INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

Série des documents de travail
de la Direction des Etudes et Synthèses Économiques

G 2000 / 01

Les déterminants des dépenses de santé : une approche macroéconomique

Ronan MAHIEU *

JANVIER 2000

L'auteur remercie particulièrement F. Maurel, P. Ralle et B. Sédillot pour leurs conseils.

* Département des Etudes Economiques d'Ensemble - Division « Redistribution et Politiques Sociales » Timbre G211 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX

Département des Etudes Economiques d'Ensemble - Timbre G201 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX - France - Tél. : 33 (1) 41 17 60 68 - Fax : 33 (1) 41 17 60 45 - E-mail : d3e-dg@insee.fr.

Ces documents de travail ne reflètent pas la position de l'INSEE et n'engagent que leurs auteurs.
Working papers do not reflect the position of INSEE but only their author's views.

Résumé

Nous présentons d'abord une étude comparative sur séries temporelles des déterminants des dépenses de santé dans plusieurs pays de l'OCDE. Dans un premier temps, nous estimons un modèle unique sur six pays en nous focalisant sur les facteurs d'offre, en particulier le progrès technique médical. Les résultats suggèrent un possible rôle des mécanismes d'incitation financière appliqués aux producteurs.

Nous poursuivons alors par une approche sur données panelisées inspirée des travaux de Gerdtham (1992, 1995), en nous concentrant sur le rôle possible des incitations financières, sommairement modélisé par un trend différencié selon le type de système (contractuel, intégré ou de remboursement). Nous estimons une élasticité-revenu proche de la valeur de 0,7 retenue par Gerdtham lorsqu'il prend en compte les spécificités institutionnelles. En outre, nous trouvons que les systèmes de remboursement (dont la France est un exemple) connaîtraient, toutes choses égales par ailleurs, une croissance annuelle moyenne des dépenses de santé supérieure d'environ 0,6 point à celle des systèmes de santé de type contractuel (cas de l'Allemagne) ou intégré (cas du Royaume-Uni).

Mots-clés : dépenses de santé, systèmes de santé, incitations financières, séries temporelles

Abstract

This paper first presents a time-series analysis of factors influencing aggregate health care expenditure in some OECD countries. We focus on supply-side factors (especially the diffusion of medical technologies) and first estimate a single model on six countries. The results suggest that financial incentives may explain a part of the discrepancies observed between these countries.

We then test this hypothesis on panel data for 20 OECD countries, following Gerdtham (1992, 1995), and thus introduce a linear trend for each type of health care system : integrated (like the British NHS), based on Fee-For-Service reimbursement (like France), or based on contracts between providers and insurers (like Germany). We compute an income elasticity around 0.7 (this is Gerdtham's estimation when introducing institutional specificities), and find that the growth of health care expenditure per capita is 0.6 point higher in countries with Fee-For-Service reimbursement than in other countries.

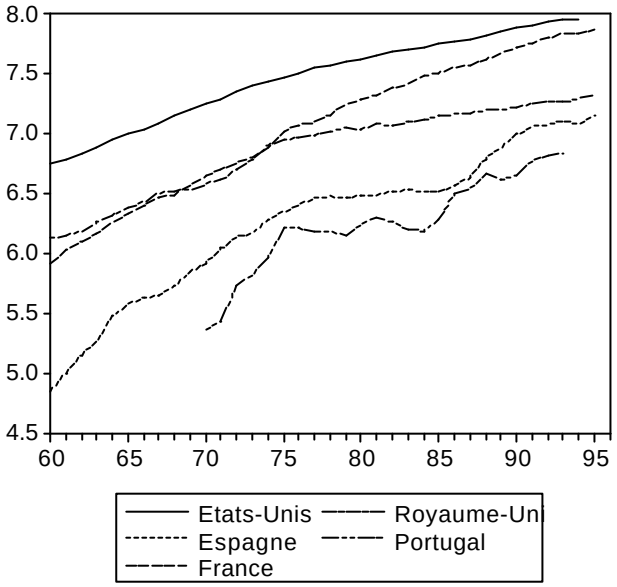
Keywords: health care expenditure, health care system, financial incentives, time series

Classification JEL : C22, D82, H51, I11

Introduction

Les dépenses de santé croissent à un rythme historiquement supérieur à celui du PIB, accréditant l'hypothèse que la santé serait l'exemple type d'un bien supérieur (Newhouse, 1977) : les dépenses de santé sont passées de 5,8 % du PIB en 1970 à 9,9 % en 1997 en France. Sur le groupe formé de l'Allemagne, des Etats-Unis, de la France, des Pays-Bas et du Royaume-Uni, le taux de croissance annuel des dépenses de santé est significativement supérieur à celui du PIB sur longue période, même si l'écart tend à diminuer : de 3,1 points dans les années 60, il est passé à 1,5 point entre 1990 et 1997 (OCDE, 1998).

Graphique 1 : Evolution du volume des dépenses de santé per capita (logarithmes), exprimées en dollars US et en parités de pouvoir d'achat, entre 1960 et 1995, pour cinq pays de l'OCDE[1].



Dans un contexte de réduction des marges de manoeuvre budgétaires, et d'un relatif échec des mesures prises jusqu'à présent pour freiner l'accroissement des dépenses de santé, il est crucial d'identifier les différents facteurs expliquant la progression de ces dernières. Il existe à ce jour des approches à la fois microéconomiques et macroéconomiques de ce sujet. Notre choix de comparer différents pays de l'OCDE les uns avec les autres nous amène à retenir une approche agrégée.

Dans toute cette étude, nous désignerons par dépenses de santé l'ensemble des dépenses à caractère médical, quelle que soit leur nature (à savoir, principalement, les dépenses d'hôpital, de médecine ambulatoire, de pharmacie et les dépenses en biens médicaux) ou leur mode de financement (dépenses socialisées, dépenses remboursées par une assurance privée ou paiements directs des ménages).

On peut schématiquement essayer de distinguer, parmi les facteurs d'évolution des dépenses de santé, ceux liés à l'offre et ceux qui relèvent de la demande de soins. Nous présentons ici, outre une revue de la littérature sur le sujet (partie 1), une étude

[1] Nous avons ici déflaté les dépenses de santé per capita par l'indice national de prix **sectoriel** (base 100 en 1990) puis converti cette somme en dollars US en utilisant le taux de change 1990 exprimé en parités de pouvoir d'achat de soins de santé. Le volume de dépenses par tête est alors, en 1993, en France, inférieur de 10,8% à ce qu'il est aux Etats-Unis. Si l'on déflate par l'indice **général** des prix à la consommation et que l'on utilise les parités de pouvoir d'achat globales, ce chiffre dépasse 45%. L'écart est donc considérable.

économétrique ciblée sur les effets d'offre : progrès technique médical et incitations financières des prestataires de soins (partie 2).

Il reste que les comparaisons internationales de dépenses de santé agrégées, comme celle que nous présentons ici, sont par nature un exercice délicat. Si la plupart des systèmes de santé sont étroitement régulés par l'Etat, les modalités de l'intervention publique et l'organisation de l'offre de soins varient fortement d'un pays à l'autre, ce qui limite la comparabilité des résultats.

1 - Expliquer les dépenses de santé : Une revue de la littérature économique

A ce jour, la plupart des travaux empiriques, notamment ceux relevant d'une approche macroéconomique, sont centrés sur les facteurs de demande.

1.1 Les travaux en coupe : effet revenu et effet prix

Faute de données longitudinales suffisantes, les premiers travaux effectués au cours des années 1970 ont adopté une approche exclusivement en coupe et ont étudié l'impact de l'effet revenu et de l'effet prix sur le volume des dépenses de santé.

1.1.1 La prépondérance de l'effet revenu

Les travaux fondateurs de Newhouse (1977), effectués en coupe sur treize pays (mais sur des années différentes par manque de données), concluent à une élasticité revenu significativement supérieure à l'unité (1,35), le PIB par tête expliquant à lui seul 92% des variations des dépenses de santé. Newhouse en tire les conclusions (discutées) que la santé est un bien supérieur et que l'achat d'une unité de soins supplémentaire contribue à améliorer davantage l'état subjectif de bien-être (*care*) que les indicateurs objectifs de morbidité ou de mortalité (*cure*). Comme l'ont souligné Parkin *et al.* (1987), ces conclusions en termes de comportements microéconomiques sont risquées lorsqu'elles s'appuient sur le seul examen de séries agrégées.

L'idée selon laquelle la santé serait un bien supérieur n'est toutefois pas sans fondements : comme l'illustre le graphique n°1, qui montre l'évolution des dépenses de santé par tête dans des pays de niveaux de vie très différents, on observe des phénomènes de rattrapage assez nets, pendant lesquels les dépenses de santé augmentent sensiblement plus vite que la richesse nationale. Ces phénomènes de rattrapage sont particulièrement sensibles pour l'Espagne et le Portugal, d'abord de 1960 à 1975 puis au cours de la deuxième partie des années 1980. En d'autres termes, l'accession d'un pays au statut d'état industriel développé entraîne un poids croissant des dépenses de santé dans la richesse nationale pendant cette phase du développement.

Un point crucial des analyses comparatives en coupe réside dans le choix du taux de change. Alors que la plupart des travaux menés à la suite de Newhouse confirment l'existence d'une élasticité revenu supérieure à l'unité, Parkin *et al*. soulignent que l'utilisation des parités de pouvoir d'achat (et non du taux de change courant, comme Newhouse) comme déflateur conduit à estimer une élasticité revenu significativement inférieure à l'unité (0,9).

Kravis *et al*. (1978) avancent néanmoins une explication : les services (notamment médicaux) étant plus intensifs en travail qu'en capital, le prix relatif des dépenses de santé est plus élevé dans les pays riches que dans les autres en raison des différences de coût du travail. De ce fait, puisque les services ne sont que faiblement échangés (par comparaison à la plupart des produits industriels), l'utilisation du taux de change courant conduit à surestimer comparativement le volume des dépenses de santé dans les pays riches et à le sous-estimer dans les pays pauvres, et donc à surestimer l'effet revenu. Les auteurs soulignent surtout la fragilité statistique des résultats si l'on ne teste pas la significativité d'autres variables que le revenu.

1.1.2 La prise en compte de l'effet prix

Gerdtham et Jönsson (1991) introduisent également directement le prix relatif des dépenses de santé comme variable explicative et concluent à une élasticité de -0,83 , non

significativement différente de -1. L'ajout de facteurs institutionnels (Gerdtham *et al*., OCDE 1992, Gerdtham *et al*., OCDE 1995) ne permet pas de trancher dans le débat sur la valeur de l'élasticité revenu (1,33 pour des données transversales sur 1985, et 0,74 pour une étude sur 24 pays de 1970 à 1991, avec des effets fixes à deux niveaux pour les pays et les années).

