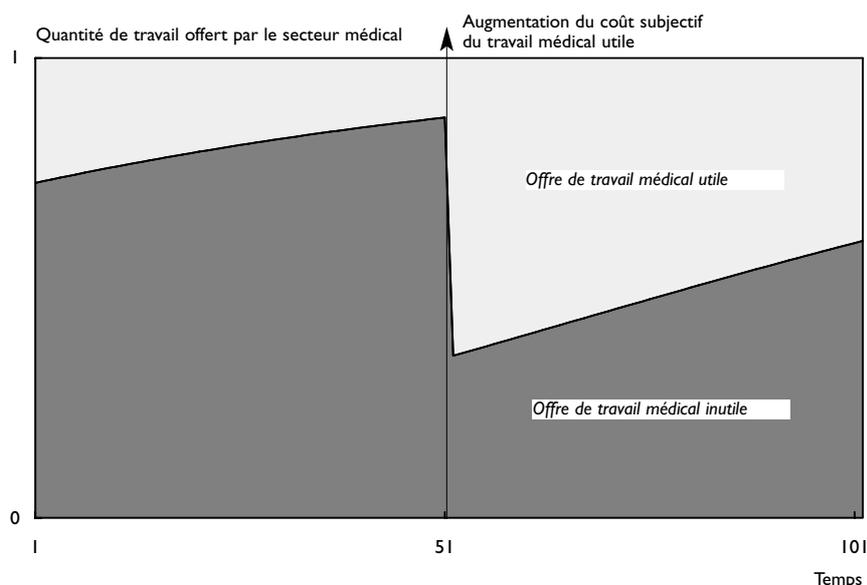


Source : Calculs des auteurs.

Arbitrage des offreurs de soins entre les deux types d'activités possibles

Une augmentation du coût subjectif de l'effort utile (respectivement inutile) des offreurs de soins conduit à une baisse de l'offre de bien utile (respectivement inutile). De même, une augmentation de l'efficacité dans la production du bien utile conduit, toutes choses égales par ailleurs, à une augmentation de l'offre de bien utile.

2. Effet d'une augmentation du coût subjectif de l'activité utile pour les offreurs de soins (avec progrès technique de productivité dans le secteur utile)



Source : Calculs des auteurs.

L'analyse des distorsions : myopie et demande induite

Taux de remboursement des soins et paiement forfaitaire

Deux cas types sont examinés : le système d'assurance par co-paiement, et un système de paiement forfaitaire pur. Il va de soi que les deux systèmes peuvent coexister dans la réalité.

Une augmentation du taux de remboursement des soins fait augmenter la demande adressée au secteur médical, car les agents voient diminuer le prix « apparent » du bien santé (ils ne perçoivent que le « ticket modérateur », qui baisse)⁸. L'offre étant rigide, on obtient une hausse du prix du bien santé et, donc, une modification du partage de la valeur ajoutée :

8. Ce mode de remboursement conduit les agents à n'internaliser que partiellement le coût de leur consommation de soin. La part socialisée du coût de la santé n'intervenant pas dans les choix des agents, il y a une distorsion dans l'allocation des ressources.

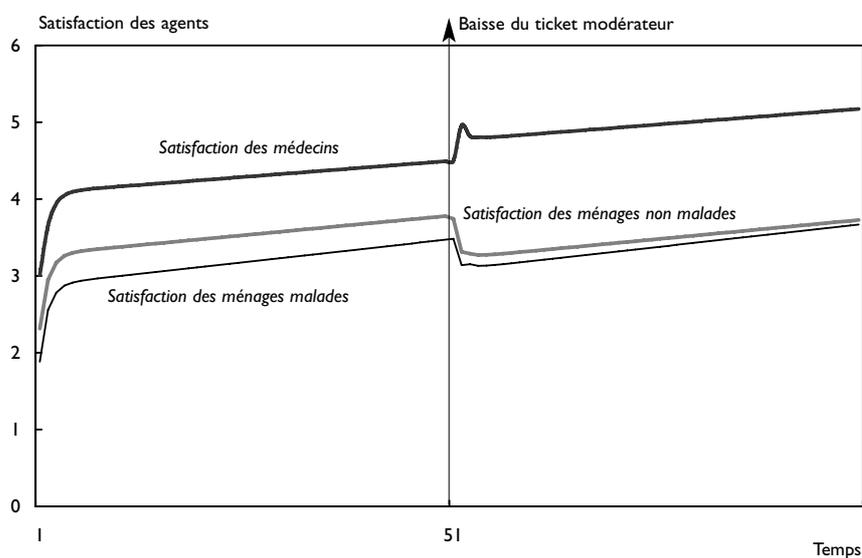
— au détriment des non-malades qui contribuent plus par unité de soin de santé (l'assurance est plus chère car plus complète), et à qui on fait couvrir une dépense de santé plus élevée (l'assurance est plus chère car les dépenses sont plus nombreuses) ;

— au profit des offreurs de soins, qui – à offre inchangée – profitent d'une augmentation du prix du bien santé⁹ ;

— en première analyse, au profit des malades, qui bénéficient plus largement de la solidarité inhérente au principe de l'assurance (la société entière cotise à l'assurance maladie, et celle-ci rembourse une partie plus importante des soins aux malades) ; néanmoins, en seconde analyse, ceux-ci sont eux aussi victimes de la distorsion causée par le co-paiement (la myopie sur le coût des soins et la surconsommation en résultant) : on peut craindre donc, à l'inverse de l'effet « assurance », une dégradation nette de leur bien-être.

Concernant les malades, dans le cadre de la modélisation retenue, c'est le second effet qui l'emporte sur le premier : la perte de satisfaction consécutive à la distorsion en bien de santé est supérieure au gain de satisfaction des assurés.

3. Effet d'une augmentation du taux de remboursement des soins (baisse du ticket modérateur) sur la satisfaction des agents

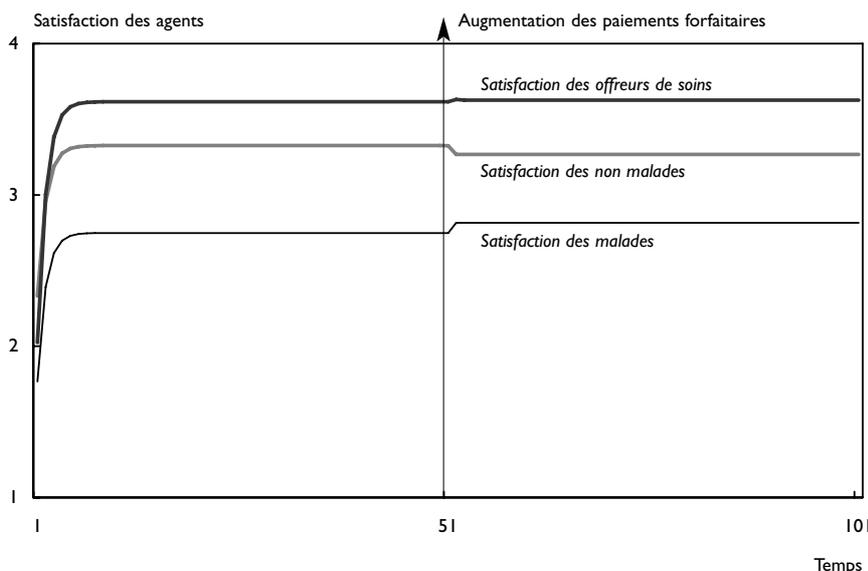


Source : Calculs des auteurs.

9. Cette surconsommation médicale est-elle la source d'une perte nette d'utilité sociale ? Pour répondre à cette question, il faudrait pouvoir comparer, de manière cardinale, la perte d'utilité des ménages (dues à la surconsommation médicale) et le gain d'utilité des offreurs de soins (dues aux mêmes raisons). En l'absence de comparaison cardinale d'utilité, il faut se contenter de constater que le co-paiement conduit à un partage de la richesse nationale en faveur des offreurs de soins plus avantageux que celui qui s'observerait avec un remboursement incitatif (de type paiement forfaitaire). On peut voir Touzé et Ventelou (2000) pour une analyse à l'aide d'une fonction d'utilité collective.

Le recours à la méthode des paiements forfaitaires a un impact très différent : au lieu de rembourser une fraction des dépenses de santé à l'assuré, l'assurance maladie donne au patient une somme fixée *ex-ante* pour chaque type de maladie qu'il subit, quelles que puissent être ses dépenses réelles relevées *ex-post*. Ainsi, l'arbitrage réalisé par le patient n'est pas altéré (pas de myopie sur le coût) : une augmentation du paiement forfaitaire n'augmente pas la demande des malades en bien de santé (aux effets revenus près). Par contre, le partage de la valeur ajoutée est modifié, toujours à cause du système d'assurance : les malades sont « mieux assurés » (ils bénéficient d'une couverture plus complète).

4. Effet d'une augmentation des paiements forfaitaires de l'assurance-maladie



Source : Calculs des auteurs.

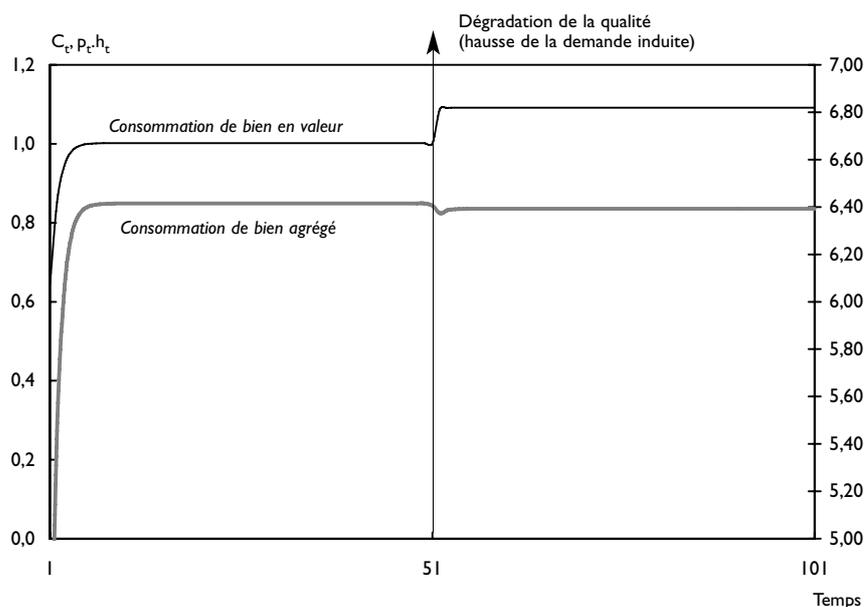
Le graphique 4 fait, en outre, apparaître une légère amélioration du bien-être du secteur médical. En redistribuant le revenu des non-malades vers les malades, l'assurance à paiement forfaitaire augmente la propension moyenne à consommer le bien santé dans l'économie, ce qui bénéficie au secteur médical.