On obtient pourtant ainsi des indicateurs intéressants sur les performances comparées des différents types de systèmes de santé : les pays dont le mode dominant de rémunération des prestataires est à l'acte (cas de la France et jusque récemment des Etats-Unis) dépenseraient de ce fait, toutes choses égales par ailleurs, en moyenne 11% de plus en soins médicaux que les pays où le mode de rémunération dominant est la capitation (cas du Royaume-Uni ou du Danemark). L'interprétation en termes microéconomiques de résultats obtenus sur données agrégées est toutefois très délicate, et ce d'autant plus que les indicateurs utilisés sont excessivement schématiques. Par exemple, au Royaume-Uni, les généralistes ne sont pas intégralement capités et sont rémunérés à l'acte pour une partie de leur activité.

1.2 Des résultats confortés par les analyses longitudinales...

Murillo *et al*. (1993) introduisent plusieurs innovations notables : cette fois, l'analyse s'appuie sur des données longitudinales grâce à l'estimation d'un modèle unique de 1960 à 1990, séparément sur chacun des dix pays formant la Communauté Européenne en 1985. Dans la foulée de Gerdtham, les auteurs ont recours aux parités de pouvoir d'achat mais testent deux modèles, en déflatant successivement les dépenses de santé per capita par un indice de prix sectoriel (*weighted quantity of health care*) et un indice de prix global (*health care expenditure*). La première mesure traduit le volume d'actes médicaux effectués, la deuxième reflète plutôt la charge financière que représentent les soins de santé.

Dans la plupart des pays le premier indicateur croît moins rapidement que le second (les prix des dépenses de santé augmentant plus rapidement que l'indice général des prix), mais c'est le contraire en France, où les prix, en particulier dans le secteur du médicament, sont très encadrés. Dans les deux cas est introduit le prix relatif des dépenses de santé comme variable explicative.

Tableau 1 : Elasticités obtenues par Murillo

Pays	*élasticité revenu*	*élasticité prix*
Allemagne	1,41	-0,51
Belgique	1,36	-0,59
Danemark	1,13	0,06[1]
Espagne	1,95	-0,60
France	1,36	-0,64
Irlande	2,17	-0,78
Italie	1,34	-0,29[1]
Pays-Bas	1,23	-0,55
Royaume-Uni	1,61	-2,21

Source : Murillo et al. (1993)

[1] : non significatif au seuil de 5%

Comme le remarquent Murillo *et al*., ce cadre méthodologique semble beaucoup plus adapté que les comparaisons transversales à la très forte hétérogénéité des systèmes de santé dans les pays considérés. La méthode économétrique employée repose sur l'estimation de relations de cointégration par la méthode d'Engle et Granger (cf. infra pour une description plus précise). Ceci leur permet d'établir l'existence de relations de long terme (sauf pour le Luxembourg) reliant le volume des dépenses de santé per capita au PIB par tête et au prix relatif des soins médicaux, lorsque le déflateur est l'indice de prix sectoriel.

Notons toutefois une particularité de la méthode : pour la plupart des pays, l'hypothèse nulle d'absence de relation de cointégration ne peut être rejetée qu'après l'ajout de ruptures de trend (correspondant a priori aux chocs pétroliers) dans la relation de long terme. Bien que les tables proposées par Perron (1989) dans le cas de séries présentant une rupture de trend correspondent à un simple test de stationnarité, les auteurs précités les utilisent dans le cadre du test des relations de cointégration.

L'analyse des relations de cointégration obtenues montre des élasticités revenu toujours supérieures à l'unité, et schématiquement d'autant moins élevées que le pays est plus riche. Ceci renforce la conjecture selon laquelle il existe des phénomènes importants de rattrapage durant lesquels les dépenses de santé croissent sensiblement plus vite que la richesse nationale. Les auteurs décrivent ainsi une **hiérarchie** des pays par élasticité revenu croissante : Danemark et Pays-Bas (1,10 à 1,20) ; Italie, Belgique, France, Royaume-Uni et Allemagne (1,30 à 1,40) ; Espagne et Irlande (de 1,90 à 2,00).

1.3 ...et enrichis par la prise en compte des spécificités du bien santé

L'Horty *et al*. (1997) utilisent une démarche comparable pour la France mais élargissent l'éventail des variables explicatives : aux effets revenus et prix considérés par Murillo *et al*., ils ajoutent des caractéristiques institutionnelles (une estimation du niveau de la prise en charge publique des dépenses de santé), démographiques (impact éventuel d'une consommation médicale plus forte aux âges extrêmes) et surtout technologiques. L'estimation ne commence qu'en 1970 du fait de l'insuffisance des données à leur disposition.

1.3.1 l'effet de la couverture sociale

L'influence du niveau de prise en charge publique des dépenses de santé est a priori relativement incertaine. Certes, les travaux en données transversales suggèrent une corrélation négative (Gerdtham, 1995). Mais il semble que ceci soit dû au contrôle assez strict qui pèse sur les prestataires dans un système intégré (de type britannique ou danois) où la part des dépenses publiques est très forte. Par contre, on peut imaginer qu'une augmentation d'une année sur l'autre de la part de la prise en charge publique dans un système dont les structures demeurent par ailleurs invariantes induise plutôt un accroissement des dépenses.

L'Horty et al. aboutissent effectivement à la conclusion que la prise en charge publique (grossièrement mesurée comme la part des dépenses totales de santé prise en charge directement par l'Etat ou les caisses d'assurance maladie) a un impact positif significatif sur le volume des dépenses : une augmentation de un point du taux de prise en charge induirait une hausse de 2% du volume des dépenses. Mais ce facteur saurait d'autant moins expliquer une part très importante de la croissance des dépenses que l'assurance maladie publique réduit continûment son niveau de prise en charge depuis une quinzaine d'années (via la hausse du ticket modérateur) pour atteindre un taux moyen de 75%.

Certes, la politique de réduction de la prise en charge, outre les restrictions dans l'accès aux soins qu'elle peut entraîner pour des personnes à bas revenus, est peu efficace en raison du remboursement du ticket modérateur par les assureurs complémentaires (qui couvraient, en 1995, 84% de la population française). L'instauration d'un ticket modérateur d'ordre public (i.e. non réassurable auprès d'une mutuelle ou d'une assurance complémentaire) éventuellement modulé en fonction du revenu aurait vraisemblablement plus d'efficacité en termes de maîtrise de la dépense totale.

L'étude de référence (*Rand experiment*) effectuée aux Etats-Unis suggère que l'instauration d'un ticket modérateur (ou coassurance) de 25% amène les assurés à consommer 20% de soins médicaux de moins que s'ils bénéficient de la gratuité totale. Au-delà, le ticket modérateur perd de son efficacité : un ticket modérateur de 95% ne permet qu'une

réduction globale de 30% mais restreint considérablement l'accès aux soins des plus pauvres.

1.3.2 les effets démographiques

L'allure des profils de consommation médicale par âge (avec un pic dans les premières années d'existence et un forte croissance après 65 ans) amène à s'intéresser aux variables démographiques, spécialement dans un contexte de vieillissement rapide de la population au début du siècle prochain.

Les variables démographiques (part des plus de 65 ans dans la population totale, et taux de dépendance, qui exprime le rapport de la population des moins de 20 ans et des plus de 65 ans au reste de la population française) se sont avérées non significatives sur la période 1970-1995 (le taux de dépendance l'est toutefois sur la période 1960-1995) et ne figurent donc pas dans les estimations de L'Horty et al. Ce constat est corroboré par les résultats de Gerdtham et al. (OCDE, 1995) qui estiment que la proportion de personnes âgées de plus de 75 ans n'influence pas significativement le montant des dépenses de santé, et que celle des enfants de moins de 4 ans n'a un impact significatif que sur les dépenses de médecine ambulatoire, pas sur les dépenses totales.

Par ailleurs, Hourriez (1992), sur des données temporelles relatives à la consommation de services médicaux des ménages français, estime que le vieillissement n'expliquerait mécaniquement, à comportements inchangés, qu'une part assez faible de l'accroissement des dépenses de santé. Entre 1980 et 1990, le vieillissement expliquerait 0,26% de croissance des dépenses de santé en volume chaque année, à comparer à un taux de croissance annuelle moyen de 3,3%.

Toutefois, il convient de distinguer **l'effet d'âge** (les besoins médicaux des personnes âgées et des très jeunes enfants sont plus importants) de **l'effet de génération** (les générations récentes sont habituées à des prestations de santé plus importantes que les générations précédentes, qui n'ont pas toutes connu un accès aux soins aussi aisé qu'il l'est aujourd'hui). Il est possible qu'à l'avenir, l'arrivée à des âges avancés de générations plus exigeantes en termes de soins de santé ait un effet accélérateur marqué. Le simple prolongement des tendances passées conduirait ainsi à minorer les conséquences du vieillissement sur la progression des dépenses.

1.3.3 Le progrès technique médical

Dans la grille d'analyse qui sépare effets d'offre et de demande, le statut du progrès technique médical est relativement incertain. On peut arguer que l'apparition de nouvelles technologies permet de soigner des pathologies auparavant non traitées (effet d'offre), mais également que la rapidité de la diffusion de ces technologies dans les cabinets et les services hospitaliers dépend de la demande des patients.

La mesure du progrès technique s'avère par ailleurs particulièrement ardue ; les auteurs envisagent deux types d'estimateurs : une première catégorie rapporte les dépenses en appareils thérapeutiques soit au nombre de lits d'hôpitaux, soit au nombre de médecins ; notons que cette mesure du progrès technique peut s'avérer inadéquate. Par exemple, une stagnation des dépenses en appareils associée à une diminution du nombre de lits (avec augmentation du taux d'occupation des lits) peut induire un accroissement de l'indicateur de progrès technique, alors qu'il faut n'y voir qu'une rationalisation de l'offre de soins.

L'autre méthode d'évaluation consiste à retenir le rapport à la population totale du nombre de personnes traitées pour une pathologie dont la fréquence d'apparition est stable sur la période considérée : insuffisance rénale par exemple. Le risque est que la croissance de ce taux de traitement soit liée autant à la demande (effet revenu, prise en charge par les organismes de Sécurité Sociale...) qu'à l'offre (progrès technique). A cet égard, la première

catégorie d'indicateurs constitue une estimation plus directe du degré de diffusion du progrès technique au sein des hôpitaux.

L'étude réalisée par L'Horty *et al.* sur la France entre 1970 et 1990 montre une élasticité revenu très forte (de l'ordre de 1,50), qui cependant décroît à environ 0,60 si l'on ajoute une tendance linéaire dans la relation de long terme. L'élasticité prix est également très importante, ce qui peut surprendre dans un système où la couverture offerte par l'assurance maladie est assez complète (notamment si l'on tient compte de l'importance de l'assurance complémentaire) : entre -1 (avec tendance) et -1,50 (sans tendance). Le taux de remboursement et l'indicateur de progrès technique sont significatifs et jouent positivement sur les dépenses. L'impact relatif de ces facteurs explicatifs montre la prépondérance de l'effet revenu (plus de 40%) devant le progrès technique et l'effet prix (environ 25% chacun).

2. Une étude économétrique des facteurs d'offre

Comme nous l'avons mentionné en introduction, notre exercice de comparaison est délicat dans la mesure où d'une part les modalités de l'intervention de l'Etat dans le processus de financement des dépenses et/ou de production des soins sont très variables, et d'autre part le poids de cette intervention publique est toujours considérable, même aux Etats-Unis.

Une autre limitation réside dans la faible taille des échantillons : les séries annuelles publiées par l'OCDE dans la base Eco-Santé (édition 1997), que nous utilisons à l'instar des autres études effectuées dans ce domaine, ne débutent pas avant 1960. Qui plus est, beaucoup de variables ne sont renseignées que vers la fin de la période 1960-1995 ou bien sont uniquement disponibles pour un nombre restreint de pays. Cette limitation n'est pas négligeable compte tenu de la méthodologie adoptée dans un premier temps (techniques de cointégration), qui suppose des échantillons de grande taille.

A côté des études s'attachant à démontrer l'influence des facteurs de demande sur le niveau des dépenses de santé (effet revenu, effet prix, prise en compte des effets démographiques ainsi que de la couverture maladie presque universelle garantie dans la plupart des pays développés), il semble intéressant de réaliser des travaux davantage ciblés sur le rôle des facteurs d'offre.

Nous retenons six pays : les Etats-Unis, la France, l'Allemagne, l'Italie, le Danemark et les Pays-Bas. Malgré sa taille, nous excluons le Royaume-Uni faute de données suffisantes.

2.1 Présentation des variables

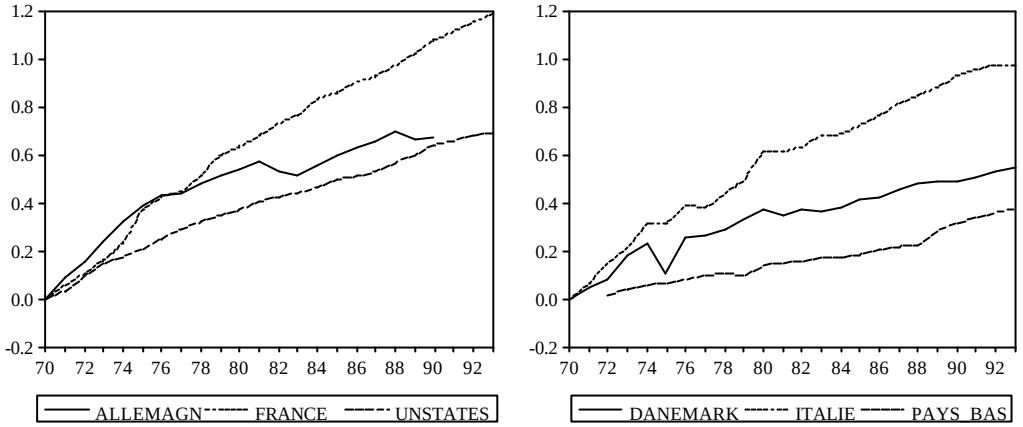
Nous nous intéressons à plusieurs indicateurs : certains se réfèrent au progrès technique (équipements lourds, prothèses, nouveaux médicaments...) dont la diffusion dépend largement de l'attitude et des exigences thérapeutiques tant des demandeurs que des offreurs de soins; d'autres essaient d'analyser plus spécifiquement, au vu des spécificités du bien santé, les comportements économiques des prestataires : l'hypothèse de l'existence d'une demande induite est-elle recevable? Comment les médecins réagissent-ils aux fluctuations des prix des soins (largement administrés dans de nombreux pays comme la France)?

L'idée sous-jacente est que les prestataires ont individuellement la possibilité d'**induire** une partie de la demande de leurs patients : dans un contexte d'excès d'offre ou bien de limitation imposée de la progression de leurs honoraires par acte, les praticiens peuvent être tentés d'user de ce pouvoir afin de prévenir une baisse de leurs revenus. Il est légitime de penser que le degré de diffusion du progrès technique dans les cabinets médicaux et les hôpitaux peut influencer à la hausse le volume des actes, soit que les praticiens eux-mêmes doivent rentabiliser des investissements réalisés, soit que les producteurs de biens médicaux ou de médicaments aient intérêt à orienter la demande vers des produits sophistiqués dégageant une marge bénéficiaire supérieure.

Le modèle retenu vise à expliquer les dépenses de santé par habitant par l'indice de prix relatif des soins médicaux, la densité médicale et un indicateur de progrès technique.

2.1.1 La variable expliquée : le volume des dépenses de santé

Graphiques 2a et 2b : Evolution du volume des dépenses de santé per capita (en log).



Pour permettre une meilleure comparaison, on impose la valeur initiale zéro pour chaque série.

Les dépenses de santé ont crû assez régulièrement en volume depuis 1970 dans la plupart des pays, mais le phénomène est particulièrement accentué en France (où l'augmentation totale dépasse les 200% et ne donne pas de signe d'essoufflement jusqu'au milieu des années 1990). Ailleurs la croissance est moins forte sur l'ensemble de la période (cas du Danemark et des Pays-Bas) ou bien a vu son rythme décroître, comme c'est le cas en Allemagne depuis la fin des années 1970.

2.1.2 Le progrès technique médical

Le volume des dépenses annuelles en appareils thérapeutiques, rapporté au nombre de médecins, permet une évaluation de la diffusion du progrès technique médical au sein tant des infrastructures hospitalières que des cabinets médicaux. Bien sûr, cet indicateur est assez grossier, puisqu'il est par exemple incapable de prendre en compte certains aspects du progrès scientifique qui ne se traduisent pas par des investissements technologiques nouveaux (comme ce peut être le cas en matière psychiatrique).

Notons qu'une diffusion moindre du progrès technique médical n'implique pas *ipso facto* des performances sanitaires médiocres : la qualité des soins fournis dépend d'un ensemble complexe de facteurs qui ne se réduisent pas aux seuls taux d'équipement en technologies de pointe.

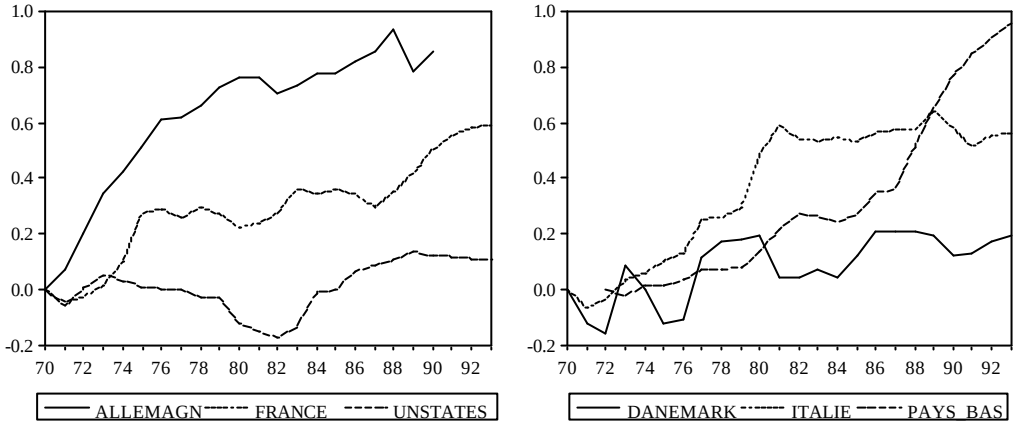
Tableau 2 : Indicateurs sanitaires

	D	*P*	*E*	*USA*	*F*	*I*	*GB*
Mortalité infantile (pour 1000 naissances, 1980)	12,7	24,3	12,3	12,6	10,0	14,6	12,1
Mortalité infantile (pour 1000 naissances, 1995)	5,3	7,4	5,5	8,0	5,0	6,2	6,0
Espérance de vie à 60 ans (femmes, 1994)	22,5	22,0	24,1	22,9	25,0	23,5	22,4
Espérance de vie à 60 ans (hommes, 1994)	18,1	18,0	19,5	18,9	19,7	19,0	18,3

L'indicateur de progrès technique subit des évolutions souvent heurtées d'une année sur l'autre mais la tendance générale est à la croissance (cf. graphique 3). Il présente

l'inconvénient d'agréger des produits dont le coût va plutôt décroissant (lunetterie) avec des produits dont le développement est plus récent et dont les coûts demeurent extrêmement élevés, mais qui répondent à un besoin qu'il était jusqu'à présent techniquement difficile de satisfaire (prothèses auditives).

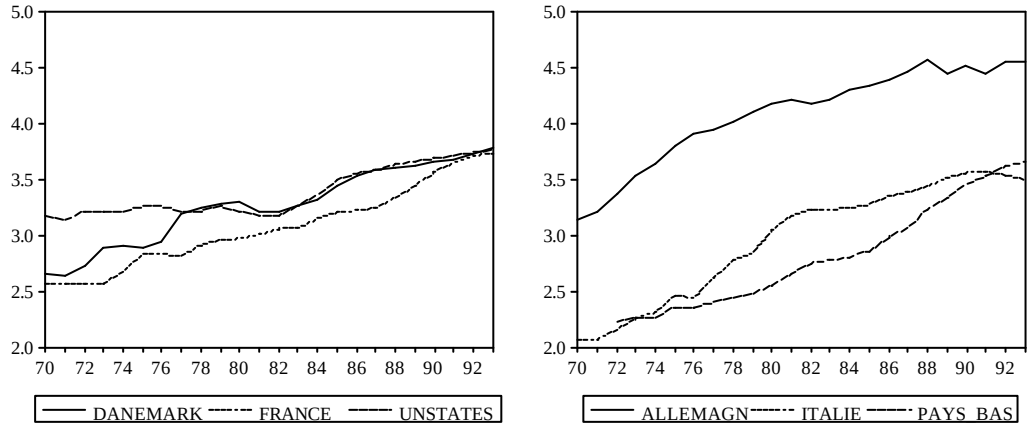
Graphiques 3a et 3b :
Evolution du volume des dépenses en appareils thérapeutiques par médecin en activité



Pour permettre une meilleure comparaison, on impose la valeur initiale zéro pour chaque série.

La tendance à la croissance que nous avons mentionnée comporte une exception notable : les Etats-Unis. En effet, dans ce pays, la croissance du volume des dépenses en appareils thérapeutiques par habitant (cf. graphiques 4) est presque compensée par l'accroissement de la densité médicale. Cependant, l'accroissement de la densité médicale dénote également la diffusion du progrès technique, puisque, en particulier aux Etats-Unis, cet accroissement est dû pour une bonne part à l'arrivée massive sur le marché des soins de médecins spécialistes utilisant des techniques nouvelles.