Rôle de la demande induite

On la mesure par le rapport production de bien médicaux inutiles sur production médicale totale. Selon nos hypothèses, le consommateur de soins ne peut différencier *ex-ante* les soins utiles des soins inutiles. Il sait qu'une part $1-\alpha$ ¹⁰ des soins prescrits sont inutiles ; et c'est *connaissant ce paramètre* (considéré comme « fatal ») qu'il établit son comportement d'acheteur de soins.

10. α étant la part utile de la consommation médicale, on appellera $(1-\alpha)$ « paramètre de demande induite ».

5. Utilisation du revenu des ménages malades en première période suite à une modification du paramètre de demande induite



Source : Calculs des auteurs.

En toute généralité, on peut envisager deux effets face à une variation positive de la « qualité » d'un bien :

— un effet de « satisfaction suffisante », une qualité moyenne élevée aura pour conséquence une demande de biens faible. Du fait de la bonne qualité des soins, les premières unités consommées fournissent déjà une utilité forte ;

— un effet « attraction », la qualité d'un bien peut avoir, à l'inverse, un effet positif : la consommation du bien santé est encouragée par sa qualité.

L'effet du paramètre α sur la demande finale de bien santé dépend de ces deux effets contradictoires. Ici, c'est le premier des deux effets qui domine (l'effet de « satisfaction suffisante »). Ce résultat est dû à nos choix de modélisation, précisément à l'utilisation d'une forme logarithmique pour les fonctions d'utilité des ménages (voir l'annexe 2) : on constate que α joue deux fois dans l'expression de la fonction de demande, une fois positivement, une fois négativement ; l'analyse mathématique de la fonction montre que l'ambiguïté est toujours résolue en faveur de l'effet négatif.

En raisonnant *a contrario*, dans le graphique 5, on constate qu'un paramètre de demande induite élevé (c'est-à-dire une qualité faible des soins prodigués) a pour conséquence un surcroît de demande de bien médical. La « demande induite », telle que nous la modélisons, entraîne donc une double distorsion :

— au premier rang, une part inutile de biens de santé est produite et achetée ;

— au second rang, cette première distorsion donne lieu à une seconde : face à la dégradation de la qualité des biens, les ménages choisissent d'en consommer plus encore afin de rétablir leur satisfaction.

Pour retrouver un état de santé donné, un malade est donc « condamné » à surconsommer du bien santé (l'effet est montré en valeur — *a fortiori* —, mais il existe aussi en volume).

Influence des taux de progrès techniques sur les arbitrages des agents et sur l'équilibre macroéconomique

Progrès technique de productivité dans la production du bien de consommation

Une augmentation du taux de progrès technique sur la production du bien de consommation courante a pour conséquence une augmentation du produit, et des revenus des agents. La consommation de bien de consommation courante, comme celle du bien de santé augmente de ce fait. Les coefficients budgétaires sont stables (l'arbitrage en valeur des ménages entre les deux postes de consommation n'est en rien modifié. Cela tient à la forme des fonctions d'utilité des agents et à l'élasticité-prix de la demande qui en résulte).

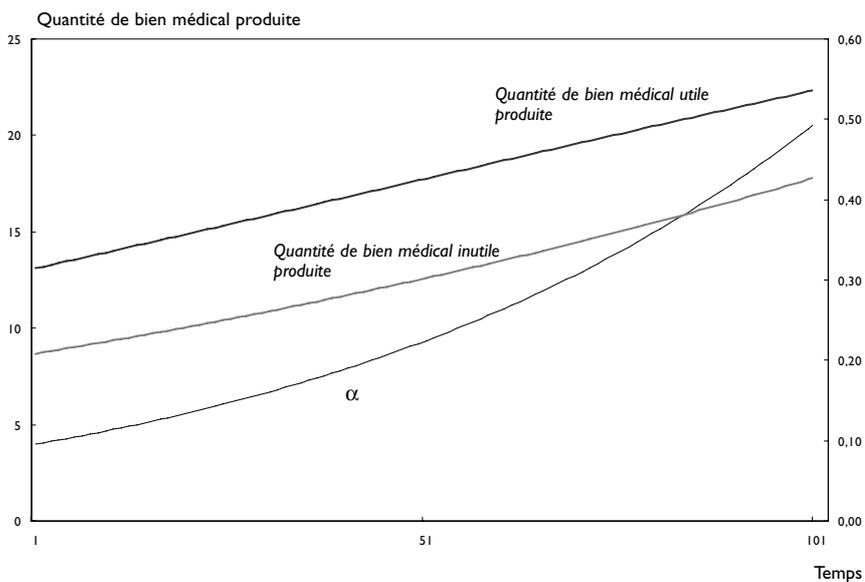
Progrès technique de productivité dans la production du bien médical

Les productions de bien de santé utile et inutile connaissent des rythmes de progrès technique différents. *Ceteris paribus*, on peut penser que peu à peu la structure de l'offre médicale se déforme vers le type de bien qui connaît le plus fort progrès technique. Le graphique 6 présente le cas où le progrès technique sur les soins utiles est le plus fort ; conformément à l'intuition, il y a effectivement déformation de la structure productive des soins en faveur des activités utiles.

En outre, toutes choses égales par ailleurs (notamment parce qu'il n'y a pas de progrès technique symétrique dans le secteur du bien de consommation standard), le prix du bien santé diminue du fait de gains de productivité, ce qui permet une augmentation des quantités consommées (tableau 7) :

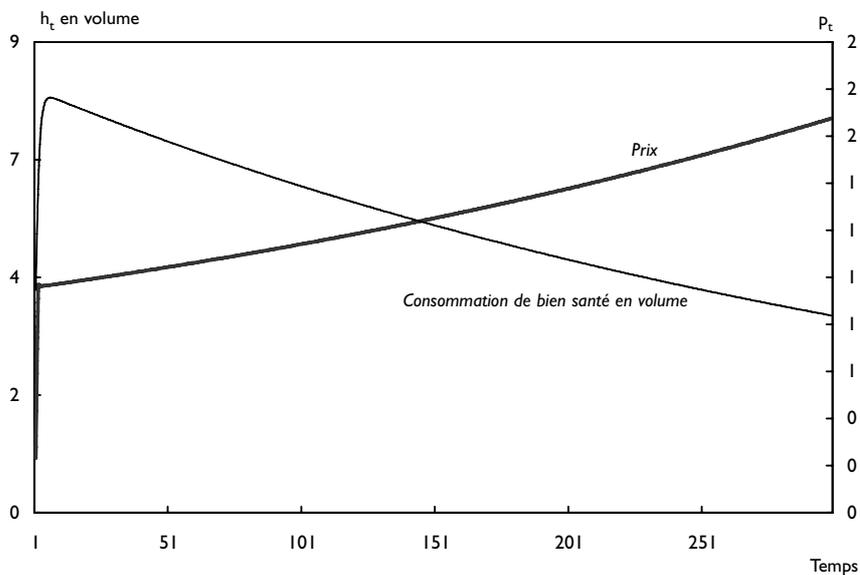
Cette baisse observée du prix du bien santé est cependant la conséquence de l'existence d'un progrès technique de productivité *dominant dans le secteur « utile »* du bien médical. En effet, celui-ci permet une modification de la part utile/inutile du bien de santé, et une demande de santé relativement moindre *ceteris paribus*, car l'effet de « satisfaction suffisante » joue (cf. *supra*). Au contraire, le progrès technique dans le secteur du bien médical inutile a pour effet paradoxal de faire augmenter le prix : la « qualité » du bien santé se dégrade, les agents malades demandent et consomment plus de bien de santé pour réparer cette dégradation. Ce qui se traduit par une augmentation du prix.

6. Arbitrage entre les activités de soins en présence d'un progrès technique de productivité fort dans la production du bien santé utile



Source : Calculs des auteurs.

7. Évolution de la consommation et du prix du bien médical avec progrès technique de productivité dans les deux secteurs médicaux

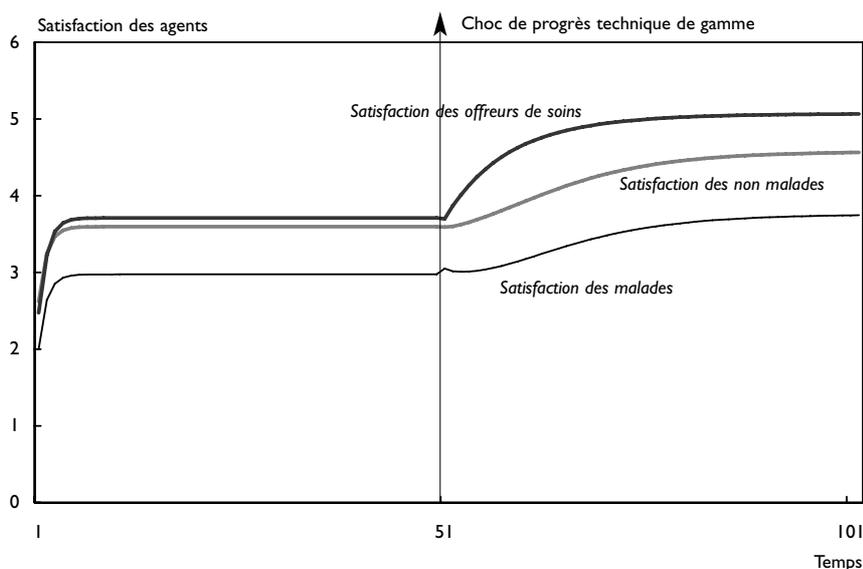


Source : Calculs des auteurs.