Graphiques 4a et 4b : Evolution du volume des dépenses en appareils thérapeutiques per capita (logarithmes) exprimées en dollars US et en parités de pouvoir d'achat.



Il est également vraisemblable que le progrès technique soit sous-évalué par notre indicateur, parce qu'il ne prend pas suffisamment en compte les équipements lourds (scanners...) sur lesquels les données sont fragmentaires et récentes. Cette sous-évaluation est encore accentuée pour les Etats-Unis par le fait que les spécialistes y représentent une fraction du corps médical beaucoup plus élevée qu'en Europe.

Un second indicateur de progrès technique est le volume des dépenses en appareils thérapeutiques *per capita* exprimées en PPA (Parités de Pouvoir d'Achat par rapport au PIB). Cet indicateur montre une remarquable convergence de cette variable vers le niveau des Etats-Unis en ce qui concerne la France, l'Italie, les Pays-Bas et le Danemark (cf graphique 4). En Allemagne, par contre, les dépenses en appareils thérapeutiques sont notablement supérieures. Nous avons également testé d'autres indicateurs comme le volume des dépenses en produits pharmaceutiques, mais ils se sont avérés statistiquement non significatifs.

2.1.3 Le comportement des praticiens

L'hypothèse de demande induite (cf. encadré 1)

L'analyse microéconomique traditionnelle prévoit que l'élasticité de l'offre au prix est positive ou nulle. Les spécificités du bien santé tout comme l'organisation du marché des soins impliquent une approche différente. Tout d'abord, l'existence d'asymétries d'information entre le prestataire de soins et le patient donne la possibilité au praticien d'influer sur la demande du patient (puisque ce dernier n'est pas toujours à même de juger de la pertinence du traitement qui lui est prescrit).

La généralisation de l'assurance maladie (presque universelle aujourd'hui dans la plupart des pays développés, à l'exception des Etats-Unis) entraîne de plus des phénomènes d'aléa moral. Toutefois, en matière de santé, il s'agit moins d'un aléa moral *ex ante* (i.e. le fait que les assurés relâchent leurs comportements de prévention, ce qui entraînerait une survenance accrue de la maladie) que d'un aléa moral *ex post* (une fois la maladie déclarée, les assurés ont tendance à surconsommer : cette surconsommation pouvant en partie résulter de leur mauvais degré d'information en matière médicale).

Le pouvoir potentiel **d'induction de la demande**, engendré par le monopole de la connaissance médicale dont bénéficient les médecins, est donc accru par la faible sensibilité des patients aux prix. Il est a priori vraisemblable que les effets d'induction soient supérieurs dans des pays où les prestataires sont essentiellement rémunérés à l'acte (comme en France, ce qui est une incitation claire à accroître le volume des soins) que dans ceux où les médecins sont rémunérés à la capitation ou bien gèrent une enveloppe financière globale (comme en Grande-Bretagne). L'encadré n°1 explicite les différentes approches théoriques envisagées pour formaliser les comportements des prestataires dans un cadre où l'asymétrie « naturelle » (car liée au différentiel de connaissances scientifiques) entre médecins et patients se double des phénomènes d'asymétrie d'information inhérentes au contexte de l'assurance.

Les études empiriques étrangères (exemple québécois, Rochaix, 1993) montrant que les prestataires auraient en pratique tendance à faire un usage raisonnable de ce pouvoir (i.e. afin de garantir le maintien de leurs revenus plutôt que de réaliser les bénéfices les plus grands possibles) font qu'il ne serait pas a priori déraisonnable d'observer une élasticité-prix *négative* de l'offre de soins.

Pour analyser ces comportements, deux variables ont été introduites parmi les régresseurs : la densité médicale et le prix relatif des soins de santé.

Encadré 1 : Demande induite et théorie microéconomique

Si le cadre théorique usuel de la théorie de l'assurance peut permettre une explication des phénomènes d'aléa moral *ex post* observés dans le domaine de la santé, l'asymétrie d'information existant entre le prestataire et le patient est formellement beaucoup plus difficile à appréhender : il s'agit en effet d'une remise en cause de l'indépendance des fonctions d'offre et de demande, et donc de la souveraineté du consommateur. Deux types de modélisations ont été développés dans un premier temps pour résoudre cette difficulté (Rochaix, 1997).

La première catégorie de modèles suppose une différence de nature entre les soins médicaux et les autres biens et implique une subordination complète de la demande des patients aux préférences des offreurs de soins. Certains modèles - passablement hétérodoxes - postulent alors que les prestataires ont pour objectif, non de maximiser leur utilité, mais d'obtenir un niveau de satisfaction « acceptable » auquel correspond concrètement un revenu-cible. D'autres modèles conservent l'hypothèse de maximisation de l'utilité mais introduisent dans la fonction d'utilité du prestataire des arguments qui tendent à limiter son pouvoir discrétionnaire, comme l'éthique professionnelle (Dionne et Constandriopoulos, 1984).

L'autre catégorie s'attache à restaurer l'indépendance du consommateur, et suppose que l'asymétrie d'information entre consommateurs et prestataires traduit non pas une différence de nature entre les soins médicaux et les autres biens, mais seulement le fait que les coûts d'information pour le patient sont particulièrement élevés dans le domaine de la santé. Par exemple, l'introduction de la dimension temporelle (le patient ignore tout des caractéristiques du prestataire lors de la première consultation, mais il peut décider de ne pas revenir) est un moyen de réhabiliter le pouvoir du consommateur : la limite au pouvoir discrétionnaire du médecin réside dans la perte de crédibilité qu'il subit s'il en abuse. Dranove (1988) montre que dans un tel cadre, une densité médicale élevée et une diffusion importante de l'assurance maladie incitent les prestataires à prescrire un niveau de traitement trop élevé.

Les dernières années ont plutôt vu le développement des modèles Principal-Agent tentant de prendre en compte les multiples relations d'agence entre les différents acteurs du marché : les patients, les prestataires, les assureurs, voire l'autorité de tutelle (instance de régulation du marché). L'application de ces modèles à la santé conduit à des formulations de plus en plus complexes :

modèles multi-task (prise en compte du dilemme prix-qualité dans les décisions de l'agent-prestataire, sachant que le principal-patient observe imparfaitement la variable qualité, Holmstrom et Milgrom, 1991) ;

modèles multi-principaux (le patient et l'assurance sont deux principaux dans deux relations différentes, mais où l'agent est le même : le médecin, Blomqvist, 1991);

modèles multi-agents (le principal, i.e. l'assureur ou la tutelle, observe les comportements de plusieurs agents et utilise cette information comparative pour deviner les coûts effectifs et fixer les prix - on parle de *yardstick competition* - ou bien instaure des quasi-marchés ; les agents peuvent être des prestataires, comme au Royaume-Uni, ou bien des assureurs, comme aux Pays-Bas, Mougeot, 1994).

La densité médicale

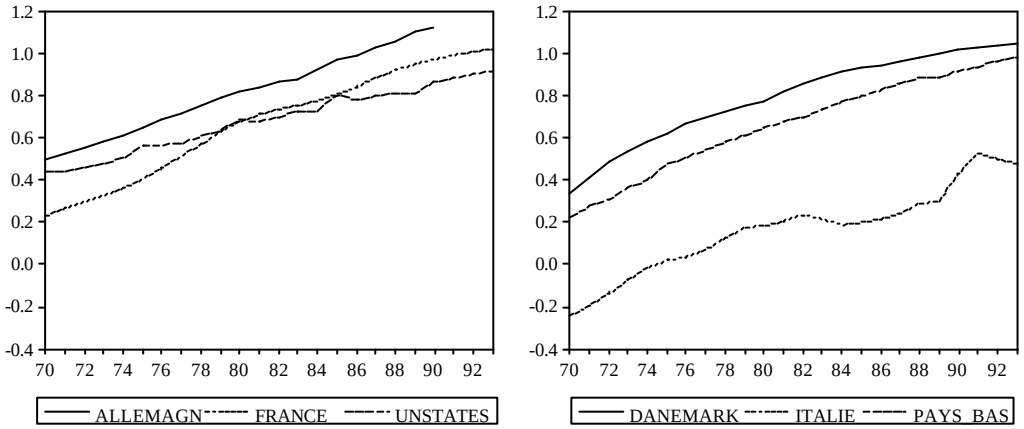
L'accroissement de la **densité médicale** (rapport du nombre de médecins à la population totale) est un phénomène ambigu : il peut certes n'être que la réponse à une demande croissante de la part des patients ou à des effets de structure démographique (fort poids des personnes âgées) ; toutefois, de nombreux travaux (Rochaix, 1993) soulignent également que les médecins ont la possibilité d'induire au moins partiellement la demande de leurs patients, afin par exemple d'augmenter le volume de leurs actes s'ils sont

confrontés à une baisse de leurs honoraires. Il serait donc intéressant de voir si l'élasticité des dépenses de santé à la densité médicale varie ou non beaucoup d'un pays à l'autre.

La densité médicale croît sensiblement dans la plupart des pays, et à des rythmes souvent voisins ; les niveaux sont par contre assez différents. Il faut noter que cette variable masque des hétérogénéités géographiques très fortes. C'est naturellement le cas dans tous les pays (en France, en 1996, la densité médicale pour 10 000 habitants variait entre 22 en Picardie et 38 en Ile-de-France et en Provence-Alpes-Côte d'Azur).

Néanmoins, l'hétérogénéité géographique est notablement accentuée aux Etats-Unis, où la faible solvabilité d'un grand nombre des personnes non assurées ou mal assurées éloigne les médecins de grandes zones économiquement en difficulté. La valeur - relativement médiane - de la densité médicale aux Etats-Unis occulte donc des phénomènes locaux de surpopulation (notamment chez les spécialistes).

Graphiques 5a et 5b : Evolution de la densité médicale (en log).



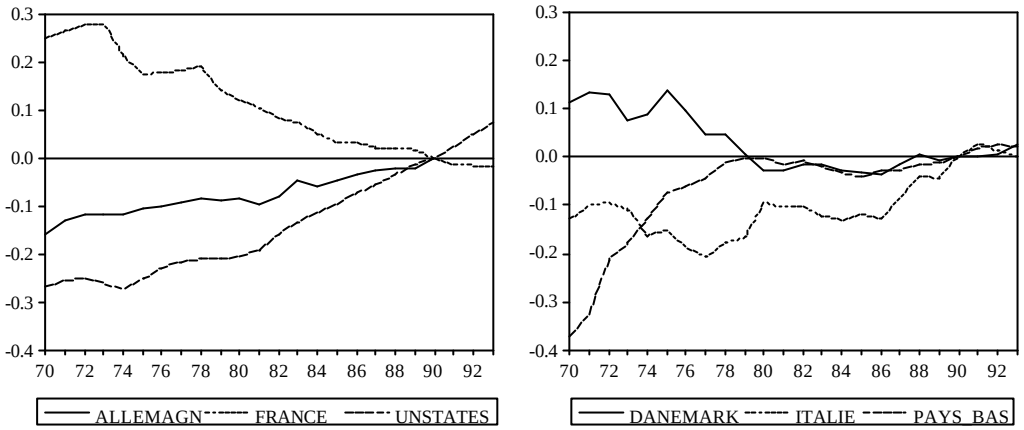
La densité médicale est définie comme le nombre de médecins pour 1 000 habitants.

La variable prix

Enfin, nous introduisons dans la régression une troisième variable explicative : **le prix relatif des dépenses de santé** (rapport de l'indice des prix des dépenses de santé à l'indice des prix de la consommation finale des ménages), qui est une proxy de la rémunération des médecins et des laboratoires. Les indices globaux fournis par les séries de l'OCDE sont en effet une base beaucoup plus pertinente pour l'analyse des facteurs d'offre que de demande : ces indices reflètent le montant effectivement perçu par un prestataire pour un acte médical, et non le coût marginal d'achat d'une unité de soins pour le consommateur, puisqu'il faudrait pour cela tenir compte de l'impact de la couverture assurancielle.

Si la croissance du volume des prestations s'avère moins forte aux Etats-Unis qu'en France, la charge financière s'y est pourtant accrue dans des proportions plus importantes en raison de la croissance soutenue du prix relatif des dépenses de santé aux Etats-Unis (avec même une accélération en fin de période) alors que le mouvement est opposé en France. Les évolutions de cet indice sont du reste très variables d'un pays à l'autre : croissance modérée en Allemagne et en Italie, relative stabilité depuis vingt ans au Danemark, croissance puis stabilité aux Pays-Bas.

Graphiques 6a et 6b : Evolution du prix relatif des dépenses de santé (en log).



Les données sont en base 1990. Les séries prennent de ce fait la valeur zéro à cette date.

Notons un écueil de taille : la pertinence de cet indicateur est discutable en raison même de son mode de construction. Bien sûr, il agrège les prix des dépenses des différents secteurs de la santé (hôpital, médecine ambulatoire, etc.) qui ont évolué de façon souvent contradictoire ; mais surtout, il n'est pas certain qu'il parvienne à prendre en compte de manière satisfaisante les évolutions technologiques.

Ainsi, aux Etats-Unis, MacClellan et al. (1997) estiment que l'indice sectoriel des produits pharmaceutiques est notablement surévalué, en raison notamment d'une sous-estimation de l'impact baissier de la diffusion des génériques. En France, la plupart des prix sont administrés et les instances de régulation ont longtemps misé sur une progression plus lente de l'indice sectoriel des prix des dépenses de santé que de l'indice global des prix à la consommation pour limiter la charge financière de l'assurance-maladie. Dans certains cas, confrontés à la réduction de la marge bénéficiaire de leurs produits, les industriels auraient cherché - et partiellement réussi - à reporter une partie de la consommation pharmaceutique vers des produits plus coûteux. En sous-évaluant ce phénomène, l'indice de l'OCDE sous-estimerait le niveau du prix relatif des soins médicaux.

2.2 Résultats économétriques

2.2.1 Méthodologie

Nous avons donc cherché à tester l'existence de relations de long terme entre les dépenses de santé et des variables dont il semble qu'elles recouvrent en partie des effets d'offre plus que de demande[2] : indice de prix relatif, indicateur de progrès technique et densité médicale, sur six pays. Ces variables présentent toutefois une non-stationnarité d'ordre 1 (au minimum). Les tests de stationnarité présentés en annexe 1 (et effectués avec la méthode de Jobert, 1993) le montrent sans ambiguïté.

La présence de non-stationnarité de nature stochastique rend plus délicate l'interprétation des relations économétriques dans la mesure où elle peut conduire à l'obtention de régressions fallacieuses, le niveau élevé du R^2 traduit non plus une forte relation causale

[2] Nous nous sommes d'abord attachés à répliquer l'estimation effectuée par L'Horty et al. sur la France : régression des dépenses de santé par tête sur le PIB par tête, le prix, le taux de remboursement de la Sécurité Sociale et un indicateur de progrès technique. Toutefois, si nous retrouvons à peu près les mêmes résultats, les valeurs critiques que nous utilisons pour les tests de cointégration ne nous permettent pas de rejeter l'hypothèse nulle d'absence de relation de cointégration.

entre les séries incriminées mais plutôt une tendance stochastique ou déterministe commune.

Nous recherchons donc des relations de long terme entre les variables par des techniques de cointégration. Une relation de cointégration correspond à une relation de long terme stable entre des variables non stationnaires (comportant une tendance stochastique et éventuellement déterministe). Le concept de **cointégration** traduit ainsi une liaison entre les trajectoires des séries dans le long terme, qui assure la stationnarité des résidus issus de la relation de long terme. La présence d'une relation de cointégration entre des séries non stationnaires valide la régression classique sur les variables en niveau dans la méthode en deux étapes d'Engle et Granger (1987).

Estimer le modèle en différences premières conduirait à des estimateurs biaisés du fait de l'omission de la relation de long terme existant entre les séries. Le théorème de représentation de Granger montre toutefois qu'en présence de variables cointégrées une **représentation à correction d'erreur (ECM)** est adaptée pour décrire non seulement la dynamique du système mais aussi l'équilibre de long terme vers lequel celui-ci converge.

La méthode d'Engle et Granger (1987)

Plusieurs méthodes ont été proposées pour l'estimation des relations de cointégration. Si des méthodes plus sophistiquées que celle que nous utilisons ont été développées (en particulier les méthodes multivariées à la Johansen), le faible nombre d'observations dont nous disposons pour chaque pays (au plus 25) ne nous permet pas d'estimer un grand nombre de coefficients. En ce sens les méthodes multivariées sont difficilement applicables.

La méthode d'Engle et Granger consiste à estimer dans une première étape par moindres carrés ordinaires la relation de long terme entre les variables en niveau. Les statistiques de Student sont à ce niveau inexploitables pour tester la significativité des coefficients des différentes variables explicatives.

Le seul moyen de vérifier la **validité** de cette relation de long terme est d'effectuer le test de cointégration, i.e. de tester si les résidus de la relation estimée entre les variables non-stationnaires sont ou non stationnaires. Pour ce faire, Engle et Granger proposent d'effectuer un test de stationnarité de Dickey et Fuller sur les résidus (ce test est éventuellement augmenté par l'ajout de retards, afin d'obtenir des résidus non autocorrélés, ce que nous vérifions à l'aide de la statistique de Ljung-Box) ; l'hypothèse nulle est celle de non-stationnarité des résidus, donc d'absence de relation de cointégration. Des tables spécifiques, différentes de celles de Dickey et Fuller, doivent être utilisées pour ce test de stationnarité ; une première version a été proposée par Engle et Granger, mais nous leur préférons les tables de MacKinnon (1991), plus précises, qui permettent une meilleure prise en compte du nombre de variables de la relation de long terme et de la taille de l'échantillon.

Si le test amène à rejeter l'hypothèse nulle d'absence de relation de cointégration, il est alors licite de procéder à une estimation du modèle en différences premières en ajoutant aux variables explicatives les résidus retardés de la relation de long terme. La valeur du coefficient des résidus retardés traduit alors l'intensité de la force de rappel qui assure la stabilité de long terme du système.

2.2.2 Résultats

Le tableau ci-dessous donne les estimations effectuées sur les six pays retenus pour l'analyse : la période retenue est en général 1970-1993, à l'exception des Pays-Bas

(problème de données en 1970-71) et de l'Allemagne (à partir de 1991 les chiffres de l'OCDE comprennent également les Länder orientaux).

Estimation de la relation de long terme :

$$y_t = \sum_i a_i x_{i,t} + e_t$$

Les statistiques entre parenthèses peuvent être comparées à la table usuelle de Student : elles ont été remaniées pour cela selon la méthode d'estimation des écarts-types développée par Hamilton (1994).

Tableau 3a : Relation de long terme

Les graphiques figurent en annexe.

	France	*Allemagne*	*Pays-Bas*	*Etats-Unis*	*Italie*	*Danemark*
Période	1970-1993	1970-1990	1972-1993	1970-1993	1970-1993	1970-1993
Prix relatif des dépenses de santé	-0,40 (-2,3)	-0,28 (-0,7)	0,07 (1,7)	-0,19 (-0,9)	0,32 (0,8)	0,03 (0,2)
Indicateur de progrès technique	0,35 (7,8)	0,59 (17,4)	0,23 (22,6)	0,08 (0,9)	0,48 (3,1)	0,24 (4,1)
Densité médicale	1,03 (15,5)	0,30 (3,1)	0,15 (7,2)	1,45 (10,7)	0,83 (3,9)	0,64 (17,3)
Test de cointégration (Statistique de Student)	**-4,77**	**-4,73**	**-5,96**	**-4,63**	**-2,51**	**-6,08**
Valeur critique à 5%	-4,58	-4,66	-4,63	-4,58	-4,58	-4,58
Valeur critique à 1%	-5,47	-5,60	-5,55	-5,47	-5,47	-5,47
R² corrigé (rel. de long terme)	0,998	0,992	0,992	0,979	0,973	0,967
Erreur moyenne	**1,70%**	**1,78%**	**0,98%**	**2,99%**	**4,93%**	**2,87%**

A l'exception de l'Italie, l'hypothèse nulle d'absence de relation de long terme est toujours rejetée au seuil de 5%. Du reste, pour l'Italie, l'erreur moyenne élevée (près de 5%) confirme l'absence de relation de long terme stable entre les quatre variables. De plus le graphique des résidus de la relation de long terme traduit bien leur non-stationnarité (cf. annexe 2). **L'ajustement est particulièrement satisfaisant pour la France, l'Allemagne, les Pays-Bas et le Danemark et les Etats-Unis (cf. graphiques).** Par contre, il est médiocre pour l'Italie.

Le modèle à correction d'erreur (ECM) confirme la bonne qualité de l'ajustement pour la France, l'Allemagne, les Pays-Bas et le Danemark : le coefficient de détermination ajusté excède 0,7 pour ces pays, mais il est particulièrement faible pour l'Italie et les Etats-Unis. Le coefficient traduisant la force de rappel prend toutefois des valeurs supérieures à l'unité (en valeur absolue) pour les Pays-Bas et le Danemark, ce qui correspond à un surajustement un peu surprenant. Toutefois, l'hypothèse nulle que cette force de rappel soit égale à -1 n'est pas rejetée au seuil de 60% pour le Danemark, et au seuil de 15% pour les Pays-Bas. Pour les Etats-Unis et l'Italie, ce terme de rappel est assez faible et non significatif au seuil de 5%.