Progrès technique de gamme dans la production du bien médical

La satisfaction des agents évolue de manière différente selon les paramètres retenus. On peut consulter les annexes 2 et 4 et le graphique 10 : selon les choix de paramètres de la fonction de production des biens de santé, il se peut que le progrès technique de gamme biaise systématiquement l'arbitrage 'utile/inutile'. Dans le cas où le biais s'avère favorable aux biens utiles, la satisfaction des agents malades et non-malades augmente en présence du progrès technique de gamme : c'est le cas retenu dans les graphiques 8 et 9. En revanche, la satisfaction des malades peut décroître, si le progrès technique de gamme sert principalement à la production de bien santé inutile (graphique 10).

8. Satisfaction des agents avec progrès technique de gamme dans le secteur médical



Source : Calculs des auteurs.

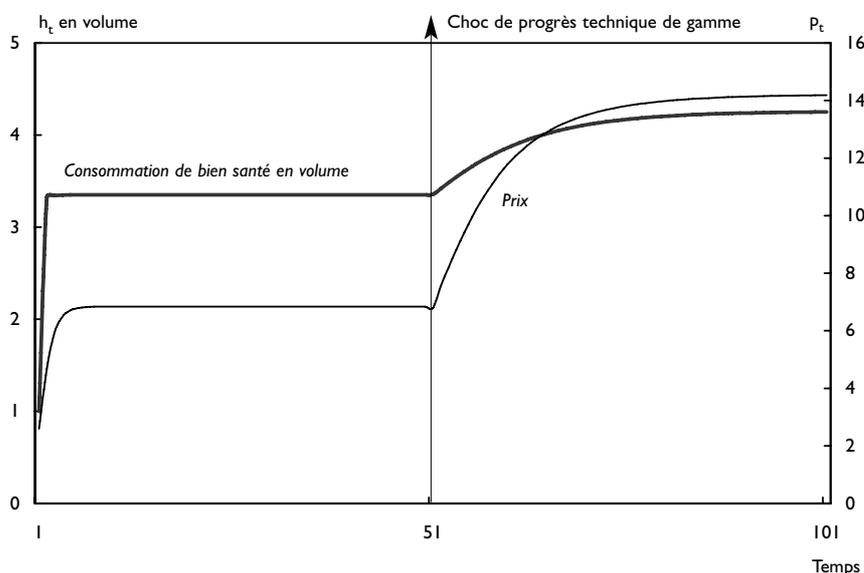
Les ménages malades et non-malades profitent des innovations médicales (les « non-malades » uniquement lorsqu'ils sont vieux). Les soignants augmentent leur chiffre d'affaire.

Le rôle du progrès technique de gamme sur le prix est *a priori* ambigu : d'une part, demande et offre en bien de santé augmentent (on peut dire que « l'aire » du secteur de santé s'étend) ; d'autre part, les offreurs de soins sont conduits à consacrer moins de temps à chaque soin, mais les incitations à travailler sont plus fortes (la hausse du chiffre d'affaire les incite à renoncer au loisir). Le prix du bien santé connaît une évolution non linéaire : selon les hypothèses retenues, sa phase de croissance est écourtée ou prolongée ¹¹.

11. A terme et en raisonnant asymptotiquement (lorsque l'extension de la gamme tend vers l'infini), le prix devient décroissant : l'offre médicale en augmentation constante (en nombre de

Dans le graphique présenté, le prix augmente tout au long de la période, ce qui conduit à une augmentation relativement forte de la satisfaction des offreurs de soins.

9. Évolution de la consommation et du prix du bien médical avec progrès technique de gamme



Source : Calculs des auteurs.

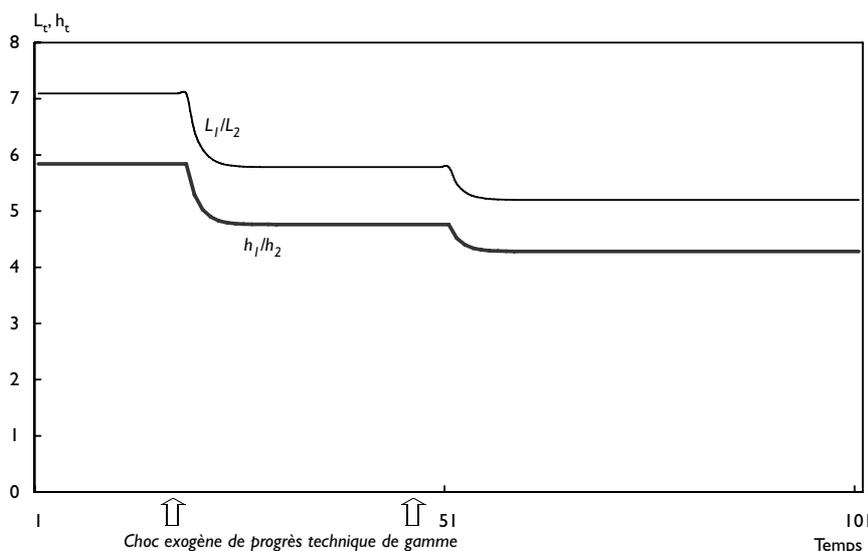
Il reste à étudier l'influence du progrès technique de gamme sur la « qualité » de biens médicaux. On observe que la déformation de la qualité sous l'effet du progrès technique de gamme dépend du signe de $b_1 - b_2$. Ces deux paramètres modélisent l'élasticité de la production des biens santé (utile et inutile) au facteur travail¹².

On note ici un biais systématique : le progrès technique de gamme conduit à une baisse tendancielle de la qualité des biens médicaux. L'inégalité $b_1 > b_2$ établit en effet une réactivité plus forte du secteur 1 aux variations de l'environnement. Devant l'augmentation du nombre de bien à produire, les offreurs de soins re-répartissent leur temps de travail, mais ils ont tendance à le faire moins vite sur le secteur 2, qui prend alors de l'importance.

biens) ne rencontre plus une demande de biens santé nouveaux. En dehors de ce comportement asymptotique, i.e. pour les I_t finis, le prix est une fonction croissante du progrès technique de gamme.

12. On a : $\bar{h}_1 = (l_1/I(t))^{b_1}$ et $\bar{h}_2 = (l_2/I(t))^{b_2}$, avec h_1 et h_2 production utile et inutile, l_1 et l_2 temps de travail alloué aux deux activités. Voir les annexes 2 et 4 pour une description précise des arbitrages réalisés par les offreurs de soins.

10. Effet du progrès technique sur l_1/l_2 et h_1/h_2 (avec $b_1 > b_2$)



Source : Calculs des auteurs.

Pour achever cette section, il convient de confronter ces premiers résultats du modèle aux faits stylisés. L'existence d'un double progrès technique, de gamme et de productivité, semble utile. En particulier, l'introduction du progrès technique de gamme dans le secteur médical permet d'expliquer la déformation progressive des coefficients budgétaires des ménages vers plus de dépenses médicales, ce que ne peut faire le progrès technique de productivité — du moins si on conserve les spécifications de ce modèle¹³. La dichotomie entre les deux types d'offre de travail et de biens médicaux rend bien compte du phénomène de « demande induite », que l'on peut retrouver dans le secteur médical réel, par le biais des sur-prescriptions, ou de la pratique du « renvoi d'ascenseur » entre médecins spécialistes.

Une hypothèse de modélisation pourrait néanmoins affaiblir l'ensemble : il tient à la rigidité de l'offre médicale au prix du bien santé (elle a été choisie pour les facilités techniques qu'elle impliquait). Toutefois, on peut justifier cette hypothèse par une étude empirique du travail médical, qui semble peu sensible aux changements du prix des consultations ou des soins médicaux malgré le statut libéral de la majorité des personnels médicaux¹⁴.

13. Une autre manière d'obtenir cette déformation aurait été d'utiliser des spécifications donnant une élasticité très forte de la demande de soins aux prix et au revenu, et, via les équations de Slutsky, d'expliquer la déformation des coefficients budgétaires par des variations de prix ou de revenus. De fait, nous proposons une démarche inverse : les coefficients budgétaires sont *a priori* indépendants de telles variations ; nous montrons alors que le progrès technique de gamme est une explication alternative plausible.

14. Sur les rapports complexes entre offre de soins de santé et prix des soins, on se reportera à Phelps [1995] : empiriquement, l'inélasticité aux prix se voit plutôt confirmée.

De l'influence des politiques de santé

Le critère d'appréciation des différentes variantes de politique de santé retenu est le bien-être. On choisit de ne pas agréger les gains de bien-être potentiel des différentes catégories d'agents. Ce parti pris est justifié par l'impossibilité de comparer « cardinalement » les satisfactions des offreurs de soins et des agents.

Le modèle ne converge pas vers un optimum parétien, du fait de sa construction. Il existe en effet deux distorsions de nature à éloigner l'économie de l'optimum : tout d'abord, l'existence d'une quantité de bien santé inutile, que les malades ne peuvent différencier du bien santé utile (asymétrie informationnelle entre offreurs de soins et malades) ; ensuite, l'existence d'un paiement proportionnel de l'assurance maladie, qui empêche une allocation optimale du revenu car elle occulte le prix payé par les agents pour la santé, les poussant une nouvelles fois à la surconsommation. Pour ces deux raisons, on pourrait être tenté d'appliquer des politiques de santé cherchant à contrôler, et encadrer, la consommation de santé¹⁵.

Nous présenterons successivement trois politiques d'encadrement : *numerus clausus*, contingentement quantitatif, et politique d'enveloppe. L'intuition première est que ces trois politiques sont équivalentes, il s'agit toujours de restreindre la consommation médicale. Une analyse plus fine, à l'aide du modèle, montre néanmoins que leurs propriétés économiques sont quelque peu différentes (les incitations et les comportements vont différer au plan microéconomique), sans toutefois que l'on soit véritablement conduit à préférer l'une ou l'autre des stratégies (les incitations microéconomiques diffèrent, mais elles ne sont jamais réellement en mesure de rétablir une optimalité sociale).