Estimation du modèle à correction d'erreur :

$$\Delta y_t = \sum_i a' \Delta_i x_{i,t} + b\hat{e}_{t-1} + h_t$$

Tableau 3b : Modèle à Correction d'Erreur

Les graphiques figurent en annexe.

	France	*Allemagne*	*Pays-Bas*	*Etats-Unis*	*Italie*	*Danemark*
Coefficients de court terme						
Prix relatif des dépenses de santé	-0,45 (-2,87)	-0,29 (-0,99)	0,26 (2,22)	-0,56 (-2,52)	0,12 (0,49)	-0,53 (-1,70)
Indicateur de progrès technique	0,32 (5,93)	0,52 (9,75)	0,04 (2,29)	0,05 (0,78)	0,18 (1,44)	0,14 (1,65)
Densité médicale	1,08 (12,28)	0,46 (3,60)	0,37 (2,32)	0,15 (0,97)	0,37 (1,72)	0,73 (2,23)
Coefficient du terme de rappel	-0,81 (-4,47)	-0,87 (-3,83)	-1,18 (-5,40)	-0,18 (-1,47)	-0,24 (-1,52)	-1,11 (-5,18)
R^2 corrigé (ECM)	**0,714**	**0,824**	**0,783**	**0,147**	**0,032**	**0,717**
Test de Ljung-Box (5 retards)	60,2%	32,5%	36,1%	43,8%	22,3%	97,9%
Test de Normalité de Jarque-Bera	41,3%	79,2%	22,9%	60,5%	64,7%	89,0%
Test d'effet ARCH	56,8%	34,1%	80,7%	54,0%	21,6%	8,8%
Test d' hétéroscédasticité de White	75,7%	13,1%	38,9%	58,9%	20,3%	39,9%

Les divers tests effectués sur les résidus du modèle à correction d'erreur (autocorrélation, normalité, hétéroscédasticité, effet ARCH, cf tableau 3b) doivent être interprétés avec prudence, les valeurs critiques utilisées étant valables asymptotiquement[3].

D'une part, l'inclusion d'une tendance linéaire dans le modèle à correction d'erreur, légitimée par le fait que pour chaque pays l'une au moins des séries serait de type $I(1)+T^2$, modifie sensiblement les valeurs des coefficients. D'autre part, les résultats ne semblent pas extrêmement robustes lorsque l'on modifie la période d'estimation.

Si l'on écarte l'Italie et les Etats-Unis (où l'ajustement est peu satisfaisant) l'élasticité de long terme des dépenses de santé à l'indicateur de progrès technique est relativement homogène : entre 0,25 (Pays-Bas et Danemark) et 0,6 (Allemagne) en passant par 0,35 (France). Par contre, les différences sont beaucoup plus marquées en ce qui concerne la densité médicale : entre 0,15 (Pays-Bas) et 1,03 (France). L'élasticité-prix est négative en France et en Allemagne, positive mais à des niveaux très faibles aux Pays-Bas et au Danemark.

Il est intéressant d'évaluer les contributions des différents facteurs à la croissance du volume des dépenses de santé : ceci est obtenu en calculant l'augmentation des dépenses induites (selon la relation économétrique) entre le début et la fin de la période d'estimation

[3] *Note sur les tests :*
Le test de Ljung-Box *teste l'autocorrélation des résidus en tenant compte des cinq premières autocorrélations, alors qu'un test de Durbin-Watson ne considère que l'autocorrélation de premier ordre. L'hypothèse nulle est celle d'absence d'autocorrélation.*
Le test de Jarque et Bera *teste la normalité des résidus (en examinant les valeurs de la skewness et de la kurtosis). L'hypothèse nulle est celle de normalité.*
Le test d'effet ARCH *teste l'hétéroscedasticité conditionnelle autorégressive des résidus. L'hypothèse nulle est celle d'absence d'hétéroscédasticité de ce type.*
Le test d'hétéroscédasticité de White *teste l'hétéroscédasticité usuelle des résidus (i.e. correspondant à une corrélation des erreurs et des régresseurs). L'hypothèse nulle est celle d'homoscédasticité.*
Pour chaque test nous indiquons le seuil limite de rejet de l'hypothèse nulle. Ainsi, si ce seuil est supérieur à 5%, l'hypothèse nulle n'est pas rejetée au seuil de 5%.

par les variations de chaque facteur en supposant que l'on maintient les autres facteurs explicatifs à leur niveau de début de période (soit en général 1970). Les chiffres des tableaux suivants correspondent aux différences des logarithmes.

Tableaux 4a à 4e : contributions estimées des variables explicatives

France	Dépenses de santé	Effet prix	Effet progrès technique	Effet densité médicale	Résidu
Dlog	1,19	0,11	0,21	0,82	0,05
Part dans l'ensemble	100	9	18	**69**	4

Allemagne	Dépenses de santé	Effet prix	Effet progrès technique	Effet densité médicale	Résidu
Dlog	0,67	-0,04	0,50	0,19	0,02
Part dans l'ensemble	100	-6	**75**	28	3

Pays-Bas	Dépenses de santé	Effet prix	Effet progrès technique	Effet densité médicale	Résidu
Dlog	0,36	0,02	0,22	0,10	0,02
Part dans l'ensemble	100	6	**61**	28	5

Etats-Unis	Dépenses de santé	Effet prix	Effet progrès technique	Effet densité médicale	Résidu
Dlog	0,70	-0,07	0,01	0,69	0,06
Part dans l'ensemble	100	-10	1	**99**	10

Danemark	Dépenses de santé	Effet prix	Effet progrès technique	Effet densité médicale	Résidu
Dlog	0,55	0	0,05	0,46	0,04
Part dans l'ensemble	100	0	9	**84**	7

Sur les cinq pays pour lesquels les tests de cointégration semblent établir l'existence d'une relation stable de long terme entre nos variables, le facteur explicatif dominant de la croissance du volume des dépenses est par deux fois le progrès technique (Allemagne et Pays-Bas), et par trois fois la densité médicale (France, Etats-Unis et Danemark).

Bien sûr, répétons-le, la prédominance de la densité médicale parmi les facteurs explicatifs n'implique pas une induction forte de la demande par les prestataires : la croissance de la densité médicale est aussi (voire d'abord) la réponse à une croissance de la demande des patients. La comparaison avec les études antérieures suggère un possible effet du PIB, les variables considérées ici comportant en effet un aspect demande.

2.2.3 Interprétation

La variabilité très forte, d'un pays à l'autre, de l'élasticité des dépenses de santé à la densité médicale traduit sans doute des modes d'organisation différents de l'offre de soins. En particulier, l'existence d'une élasticité unitaire en France, très supérieure à celle que l'on observe chez des pays voisins comme l'Allemagne, soulève certaines interrogations.

Il est intéressant de rapprocher ces résultats d'une variable-clef non présente dans notre modélisation : **le mode de rémunération dominant des prestataires.**

Il paraît a priori logique que l'incitation pour les prestataires à induire la demande de leurs patients soit plus forte dans des systèmes où le mode de rémunération est à l'acte (sans limitation vraiment contraignante du volume d'actes effectué par chaque médecin) que dans des systèmes où une certaine contrainte s'exerce sur la rémunération des prestataires. **Cette contrainte peut prendre divers aspects.**

Les aspects incitatifs de la rémunération des prestataires (Bocognano et al., 1998)

Les prestataires peuvent être rémunérés à la capitation, c'est-à-dire qu'ils reçoivent un forfait annuel pour chaque patient officiellement inscrit dans leur clientèle, ce forfait étant modulé en fonction des caractéristiques de l'individu (âge, sexe, antécédents médicaux...). C'est ainsi qu'étaient rémunérés la plupart des prestataires britanniques jusqu'à la fin des années 1980. Dans ce contexte il n'y a pas d'incitation à augmenter le volume des actes au-delà de ce que les besoins médicaux exigent, mais il n'y pas non plus d'incitation à une plus grande efficacité.

Les réformes du NHS britannique instaurées à la fin des années 1980 ont maintenu la rémunération des prestataires à la capitation, mais les médecins généralistes peuvent devenir gestionnaires de fonds (fundholders), c'est-à-dire que l'autorité de régulation leur alloue un budget global qui couvre non seulement les frais de consultation, mais aussi les examens, consultations externes et certains actes de chirurgie non vitale effectués à l'hôpital, et les dépenses de prescription pharmaceutique. L'objectif était de sensibiliser les généralistes aux coûts hospitaliers. Pour le médecin, ce mode de rémunération incite non pas à multiplier le volume des consultations ou des prescriptions (les économies réalisées à l'année n pouvant être reportées sur l'année n+1), mais à dispenser, pour un budget donné, des soins de meilleure qualité (en cherchant le meilleur rapport qualité-prix auprès des hôpitaux), afin de fidéliser sa clientèle et d'en accroître l'importance (ce qui a une incidence directe sur sa rémunération).

D'autres modes de régulation sont envisageables : par exemple, en Allemagne, jusque 1997, les caisses d'assurance maladie négocient au préalable avec les associations de médecins une enveloppe financière globale. Celle-ci est redistribuée ensuite aux médecins au pro rata de leur activité. Si ce système garantit ex ante que les dépenses ne dépasseront pas les prévisions, son efficacité pour la maîtrise des dépenses dépend beaucoup du pouvoir de négociation respectif des caisses et des médecins lors des négociations annuelles. De plus, chaque praticien a individuellement intérêt à effectuer le plus grand nombre d'actes possible, même si c'est au détriment de ses collègues. La régulation est alors interne à la profession : les associations de médecins ont la possibilité de sanctionner les dérives éventuelles de leurs adhérents.

Demande induite et incitations à l'économie

Parmi les cinq pays pour lesquels le test de cointégration valide l'hypothèse d'existence d'une relation stable de long terme, deux connaissent, sur la période 1970-1993, un mode de rémunération à l'acte où il n'y a pas réellement d'incitation à limiter le volume des actes médicaux : ce sont **la France** et **les Etats-Unis**.

Ceci est vrai essentiellement de la médecine ambulatoire : les hôpitaux publics français sont financés par une dotation forfaitaire depuis 1984, et l'assurance maladie publique pour les personnes âgées Medicare indemnise les établissements hospitaliers sur la base d'un coût prospectif modulé selon la pathologie concernée. De plus, les transformations très rapides du système de soins américains depuis 1993 font qu'il n'est plus possible, en 1999, de dire que les médecins américains exerçant en cabinet n'ont pas d'incitation à limiter le volume de leurs actes (développement du managed care avec l'extension des formules de type HMO ou PPO).

En **Allemagne**, on l'a vu, le mode de rémunération à l'acte coexiste avec la possibilité pour les associations professionnelles de limiter la propension de leurs adhérents à accroître excessivement le volume de leurs actes et de leurs prescriptions.