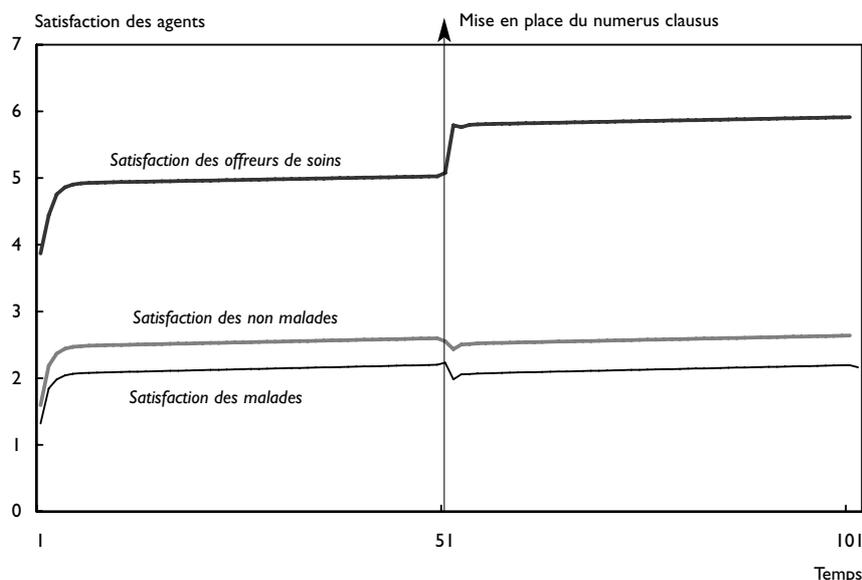
Limitations de l'offre médicale et politique de *numerus clausus*

On retrouve cette technique utilisée en France pour fixer le nombre de nouveaux médecins autorisés annuellement à pratiquer des honoraires libres en France. Dans le modèle, la mise en place d'un *numerus clausus* est représentée par la baisse du nombre de médecins dans la population totale.

Aucun changement dans les coefficients budgétaires des ménages n'intervient : la baisse de la quantité offerte est compensée par une hausse du prix des biens santé, ce qui empêche les ménages de réallouer une quelconque fraction de leur revenu vers d'autres consommations ou épargne. On observe que la seule catégorie d'agents à profiter d'une telle politique est la catégorie des offreurs de soins, qui augmentent leurs tarifs sans perdre de clients du fait de la raréfaction de l'offre (ils bénéficient d'une rente, proche de celle du monopole). Le rapport α n'est toutefois pas modifié par une telle politique, puisqu'il s'agit d'un arbitrage, microéconomique, que réalise chaque offreur de soins : la « qualité » d'une unité de soins n'est ainsi pas sensible à une mesure de limitation de l'offre médicale de type *numerus clausus* (cependant, la société est quantitativement limitée dans son accès aux soins).

15. Voir Batifoulier et Touzé (2000) sur les politiques de santé effectivement appliquées.

II. Mise en place d'un numerus clausus et évolution de la satisfaction des agents



Source : Calculs des auteurs.

Mesures de contingentement quantitatif pur

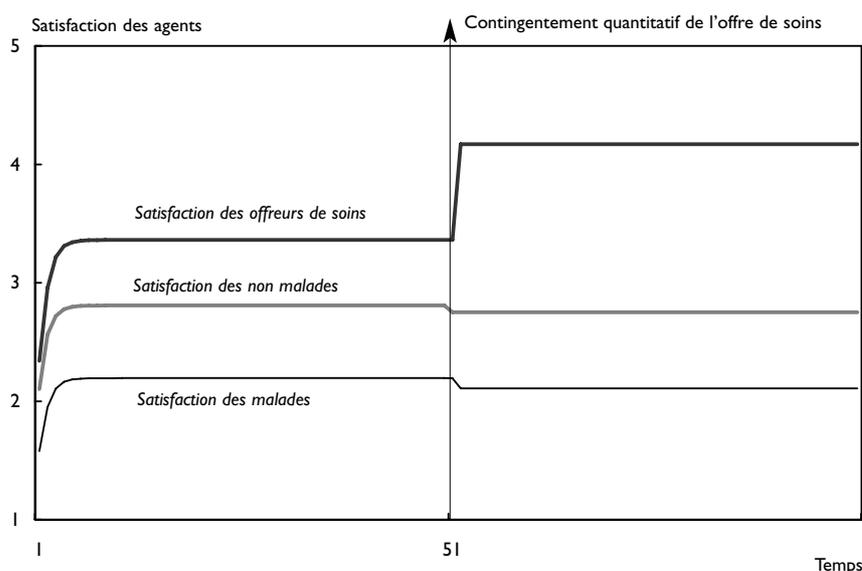
Dans ce cas, le législateur — en charge de l'assurance maladie — contrôle la quantité de soins offerts aux malades, laissant le prix s'établir selon les règles du marché.

Le premier effet d'une telle politique est l'augmentation du prix d'équilibre (les ménages choisissent de conserver inchangés leurs coefficients budgétaires : l'ajustement à une raréfaction de l'offre ne peut passer alors que par une hausse du prix des biens médicaux).

La conséquence d'une telle mesure de politique de santé est de diminuer la satisfaction des agents malades, et ce, quelle que soit la variante d'assurance maladie retenue (ticket modérateur fort ou faible, voire paiement forfaitaire) : ils dépensent autant, mais les quantités obtenues sont moindres. Les agents non-malades sont eux aussi affectés négativement : le contingentement devrait *a priori* rendre moins coûteuse l'assurance ; mais la hausse du prix qu'il génère se répercute négativement sur leur satisfaction. Selon nos hypothèses, le contingentement a pour effet une dégradation nette de l'état de santé moyen de la population des malades et des non-malades : les soins utiles consommés sont au total moins nombreux.

Seuls les offreurs de soins tirent profit de la politique de contingentement, du fait de la hausse du prix du bien santé. Ce profit n'est pas monétaire : en valeur, les agents malades dépensent la même somme pour la santé. Le secteur médical conserve donc le même revenu, mais à effort plus faible (le temps de

I2. Évolution de la satisfaction des agents à la suite d'une mesure de contingentement quantitatif



Source : Calculs des auteurs.

loisir des offreurs de soins augmente). Comme le prouve le graphique 10, paradoxalement, les offreurs de soins sont les grands bénéficiaires de politiques de contingentement quantitatif. Nous allons voir, en outre, que le rapport α peut se voir inopportunistement modifié du fait du contingentement quantitatif de l'offre de biens médicaux¹⁶ (ce qui n'était pas le cas dans un contingentement de type *numerus clausus*).

Contingentement quantitatif et contrôle du prix

Sachant ce qui précède, le gestionnaire de l'assurance maladie pourrait être tenté de croiser un contingentement quantitatif avec un contrôle des prix, de manière à éviter les effets néfastes, déjà identifiés, d'une hausse indésirable du prix des soins.

Le contingentement quantitatif conjugué à un contrôle des prix est la méthode privilégiée en France pour contenir l'évolution des dépenses de santé. C'est le cas des hôpitaux français ou, de manière partielle, le cas du secteur ambulatoire. L'autorité de tutelle fixe discrétionnairement prix et quantités produites (via le nombre de lits, par exemple) ; la consommation médicale en volume chute, mais le prix n'augmente pas, au contraire (le prix fixé par le législateur sera par hypothèse inférieur au prix de marché), ce qui permet aux

16. Cf. annexe 4 « Réaction des offreurs de soins à une politique de contingentement de l'offre médicale ».

malades de réallouer une fraction de leur revenu à la consommation et à l'épargne. On obtient donc une consommation médicale plus faible en volume, et en valeur. Les effets d'une telle modification des comportements sont multiples :

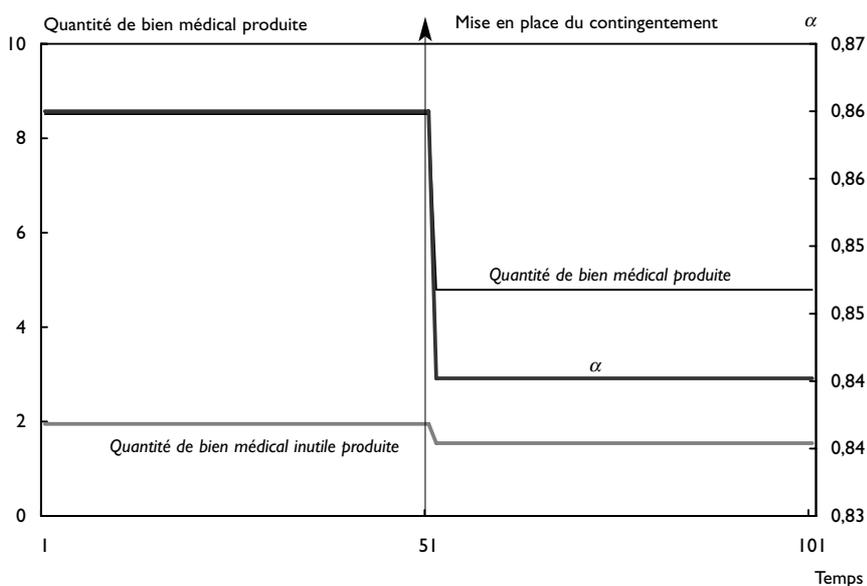
— la propension à épargner des malades s'élève : les malades, rationnés, ne tiennent plus compte de leur goût pour la santé au moment de leur arbitrage intertemporel.

— il se produit des ré-allocations de revenu, du fait de la libération de pouvoir d'achat des ménages, elle-même consécutive à la double réduction, en volume et en valeur, des dépenses de santé.

Compte tenu de la « surconsommation » de soins de santé préexistante, on peut donc s'attendre à des effets de bien-être positifs ; *a priori* la distorsion est mieux contrôlée.

Néanmoins, une autre évolution va entrer en contradiction avec ces effets positifs : sauf cas très particuliers, *la quantité de bien de santé utile, produite et consommée dans l'économie, va décroître sous l'effet de la mesure de contingentement*, le contingentement — de nature macroéconomique — n'étant, en aucun cas, à même de différencier les deux types de biens et d'adresser la restriction exclusivement aux biens inutiles¹⁷.

13. Effet du contingentement sur le rapport



Source : Calculs des auteurs.

17. On en trouvera la démonstration mathématique dans l'annexe 4.

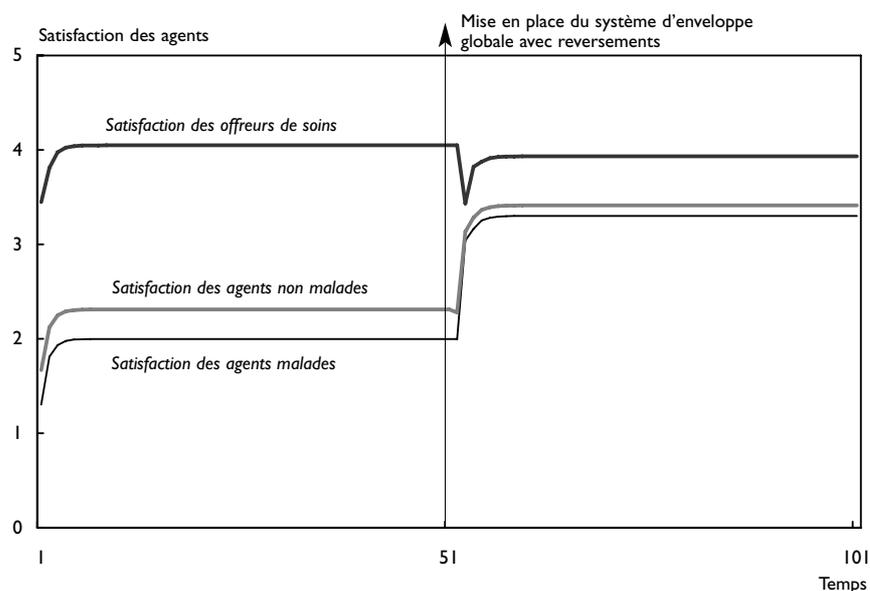
On observe que le contingentement diminue les deux offres de biens : on ne pourra pas affirmer, donc, qu'il s'adresse uniquement à la partie *inutile* de la dépense de santé. De plus, le contingentement peut avoir pour conséquence de modifier le rapport α (utile/inutile) ; ainsi, l'évolution des produits de santé utiles (h_1) peut être : i) soit strictement proportionnelle au contingentement (10 % de contingentement, 10 % de réduction de soins utiles) ; ii) soit moins que proportionnelle, iii) soit plus que proportionnelle. Dans ce dernier cas, le contingentement aggrave la distorsion (mesurée en pourcentage), ce qui est, bien sûr, très gênant pour qui se soucie de l'état de santé des ménages. (Voir annexe 4).

Enveloppe globale avec reversements

Le législateur fixe un « plafond » de dépenses de santé agrégées. S'il est avéré que la dépense médicale nationale dépasse ce seuil, il modifie les lettres-clés du praticien en vue de contenir *a posteriori* les dépenses de santé dans l'enveloppe prévue. Il s'opère donc un transfert monétaire des offreurs de soins vers les pouvoirs publics (en l'occurrence l'assurance maladie), appelé « reversement ». L'assurance maladie peut donc, du fait de ces recettes nouvelles, diminuer les taux de cotisation des assurés, tout en respectant l'objectif d'équilibre de ses comptes.

La caractéristique de cette variante de politique de santé est qu'elle laisse place à un équilibre de marché, à la différence des deux précédentes. En effet, les ménages consomment la quantité qu'ils désirent de bien santé, les reversements éventuels s'effectuant *ex-post*.

14. Évolution de la satisfaction des agents à la suite de la mise en place d'un système d'enveloppe globale avec reversements



Source : Calculs des auteurs.

Le graphique 14 permet de montrer la chute de la satisfaction des offreurs de soins, leurs revenus étant plafonnés par le législateur à un niveau inférieur au niveau prédéterminé. Parallèlement, on note une augmentation de la satisfaction des malades, et des non-malades : les versements introduits à la période 50 permettent une baisse de la cotisation d'assurance, un surcroît de revenu pour les ménages, et donc une satisfaction en hausse. De plus, la contrainte d'enveloppe permet, à la différence des politiques de contingentement, de ne pas contraindre la production de bien santé, et donc de ne pas limiter la production de bien santé utile. Il n'y a pas ici de contraction de la quantité de biens produits, pas plus que de leur « qualité ». Ceci est une conséquence de l'inélasticité de l'offre médicale au prix. (A *contrario*, si les offreurs de soins effectuaient un calcul économique intégrant la révision du prix du bien santé dans leur fonction d'offre, les versements pousseraient les offreurs de soins à diminuer les quantités offertes ; il existerait bien, ainsi, une contrainte, réellement « incitative », et/mais, en même temps, une réduction de l'offre nationale de santé, sans doute pénalisante pour l'état de santé et le bien-être¹⁸).

Il s'opère aussi un transfert vers les non-malades : les ménages dans leur ensemble paient moins, car les versements constituent une nouvelle ressource de l'assurance maladie. La charge de l'assurance est répartie sur une population plus grande, au bénéfice principal des « non-malades » (voir plus loin, les « malades » n'en bénéficient pas autant car ils sont confrontés plus largement à une hausse du prix « notionnel » — i.e. le prix pratiqué avant la révision des tarifs ; nous développerons cet effet).

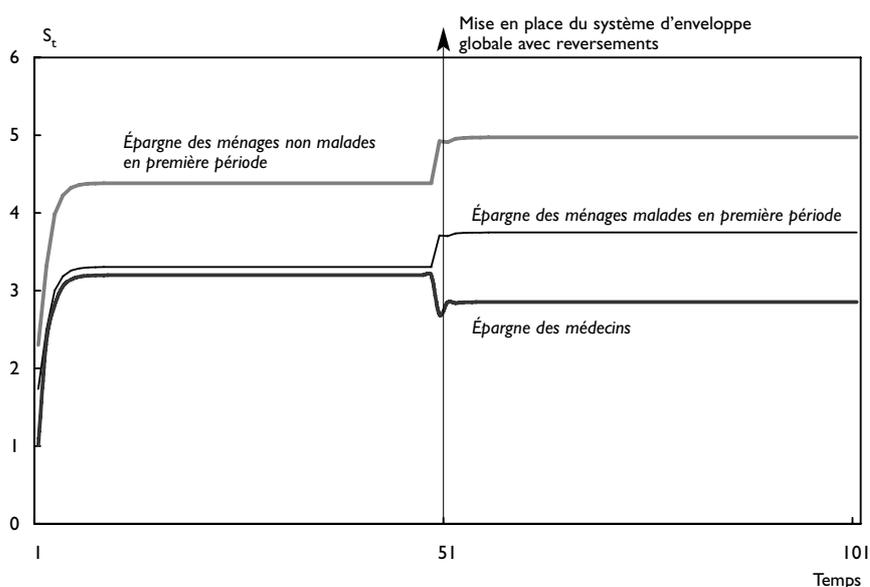
En ce qui concerne les comportements d'accumulation, on observe (graphique 15) que la mise en place d'une enveloppe globale des dépenses de santé conduit les personnels médicaux à réduire le niveau de leur épargne (leur taux d'épargne reste constant et leurs revenus chutent). Quant aux ménages, leur épargne progresse, de manière symétrique, du fait du transfert positif de revenu. Au niveau macroéconomique, à la condition que les taux d'épargne des offreurs de soins soient inférieurs aux taux d'épargne des ménages, on note une progression de l'épargne agrégée, qui a pour conséquence une accumulation supplémentaire de capital, et en définitive une augmentation du produit d'équilibre.

La politique de santé qui consiste à fixer une enveloppe globale, et à infliger des sanctions pécuniaires au secteur de santé semble donc atteindre certains de ses objectifs : la satisfaction de la population hors secteur médical augmente, du fait principalement de la baisse des cotisations d'assurance¹⁹ ; le taux

18. Qu'en est-il en réalité ? Le secteur médical français réagit-il aux révisions de tarifs ? On peut le penser à voir les grèves de médecins libéraux intervenant en fin d'année budgétaire (quand il est possible de calculer le dépassement), manière pour eux de réduire leur activité *in extremis*. On peut aussi penser, à l'inverse, que la nature concurrentielle du secteur empêche la coordination, on verra ainsi plus loin que l'hypothèse d'inélasticité de l'offre aux versements est cohérente avec l'idée — fournie par la parabole du passager clandestin — selon laquelle les calculs individuels ne tiennent pas compte (n'ont pas de raison de tenir compte) des sanctions collectives.

19. Cf. annexe 3 « Calcul de l'équilibre de l'assurance maladie, en présence de co-paiements, de paiements forfaitaires et de versements », pour le lien entre versements des médecins et baisse des cotisations des assurés.

15. Effet d'une politique d'enveloppe globale sur l'épargne des agents



Source : Calculs des auteurs.