Au **Danemark**, le système de soins est de type vertical intégré, dont l'exemple le plus typique était le NHS britannique (National Health Service) d'avant les réformes de la fin des années 1980. Les prestataires n'ont donc pas intérêt à accroître le volume de leurs prestations au-delà des nécessités médicales.

Le cas des **Pays-Bas** est plus complexe : les hôpitaux sont incités depuis 1983 à ne pas accroître le volume des prestations, puisque depuis 1983, ils reçoivent des budgets annuels prospectifs négociés avec les assureurs locaux publics et privés. S'agissant de la médecine ambulatoire, il faut distinguer deux catégories d'assurés : ceux relevant d'une assurance publique s'inscrivent auprès d'un généraliste qui est alors rémunéré à la capitation ; par contre, les fonctionnaires et les personnes à revenus élevés exclues de l'assurance publique ont la liberté de choix de leur généraliste, qui est rémunéré à l'acte. Les spécialistes sont généralement rémunérés à l'acte, mais l'accès à leurs services suppose l'accord préalable d'un généraliste (qui, comme au Danemark, joue le rôle de gatekeeper).

Ainsi, les deux pays pour lesquels les incitations à limiter le volume des actes médicaux sont les plus faibles sont également ceux pour lesquels l'élasticité du volume des dépenses de santé à la densité médicale serait la plus forte (supérieure ou égale à l'unité).

De plus, la France est le seul pays pour lequel l'élasticité-prix est significativement négative. Cette élasticité agrégée négative, dans la mesure où le degré de couverture des dépenses par les assurances (Sécurité Sociale et assurances complémentaires) est élevé, ne traduit pas forcément la réaction de la demande aux variations du prix. Ceci pourrait s'interpréter comme une réaction de l'offre confrontée à une structure de prix administrée : afin de maintenir son chiffre d'affaires, l'offre de soins pare la baisse des prix unitaires en stimulant la demande via son pouvoir d'induction (pour autant que l'on en accepte l'existence). On pourra noter que l'élasticité-prix n'est pas significative pour des pays (Pays-Bas et Danemark) où les mécanismes de rémunération des producteurs incitent fortement à limiter la croissance du volume des prestations.

2.3 Approche avec données panelisées

Les résultats précédents suggèrent que les modes de rémunération des offreurs pourraient pour partie expliquer la forte variabilité du taux de croissance des dépenses de santé d'un pays à l'autre. Nous en cherchons maintenant la confirmation quantitative en adoptant une approche panelisée qui permet d'introduire comme régresseurs les différents types d'incitations subies par les offreurs. Par rapport à la section précédente, nous intégrons un plus grand nombre de variables explicatives (en contraignant les élasticités à être les mêmes pour chaque pays) et cherchons à améliorer la prise en compte du progrès technique.

2.3.1 Méthodologie

Nous reprenons d'abord la démarche de Gerdtham (1992, 1995) en fusionnant dans une même base les données relatives à vingt pays de l'OCDE : les quinze pays de l'Union Européenne, les Etats-Unis, le Canada, la Suisse, l'Australie et l'Islande. En définitive, nous disposons de 477 points, correspondant à la période 1970-1993, exception faite de l'Allemagne (1970-1990).

L'approche par les panels nous amène à faire le choix d'un taux de change ; nous disposons à cet effet des Parités de Pouvoir d'Achat (PPA) en dollars US en 1990, calculées soit par rapport au prix du PIB, soit par rapport au prix des dépenses de santé.

Les valeurs de ces deux types de PPA diffèrent souvent notablement, en particulier à cause du prix élevé des dépenses de santé aux Etats-Unis.

Tableau 5 : Valeur d'un dollar US en unités monétaires nationales (1990)

Pays	D	DK	F	I	NL	UK
PPA PIB	2,09	9,39	6,61	1421	2,17	0,60
PPA Santé	1,82	8,60	4,49	1087	1,69	0,42

Nous retenons dans un premier temps un nombre limité de variables explicatives : le PIB par tête (en volume) et le prix relatif des dépenses de santé. Des indicateurs démographiques (parts respectives des moins de quatre ans et des plus de soixante-cinq ans dans la population totale, ces populations étant a priori plus grosses consommatrices de soins) s'étant avérés non significatifs, nous tentons de prendre en compte grossièrement l'impact de la démographie en corrigeant la variable expliquée selon l'hypothèse suivante : les moins de quatre ans consommeraient deux fois plus de soins de santé que le reste de la population, et les plus de soixante-cinq ans trois fois plus.

Cette approximation grossière vise à prendre (très partiellement) en compte l'impact de la démographie (compliqué du reste par les effets de génération : aujourd'hui, la consommation des plus âgés relativement à la consommation moyenne tend à s'accroître). Il est intéressant d'examiner la sensibilité de nos résultats à cette variante.

En définitive, chaque modèle peut être estimé quatre fois, selon que les dépenses de santé sont estimées en PPA PIB ou en PPA Santé, et selon qu'elles sont affectées d'une correction démographique ou non. En réalité, il est apparu qu'en présence d'effets fixes par pays, le choix de la parité de pouvoir d'achat a peu d'incidence sur les résultats des estimations.

2.3.2 Estimation du modèle de demande

Notons tout d'abord que l'impact de la démographie, sous les hypothèses (très fortes) que nous avons faites, est marginal. Le tableau ci-dessous compare le taux de croissance moyen annuel du volume des soins de santé à la croissance annuelle moyenne qu'expliquerait la déformation de la structure démographique (baisse de la part des plus jeunes et accroissement de celle des plus âgés). A titre comparatif, l'OCDE (1995) estime que le vieillissement explique 0,2% de croissance annuelle dans les années 80, 0,3% dans les années 90, et que ce chiffre pourrait se situer entre 0,4% et 0,7% entre 2000 et 2020.

Tableau 6 : Impact de la démographie

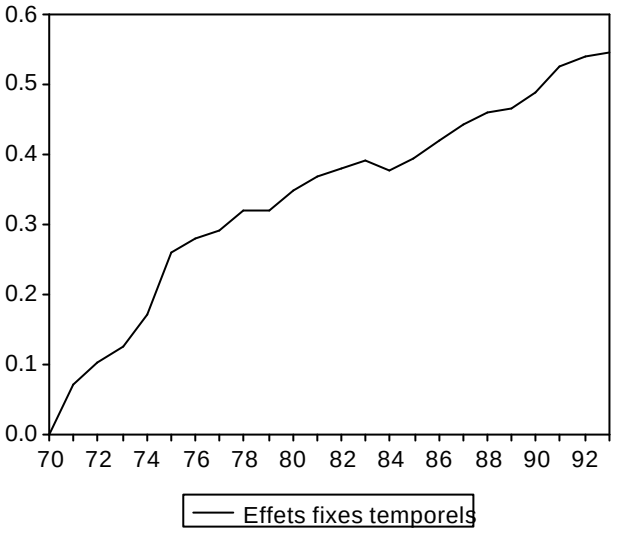
Pays	D	DK	F	I	NL	UK	US
Croissance annuelle moyenne	3,4%	2,4%	5,2%	4,2%	1,6%	3,0%	3,0%
Impact de la démographie	0,04%	0,15%	0,05%	0,18%	0,10%	0,11%	0,17%

La première régression est effectuée avec des effets fixes à deux niveaux (par année et par pays). santet désigne le logarithme du volume des dépenses de santé par tête, en PPA PIB ou PPA Santé, corrigées ou non des facteurs démographiques ; pibtet représente le logarithme du PIB par tête en volume, prixrel le logarithme du prix relatif des dépenses de santé :

$$santet_{i,t} = a\ pibtet_{i,t} + b\ prixrel_{i,t} + m_i + l_t + e_{i,t}$$

On estime alors l'élasticité-revenu à 0,76 et l'élasticité-prix à -0,77. L'observation des effets fixes temporels estimés dénote une tendance croissante, affectée d'une rupture vers 1975, donc contemporaine de l'inflexion générale de la croissance économique dans les pays de l'OCDE.

Graphique 7 : Estimateurs des effets fixes temporels



2.3.2 Intégration du progrès technique

Nous introduisons maintenant une mesure du progrès technique plus sophistiquée que précédemment. Nous avons vu que les dépenses en appareils thérapeutiques ne constituaient pas un indicateur satisfaisant de ce point de vue, notamment parce qu'elles ne prennent pas en compte les investissements coûteux réalisés en équipements lourds (scanners, RMN, etc.). En outre, les quelques données (souvent incomplètes) dont nous disposons sur ces types d'équipements ne permettent pas de remonter jusqu'au début de notre période d'estimation.

Nous considérons d'abord, dans la foulée de L'Horty (1997), les taux de dyalises rénales, dont les séries débutent en 1970. Dans la mesure où la prévalence des affections rénales peut en première analyse être considérée comme constante dans le temps et relativement homogène d'un pays à l'autre, les taux de dyalises reflètent assez fidèlement le progrès technique médical dans ce domaine. Cependant, nous faisons implicitement l'hypothèse que la croissance du taux des dyalises ne résulte pas tant de facteurs de demande que du progrès technique : à possibilités technologiques constantes, une couverture sociale plus étendue, de même qu'un revenu plus élevé, pourraient expliquer un recours plus fort aux dyalises (ce qui correspond à des effets de demande).

Tableau 7 : Scanners (par million d'habitants)

	Allemagne	Portugal	Espagne	Etats-Unis	France	Italie	Royaume-Uni
1990	10,1	4,6	7,2	25,2	6,7	n.d.	4,3
1993	12,9	n.d.	7,2	26,0	7,5	16,0	6,3

Tableau 8 : RMN (par million d'habitants)

	Allemagne	Portugal	Espagne	Etats-Unis	France	Italie	Royaume-Uni
1990	1,9	0,8	n.d.	7,8	0,8	n.d.	1,0
1993	3,7	n.d.	0,2	12,6	1,4	3,2	2,3

Tableau 9 : Greffes du coeur (par million d'habitants)

	Allemagne	Portugal	Espagne	Etats-Unis	France	Italie	Royaume-Uni
1993	7,1	0,9	7,3	9,0	10,4	4,6	5,8

Tableau 10 : Dyalises (par million d'habitants)

	Allemagne	Portugal	Espagne	Etats-Unis	France	Italie	Royaume-Uni
1993	297	247	264	506	270	296	81