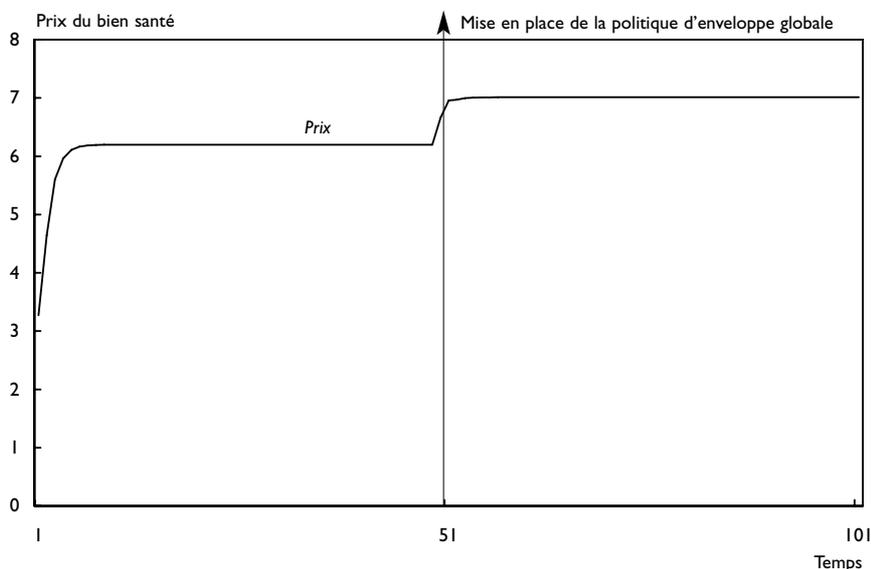
d'épargne macroéconomique augmente — à condition que les propensions à épargner des ménages soient bien supérieures à la propension à épargner du secteur de santé. Si l'on veut, dans ces circonstances (celle notamment d'une réelle inélasticité de l'offre médicale à la révision des tarifs), la politique d'enveloppe globale se présente comme une réponse adéquate à la surconsommation médicale de l'économie.

Néanmoins, comme le montre le graphique 16, un effet induit de la politique d'enveloppe globale s'avère négatif : à offre inchangée, on constate que le prix (« notionnel ») des transactions de santé augmente, ce qui aura pour conséquence de limiter la soi-disant neutralité (marchande) de la politique menée. Il faut, pour expliquer ce phénomène, avoir en tête le bouclage macroéconomique du modèle : *après transferts*, la politique d'enveloppe globale va augmenter les revenus des agents non médicaux ; puis, ce revenu étant consommé, on assiste à une hausse de la demande de santé dans l'économie. Il s'agit là, typiquement, d'un « effet richesse » (l'assurance maladie et, par là même, les assurés s'enrichissent grâce au transfert).

Notons en outre que, dans le modèle, l'offre médicale ne change pas à la suite d'une politique d'enveloppe globale, l'offre est inélastique au prix. De fait, le modèle reproduit une réalité *microéconomique*, caractéristique de ce type de politiques *macroéconomiques*, insuffisamment fondées sur les incitations individuelles : les offreurs de soins se trouvent dans la situation classique du « passager clandestin » : s'ils réduisent seuls leur offre de biens médicaux face à une politique d'enveloppe, ils devront contribuer tout de même aux versements fixés par l'assurance maladie (puisque les versements sont établis

au niveau de chaque spécialité, ou unité de soins, et non en fonction de l'effort de limitation de l'offre, individuelle, de chaque médecin). L'offre médicale *individuelle* n'a donc aucune raison de baisser en réponse à une révision *collective* des tarifs. Ainsi, dans ce jeu d'hypothèse, la politique d'enveloppe aura pour conséquence de maintenir inchangées les quantités de soins produites et consommées...

16. Effet d'une politique d'enveloppe globale sur le prix du bien santé

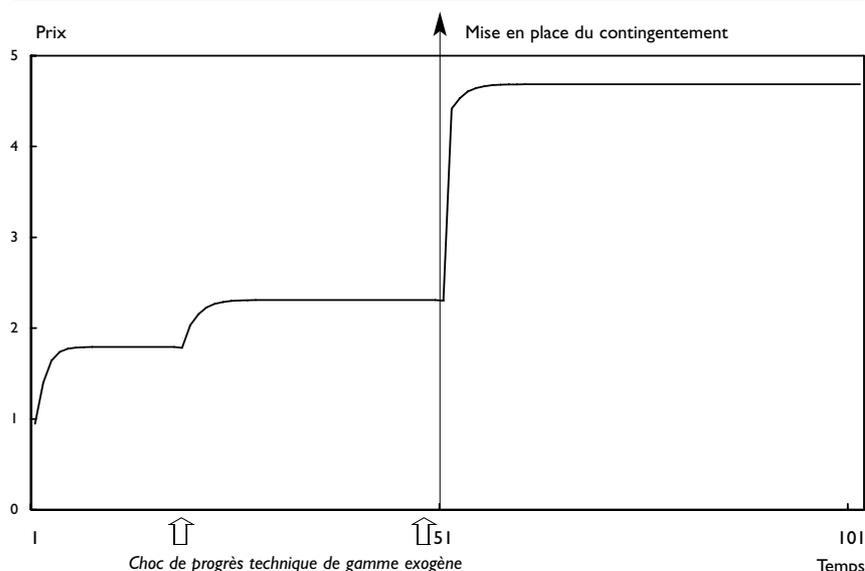


Source : Calculs des auteurs.

Politique d'enveloppe avec progrès technique de gamme dans le secteur médical

Cette section a pour but d'étudier les effets d'une politique d'enveloppe globale en présence de progrès technique de gamme dans le secteur médical. On croise ces deux phénomènes, car, dans certains discours, la politique d'enveloppe est présentée comme la réponse appropriée à l'extension de la gamme des soins. On l'a vu, l'extension de la gamme provoque une déformation des coefficients budgétaires en faveur de la santé, se reportant sur l'équilibre des comptes de l'assurance santé. On sait aussi qu'elle peut tendre à biaiser la répartition du bien-être, des ménages vers les offreurs de soins (lorsque l'extension de la gamme est cosmétique et qu'il y a un biais systématique vers « l'inutile »). La politique d'enveloppe pourrait donc intervenir pour compenser ces différentes évolutions, jugées indésirables ; en effet, la politique d'enveloppe a pour mérite d'assurer de nouveaux financements à l'assurance, tout en retransférant du revenu (et donc *a priori* du bien-être) du secteur médical vers les ménages.

17. Effet d'une politique de contingentement sur le prix du bien santé



Source : Calculs des auteurs.

On obtient à nouveau, dans le contexte d'un progrès technique de gamme, le résultat mis en évidence à la section précédente, à savoir « l'effet richesse » d'un reversement : le corollaire indésirable d'une politique d'enveloppe globale est que, loin de compenser l'effet du progrès technique sur le prix, elle amplifie la hausse du prix « notionnel » (sur le marché, à offre constante, la demande augmente).

Ce corollaire ne peut toutefois intervenir que dans la mesure où le tarif médical n'est pas conventionné dans son intégralité. Techniquement, notre modèle instaure un clé de révision *ex-ante* sur un tarif (notionnel) libre *ex-post* (on fait porter une clé de révision τ sur un prix p libre, le tarif *ex-post* est $(1-\tau)p$). Or, en France, ce n'est pas le cas : pour chaque activité médicale conventionnée, l'intégralité du prix est contrôlé, et le prix p^r est fixé une fois pour toutes, sans degré de liberté. Le raisonnement présenté ci-dessus reprend néanmoins de la validité si l'on réfléchit en termes de quantités, et si l'on admet une certaine substitution entre les activités de soins : d'une part, « l'effet richesse » du reversement reste valide pour expliquer une hausse *en volume* de la demande de soins face aux révisions de tarifs ; d'autre part, l'effet richesse peut se reporter sur les prix des autres sous-secteurs de la santé (par exemple, les tarifs non conventionnés). En contrôlant d'une main la dépense de santé dans un sous-secteur de santé, on suscite donc, en même temps, les incitations pour un laisser-aller dans les autres sous-secteurs²⁰.

20. Empiriquement, il semble bien que « l'enrichissement » de l'assurance maladie, et aussi celui des ménages, déclenche une réaction à la hausse de la consommation médicale. Le premier semestre 2000 a montré une très forte sensibilité de la dépense maladie (soins de ville) à la bonne conjoncture et aux cagnottes.

Conclusion

Ce modèle macroéconomique nous a permis d'étudier les conséquences de quelques variantes des politiques de santé sur la satisfaction des agents, la quantité, et le prix, du bien santé échangé. Le modèle nous a aussi permis d'analyser plusieurs phénomènes : progrès technique de gamme et/ou de productivité, chocs exogènes de demande ou d'offre médicale, etc.. Il convient, néanmoins, d'attirer l'attention du lecteur sur le fait que cet article présente les résultats obtenus sur un modèle *simulé*, qui n'a pas fait l'objet de calibrage le mettant directement en rapport avec le réel (en France ou ailleurs). Ce travail resterait à faire, et nous aiderait à porter un diagnostic plus fin sur la robustesse réelle du modèle. A court terme, cet effort semble problématique, du fait du nombre élevé de paramètres « psychologiques » à estimer. Cela n'obère pas pour autant les apports du modèle, qui permet d'étudier *in abstracto* la réaction du secteur médical et des agents économiques à diverses variations de leur environnement naturel ou institutionnel. Reprenons les principaux résultats :

L'intégration du progrès technique de gamme dans le modèle propose une voie nouvelle pour expliquer la déformation progressive des coefficients budgétaires des ménages vers la dépense santé (chose que ne peut pas faire un progrès technique classique, dit de productivité, dans ce même secteur). Elle permet ainsi d'enrichir la compréhension du phénomène : si les consommations médicales sont souvent qualifiées de « bien de consommation supérieur », puisqu'elles possèdent empiriquement une élasticité-revenu supérieures à 1, c'est peut-être aussi, pour une part, la conséquence d'un « effet de gamme » plus fort dans le secteur de santé que dans les autres secteurs de l'économie. Au delà de l'élasticité-revenu supérieure à 1 et du simple constat économétrique, le goût *pour la variété des soins*²¹ deviendrait un facteur explicatif de l'évolution de la part de santé dans le PIB, explication jusqu'ici assez peu exploitée (elle vient de la *théorie de la croissance endogène*, et, elle a été appliquée surtout au versant production²² plutôt qu'au versant consommation). Par ailleurs, on a pu noter que ce type de progrès technique profite *a priori* aussi bien aux offreurs de soins qu'aux ménages.

On a pu rendre compte des modifications introduites dans le système assurantiel français (hausse du ticket modérateur...), et envisager l'effet du passage à un remboursement forfaitaire. En pure technique assurantielle, il est clair que le remboursement forfaitaire domine le co-paiement, car il évite l'effet de myopie sur le coût des soins. Dans l'assurance santé, il nécessiterait néanmoins une information, très complète, sur les différentes maladies et sur leurs coûts (explicite et implicite). Il nécessite aussi un contrôle étroit des diagnostics.

Nos hypothèse sur la « demande induite » reviennent à formaliser, sans doute imparfaitement, les aspects *informationnels* et *qualitatifs* de la transaction de santé. Travail utile et travail inutile, associés à des productions utiles et inutiles, sont des opérateurs d'une réflexion sur la « qualité » d'une unité de bien

21. i.e. l'absence de décroissance de l'utilité marginale face à l'extension de la gamme.

22. Dans le modèle de Romer (1991), c'est la variété des inputs qui permet de contre-carrer la loi des rendements décroissants.

fourni et le rapport a une mesure de cette qualité (α = pourcentage de « bien utile » incorporé dans le total de la transaction de soins). Le premier résultat de cette modélisation est que, face à une diminution de α , les ménages sont parfois contraints de consommer plus encore pour réparer cette dégradation ; ce qui contribue, à nouveau, à faire croître la dépense. Le second résultat réside dans l'observation de la détermination du rapport α : la qualité du soin unitaire varie (elle résulte d'un calcul), elle est fonction, notamment, des variations de l'environnement institutionnel des offreurs de soins. L'intérêt d'une telle démarche est, bien sûr, de boucler ces réflexions avec les politiques mises en œuvre pour contrôler la surconsommation.

Les politiques de « contingentement pur » semblent peu à même de répondre de manière efficiente à la hausse de la consommation médicale, car elles induisent des effets prix et confèrent des rentes de monopoles. Couplées avec un contrôle des prix, elles sembleraient plus efficaces. Néanmoins, l'effet recherché n'est pas atteint, dans la mesure où le contingentement — de nature macroéconomique — n'est, en aucun cas, à même de différencier les deux types de soins, utiles et inutiles, et d'adresser la restriction aux biens inutiles exclusivement (parfois même, le contingentement a pour effet de dégrader le contenu en bien utile incorporé dans la transaction de santé). Enfin, bien sûr, un contingentement quantitatif qui voudrait améliorer le bien-être se doit de suivre l'évolution du progrès technique. Dans le cas d'un progrès technique de gamme, on a vu que la déformation à la hausse de la part du secteur de santé dans le PIB pouvait être « optimale », i.e., qu'elle répondait à des logiques de bien-être (lorsque le progrès technique de gamme ne biaise pas le rapport utile/inutile) ; il n'y a donc aucune raison *logique* pour figer cette part dans le temps comme le législateur français est parfois tenté de le faire.

Enfin, la politique d'enveloppe globale avec reversements peut apparaître plus avantageuse car elle laisse place à un équilibre de marché (les biens « utiles » ne sont pas *a priori* restreints). Elle est pertinente pour qui veut retransférer une part du revenu, et du bien-être, du secteur de santé vers les ménages. Elle ne fonctionne néanmoins pas intégralement, car elle crée un effet en retour sur la demande de soins : « l'effet richesse » de l'assurance maladie, qui accroît le prix de marché et affecte le bien-être. En outre, elle néglige les aspects incitatifs microéconomiques de l'offre : les offreurs peuvent jouer des stratégies de « passagers clandestins » de sorte que, en l'absence d'une coordination du groupe, ils seront insensibles aux révisions de tarifs : les quantités offertes globalement restent insensibles au signal de prix qu'avait pu vouloir donner le législateur. *A contrario*, si les offreurs se coordonnent (ils anticipent la révision de tarifs et ajustent leurs offres en conséquences – cas que notre modèle n'a pas pris en compte), la politique d'enveloppe s'analyse comme un contingentement classique, retrouvant par là même les avantages mais, surtout, les inconvénients de la méthode : il y a alors, de fait, un rationnement des soins offerts à la société.

Aucune de ces variantes ne semble donc entièrement satisfaisante, mais peut-être cela tient-il aux particularités du secteur médical, peut-il (et doit-il ?) être considéré comme devant être soumis à une maîtrise comptable *macroéconomique* sachant que les logiques sous-jacentes s'avèrent essentiellement *microéconomiques* ? La conséquence à tirer de cette question, quelque peu insidieuse, est claire : la bonne réponse, donc, est microéconomique. Une piste

concrète réside dans ce qu'il est convenu d'appeler les « schémas de paiement », modalités de réponse réellement *microéconomique* à une distorsion du même type. Il faut bien voir alors, qu'en France, ce seront les règles actuelles de rémunérations des actes médicaux qui seront à remettre en cause, au profit par exemple d'une logique *d'abonnement forfaitaire* comme pratiquée dans d'autres pays. L'honnêteté exige de signaler que, de ce point de vue, le traitement actuel du modèle ne dit rien (comment la qualité des soins réagit-elle aux schémas de paiement ? C'est une question à explorer). Il y a donc encore du travail à faire, tant pour l'économiste que pour le législateur.

Références bibliographiques

- BATIFOULIER P., V. TOUZÉ, 2000 : *La protection sociale*, Dunod, Paris.
- DIAMOND P., 1965 : « National Debt in a Neoclassical Growth Model », *American Economic Review* 55 (Dec.), p. 1126-1150.
- EVANS, R., 1974 : « Supplier-Induced Demand : some Empirical Evidence and Implications », *The Economics of Health and Medical Care*, p. 163-173.
- PHELPS C., 1995 : *Les fondements de l'économie de la santé*, Publi-Union.
- ROCHAIX L., C. JACOBZONE, 1997, « L'hypothèse de demande induite : un bilan économique », *Economie et prévision*, n°129-130.
- ROMER, P., 1991 : « Progrès technique endogène », *Annales d'Eco et Stat*, n°22.
- TOUZÉ V., B. VENTELOU, 1999 : « Politiques de santé dans un modèle macro-économique », Document de travail – OFCE, disponible auprès des auteurs.
- TOUZÉ V., B. VENTELOU, 2000 : « The Macroeconomics of Health Care », OFCE Working Paper, disponible auprès des auteurs.
- WEISBROD B., 1991 : « The Health Care Quadrilemma : en essay on technical change, insurance, quality of care and cost containment », *Journal of Economic Literature*, vol. XXIX, pp. 523-52.

ANNEXE I

Index des notations utilisées dans le modèle

X^+, X, X^-	Valeur de X en t+1, en t, et en t-1
X^*	X fait l'objet d'un arbitrage de l'agent
X'	X concerne les agents non malades en première période
X_m	X concerne les offreurs de soins
X_y	X concerne les agents en activité
X_o	X concerne les agents retraités
<hr/>	
C	Consommation de l'agent lors de sa période d'activité
Z	Consommation de l'agent lors de sa période d'inactivité
S	Épargne
K	Stock de capital agrégé
h_p , avec $i=1,2$	Activité de soins utiles (resp. inutiles)
l_p , avec $i=1,2$	Temps de travail consacré aux soins utiles (resp. inutiles)
d_y, d'_y	État de santé des malades (resp. non malades) en première période
$I(t)$	Nombre de biens et services composant le secteur santé
I_{t+1}/I_t	Progrès technique de gamme
<hr/>	
p	Niveau général de prix de la santé
w	Salaire
<hr/>	
β	Facteur d'escompte du temps
δ_p , avec $i=1,2$	Coût subjectif associé à l'effort de travail médical utile (resp. inutile)
α	Proportion des biens médicaux utiles sur l'ensemble des biens médicaux
π	Prime d'assurance maladie
λ	Taux de remboursement des soins
J_1	Remboursement de l'assurance maladie
J_2	Païement forfaitaire de l'assurance maladie
Ψ	Proportion des ménages malades en première période
$\gamma, \gamma^\#$	Goût pour la santé des agents actifs (resp. des agents inactifs)
<hr/>	
q	Proportion des offreurs de soins dans la population totale (1+q)
h^c	Offre de soins contingentée
E_y	Enveloppe de dépense médicale
R^y	Valeur des versements par les offreurs de soins en cas de dépassement de l'enveloppe globale
τ	Taux de versements des offreurs de soins à l'assurance maladie

ANNEXE II

Choix de modélisation

Cette annexe présente les hypothèses de modélisation.

Les calculs détaillés des comportements optimisés des différents agents sont disponibles dans l'article Vincent Touzé et Bruno Ventelou (2000) : « Politiques de santé dans un modèle macroéconomique », *Document de travail OFCE*, disponible auprès des auteurs.

Le bloc demande du modèle

Les agents « malades » (en première période)

Ce type de ménage représente une fraction ψ de la population des ménages qui travaillent à la production du bien de consommation agrégé (on fait l'hypothèse que la maladie ne diminue pas la productivité des agents...).

Soit une fonction de consommation des agents de la forme (contraintes budgétaires intertemporelles :

$$\begin{cases} C = w^N - ph - S \\ Z^+ = R^+S - p^+h^+ \end{cases}$$

avec h quantité de bien médical consommée, w^N revenu net des cotisations et remboursements effectués, et $R^+ = (1+r^+)$, r taux d'intérêt.

Soit une fonction d'utilité des agents malades de la forme :

$$U(C, Z^+, h_y, h_o^+) = \ln C + \gamma \sum_{c=1}^{I(t)} \ln(h_{y,c} - h_{y,c,2}) + \beta \left(\ln Z^+ + \gamma^\# \sum_{c=1}^{I(t+1)} \ln(h_{o,c}^+ - h_{o,c,2}^+) \right)$$

En réduisant l'hétérogénéité des soins de santé et en définissant \bar{h} comme le bien médical moyen, on obtient :

$$U(C, Z^+, h_y, h_o^+) = \ln C + \gamma I \ln \bar{h}_y + \beta (\ln Z^+ + \gamma^\# I^+ \ln \bar{h}_o^+),$$

avec γ goût pour la santé des agents ne travaillant pas dans le secteur médical.

Les agents « non malades » (en première période)

Ce type de ménage représente une fraction $(1-\psi)$ de la population des ménages qui travaillent à la production du bien de consommation agrégé. Ces

agents ont les mêmes comportements que les agents malades en première période, avec une spécificité, traduite par la contrainte $\gamma = 0$, qui implique que les agents « non malades » ne consomment pas de bien santé en première période.

Les offreurs de soins

Par hypothèse, les offreurs de soins ne consomment pas de bien santé, à aucune période de leur vie. Ils arbitrent donc entre consommation présente et consommation future, par le biais de l'épargne. Ils maximisent une fonction d'utilité de la forme :

$$U(C_m, Z_m^+) = \ln C_m + \beta_m \ln Z_m^+ - \delta_1 l_1 - \delta_2 l_2$$

$$\text{s.c. } C_m + \frac{Z_m^+}{R^+} = p h(l_1^*, l_2^*)$$

Le bloc offre du modèle

La production du bien de consommation

Le secteur productif est modélisé par une fonction de production Cobb-Douglas classique, avec progrès technique de productivité au sens de Hicks.

$$F = A K^a L^{1-a}, \text{ avec } a \in [0,1].$$

Le progrès technique A_t est exogène, et dA/dt , avec $\rho \in \mathbb{R}_+$.

On suppose que la taille de la population active est constante ($L_t=1$). On obtient donc, en simplifiant :

$$F = A K^a$$

Les facteurs de production sont rémunérés à leur productivité marginale.

$$\begin{cases} w = F'_K(K, L) = (1-a)A .K^a \\ r = F'_L(K, L) = aA .K^{a-1} \end{cases}$$

L'accumulation du capital est fonction de l'épargne de la période précédente :

$$K^+ = S_{TOT,t}, \text{ avec } S_{TOT,t} = S_t + S_t' + S_{m,t}. \text{ On a donc :}$$

$$K^+ = \psi S + (1-\psi)S' + qS_m$$

La production des biens médicaux

Une part $q/(1+q)$ de la population travaille dans le secteur de la santé. On fait l'hypothèse que les activités médicales sont au nombre de deux, l'une utile,

l'autre inutile. On note h_1 l'activité de soins utile (et l_1 le temps de travail consacré à cette activité), h_2 l'activité inutile (resp. l_2). On a :

$$\bar{h}_i^s = B_i \left(\frac{l_i}{I} \right)^{b_i} \text{ pour } i=1, 2.$$

$B_{i,t}$ est le progrès technique de productivité, qui croît à un taux exogène et stable.

Par agrégation des biens médicaux et prise en compte du progrès technique de gamme exogène I_t , l'offre totale de biens médicaux est :

$$h^s = \sum_{c=1}^{I(t)} h_{1,c}^s (l_{1,c}) + h_{2,c}^s (l_{2,c}) = I \left(\bar{h}_1^s \left(\frac{l_1}{I} \right) + \bar{h}_2^s \left(\frac{l_2}{I} \right) \right)$$

On pose (au plus simple) :

$$\bar{h}_1^s = B_1 \left(\frac{l_1}{I} \right)^{b_1}, \quad \bar{h}_2^s = B_2 \left(\frac{l_2}{I} \right)^{b_2}$$

Bouclage du modèle, situation à l'équilibre

Pour $h_t^s = h_t^d$, on obtient le prix d'équilibre :

$$p^* = \frac{(\psi Z^* + (1-\psi)Z^*)\gamma^\# + \frac{\gamma}{1-\lambda} \psi C^*}{\alpha q \left(\bar{h}_1 \left(\frac{l_1}{I} \right) + \bar{h}_2 \left(\frac{l_2}{I} \right) \right)}$$

ANNEXE III

Calcul de l'équilibre de l'assurance maladie (en présence de co-paiements, de paiements forfaitaires et de reversements)

On a : $p_t h_{y,t} = X(w_t - \pi_t + J_{2,t})$, avec X coefficient budgétaire de la consommation de bien santé (pour les ménages malades en première période). Par définition :

$$\begin{cases} \pi_t = \psi (J_{t,1} + J_{t,2}) - R_t^v \\ J_{t,1} = \lambda_1 X I_t (w_t - \pi_t + J_{t,2}) \\ J_{t,2} = \lambda_2 X I_t (w_t - \pi_t + J_{t,2}) \\ R_t^v = \psi X I_t (w_t - \pi_t + J_{t,2}) - E_{t,y} \end{cases}$$

Après résolution du système, on obtient en définitive :

$$\begin{cases} J_{t,1} = \frac{X \lambda_1 (w_t - E_{y,t})}{\frac{1}{I_t} + \psi X (\lambda_1 + \lambda_2) - \psi X - \lambda_2 X} \\ J_{t,2} = \frac{\lambda_2}{\lambda_1} J_{t,1} \\ R_t^v = \psi \frac{J_{t,1}}{\lambda_1} - E_{t,y} \end{cases}$$

Les reversements effectués par les offreurs de soins à l'assurance maladie viennent en déduction des cotisations payées par les assurés. Les cotisations des assurés ne sont donc plus calculées par rapport aux dépenses de l'assurance maladie, mais en fonction du solde « dépenses de l'assurance maladie–reversements ».

ANNEXE IV

La réaction des offreurs de soins à une politique de contingentement et au progrès technique de gamme : les biais sur l'arbitrage interne l_1/l_2

L'hypothèse implicite retenue par tous les défenseurs d'une politique de contingentement est que celle-ci n'affecte pas la production de biens médicaux utiles, mais réduit uniquement la « demande induite ». Nous montrons que les deux catégories de biens sont contraintes. Le contingentement agit donc aussi sur la consommation de biens médicaux utiles, parfois même dans un rapport plus que proportionnel. Au passage, on traite aussi des biais introduits par le progrès technique de gamme.

Cas où le contingentement n'est pas effectif : l'effet du progrès technique

Les comportements des offreurs de soins sont définis par :

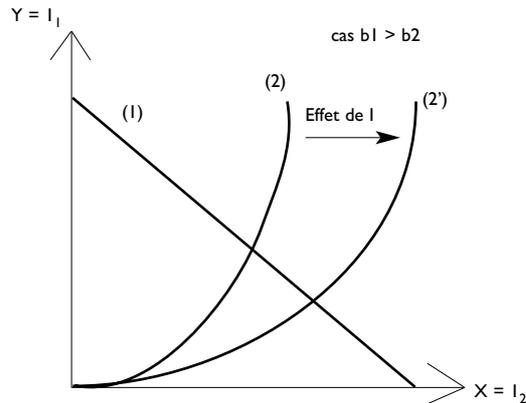
$$\begin{cases} \frac{\bar{h}'_1}{\bar{h}} = \delta_1 \\ \bar{h}'_1 \left(\frac{l_1}{I} \right) = \frac{\delta_1}{\delta_2} \bar{h}'_2 \left(\frac{l_2}{I} \right) \end{cases}$$

Ce qui équivaut à :

$$\begin{cases} l_1 = \frac{b_1}{\delta_1} \left(\frac{1 + \beta_m}{\beta_m} \right) - \frac{b_1 \delta_2}{b_2 \delta_1} l_2 & \text{Équation (1) du graphique} \\ \frac{B_1 b_1}{B_2 b_2} \left(\frac{1}{I} \right)^{b_1 - b_2} \frac{l_1^{b_1 - 1}}{l_2^{b_2 - 1}} = \frac{\delta_1}{\delta_2} & \text{Équation (2) du graphique} \end{cases}$$

On peut représenter les deux équations sur le plan (l_1, l_2) :

Choix interne l_1/l_2



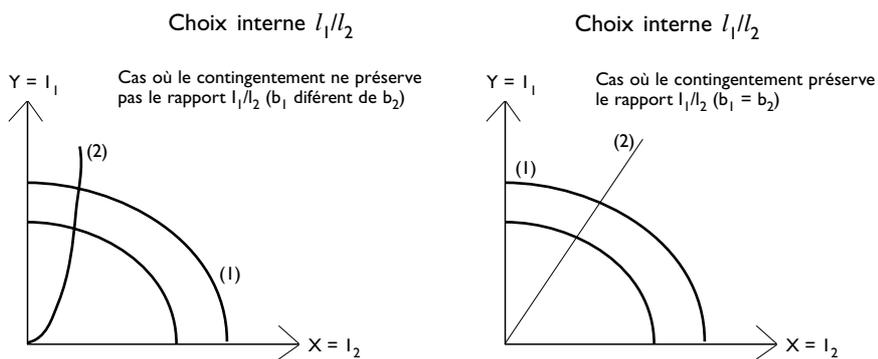
Une simple exercice de statique comparative permet de montrer dans le cas $b_1 > b_2$ que le progrès technique biaise systématiquement l'arbitrage l_1/l_2 en faveur de l_2 . L'inégalité $b_1 > b_2$ établit en effet à une réactivité plus forte du secteur n° 1 aux variations de l'environnement. Devant l'augmentation du nombre de bien à produire, les offreurs de soins révisent à la baisse leur temps de travail alloué à chaque soin unitaire, mais ils ont tendances à le faire moins vite sur le secteur n° 2, qui prend alors de l'importance.

Cas où le contingentement est effectif

Cette fois, pour h^c valeur de la contrainte, les comportements des offreurs de soins sont définis par :

$$\left\{ \begin{array}{l} \bar{h}^c = \bar{h}_1 \left(\frac{l_1}{I} \right) + \bar{h}_2 \left(\frac{l_2}{I} \right) \quad \text{Équation (1) du graphique} \\ \bar{h}_1 \left(\frac{l_1}{I} \right) = \frac{\delta_1}{\delta_2} \bar{h}_2 \left(\frac{l_2}{I} \right) \quad \text{Équation (2) du graphique} \end{array} \right.$$

On obtient :



Le contingentement (une baisse de h^c) équivaut à un déplacement de la frontière de production (courbe (1)) vers l'origine. La modification du rapport l_1/l_2 et, surtout, le sens de cette modification (amélioration ou détérioration de la qualité moyenne des soins prodigués) dépendent de b_1 et b_2 .

On voit donc que, non seulement, le contingentement quantitatif n'améliore pas la qualité des soins produits (il n'agit pas uniquement sur la demande induite), mais peut, sous certaines conditions, la détériorer.