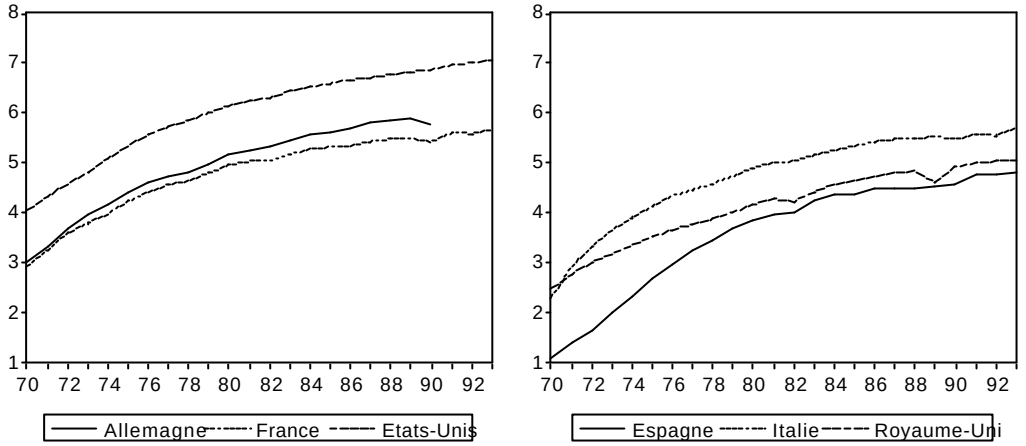
Nous utilisons également trois autres indicateurs, qui ne sont disponibles qu'à partir de la fin des années 1980 : le taux de greffes du coeur (susceptible d'encourir les mêmes critiques que le taux de dyalises quant au partage entre effets de demande et progrès technique) ; les taux d'équipement en scanners et en appareils à Résonance Magnétique Nucléaire (RMN), qui constituent une mesure plus directe (quoique partielle) du progrès technique. Notons que, suivant l'indicateur retenu, la France, en matière d'assimilation du progrès technique médical, se situe en pointe (c'est, avec la Belgique, le pays où l'on pratique le plus de greffes du coeur), dans la moyenne (cas des dyalises) ou dans le bas de la distribution (les taux d'équipement en scanners et en appareils de RMN sont relativement faibles). Nous proposons ici un indicateur composite pour minimiser cette variabilité.

Concrètement, nous extrayons la tendance commune (notée $\overline{dyal_t}$) aux vingt séries de taux de dyalises en appliquant un filtre de Hodrick-Prescott sur la moyenne des taux de dyalises nationaux $dyal_{i,t}$. Nous notons $e_{i,gc}$, $e_{i,scan}$, $e_{i,RMN}$ les taux relatifs de greffes du coeur, d'équipement en scanners et en appareils de RMN du pays i au début des années 1990, en normalisant à 1 chacun de ces indices pour la France. Nous construisons alors un indice de progrès technique :

$$PT_{i,t} = a\,.\log\left(dyal_{i,t}\right) + (1-a)\,.\log\left[(e_{i,gc}\,e_{i,scan}\,e_{i,RMN})^{1/3}\,\overline{dyal_t}\right]$$

Naturellement le choix des coefficients est largement arbitraire ; nous optons pour α=0,5. Notre indicateur donne alors des niveaux de diffusion du progrès technique relativement différenciés (cf graphiques 8a et 8b), mais des tendances plus proches qu'auparavant (cf. graphiques 3a et 3b).

Graphiques 8a et 8b : Evolution de l'indicateur de progrès technique



La formalisation est la suivante :

$$santet_{i,t} = a\ \ pibtet_{i,t} + b\ \ prixrel_{i,t} + d\ \ PT_{i,t} + l_{\ t} + m_i + e_{\ i,t}$$

Nous estimons une élasticité-revenu de 0,75 ; une élasticité-prix de -0,78 ; une élasticité à l'indicateur de progrès technique de 0,13. Le R² ajusté s'élève à 0,9705 (chiffre élevé traduisant vraisemblablement l'existence d'une relation de cointégration), mais les effets temporels ne sont plus significatifs au seuil de 5%. Le progrès technique capte maintenant l'essentiel du trend.

Nous ôtons donc les effets temporels et nous nous restreignons à la période 1975-1993, afin que l'estimation ne porte que sur la période postérieure au premier choc pétrolier (l'inflexion de la croissance économique a pu altérer profondément la dynamique de l'évolution des dépenses de santé et ses déterminants). Les estimations sont assez peu modifiées. Il ressort que la tendance croissante des dépenses de santé est expliquée pour l'essentiel par le PIB et l'indicateur de progrès technique.

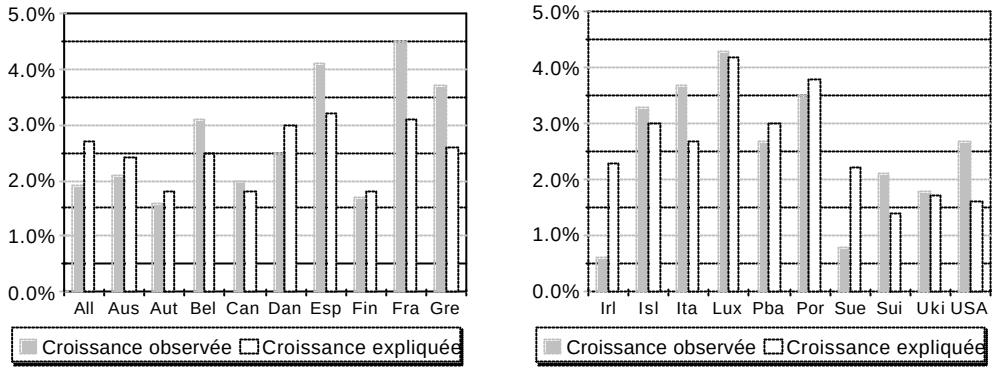
$$santet_{i,t} = a\ \ pibtet_{i,t} + b\ \ prixrel_{i,t} + d\ \ PT_{i,t} + m_i + e_{\ i,t}$$

Tableau 11 : Estimation du modèle général

Variables	PPA Santé
PIB	0,91
	(13,8)
Prix	-0,63
	(-9,7)
Progrès	0,12
Technique	(8,8)
R² ajusté	0,9704

Le progrès technique ressort de manière très significative. A titre de comparaison, si l'on enlève du modèle précédent l'indicateur technologique (et toujours sans effets fixes temporels), l'estimateur de l'élasticité-prix (-0,67) varie peu, ce qui n'est pas le cas de l'élasticité-revenu (1,40). Le R² ajusté fléchit à 0,9639.

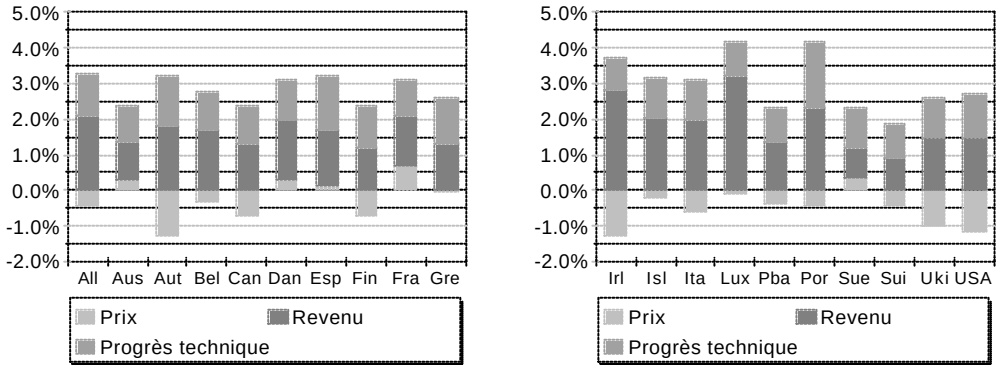
Graphiques 9a et 9b : Croissance annuelle réelle et expliquée



Toutefois, cette formalisation ne prend pas suffisamment en compte les spécificités structurelles de chaque pays. L'analyse des contributions des variables explicatives à la croissance du volume des soins dispensés (estimées par notre modèle) montre que pour plusieurs pays, il demeure une part importante d'inexpliqué.

La croissance des dépenses est largement surestimée en Suède et en Irlande, et à un degré moindre en Allemagne et au Danemark. A l'inverse, elle est largement sous-estimée en France, en Espagne, en Grèce, en Italie et aux Etats-Unis.

Graphiques 10a et 10b : Contribution des facteurs à la croissance



Nous essayons maintenant d'appréhender les différences de type institutionnel qui existent entre les systèmes de santé ; une première approche consiste à prendre en compte le **mode de rémunération** des prestataires de premier recours. Nous distinguons donc :

les pays où le mode de rémunération dominant est à l'acte : Allemagne, Australie, Autriche, Belgique, Canada, Etats-Unis, France, Grèce, Irlande (1970-1988), Italie (1970-1977), Luxembourg, Suisse.

les autres pays, où le mode de rémunération des généralistes comporte certaines incitations à l 'économie, que ce soit par le salariat ou la capitation : Danemark, Espagne, Finlande, Irlande (1989-1993), Islande, Italie (1978-1993), Pays-Bas, Portugal, Suède, Royaume-Uni.

Gerdtham (1995) utilise pour ce faire des variables muettes : ainsi sa variable pubinteg vaut 1 si le système du pays considéré est de type public intégré, et 0 sinon. Dans la mesure

où, tout au moins pour les pays européens, on observe plutôt de très forts contrastes dans les taux de croissance que des écarts constants en niveaux entre les différents pays[4], nous avons testé l'impact des différences d'organisation sur le **taux de croissance** plutôt que sur le **niveau** des dépenses. Aussi, nous introduisons des tendances linéaires dans le modèle et estimons une élasticité différente pour les pays du premier et du second groupe. Une démarche alternative consisterait à évaluer des élasticités au PIB et aux prix différentes pour chaque groupe.

$$santet_{i,t} = a\ \ pibtet_{i,t} + b\ \ prixrel_{i,t} + d\ \ PT_{i,t}$$
$$+ \left(g_1.\mathbf{1}_{acte} + g_2.\mathbf{1}_{capitation}\right).t + m_i + e_{i,t}$$

Alternativement nous pouvons reprendre la typologie employée par l'OCDE pour classer les différents systèmes de santé. On distingue donc les systèmes de type **intégré**, où le financement et les prestations incombent à un organisme unique ; les systèmes de **remboursement,** dans lesquels la délivrance d'une prestation médicale entraîne après coup le paiement : l'assureur rembourse le prestataire ou bien le patient si celui-ci est tenu de faire l'avance de frais (comme c'est le cas en France) ; les systèmes de **contrat**, dans lesquels un accord prospectif entre les tiers payants et les prestataires fixe les clauses et conditions du paiement des services. L'idée est que la nature très différente des incitations financières inhérentes à ces systèmes peut expliquer que les taux de croissance varient fortement d'un pays à l'autre.

Naturellement, aucun pays ne relève stricto sensu de l'une de ces trois catégories : ainsi, en France, la médecine ambulatoire relève essentiellement du système du remboursement, l'hôpital public du système intégré, et les cliniques privées du modèle contractuel.

Nous reprenons la classification de Gerdtham :

- système du remboursement : Australie, Belgique, Etats-Unis, France, Italie (1970-78), Luxembourg, Suisse ;

- système contractuel : Allemagne, Autriche, Canada, Espagne (1970-83), Grèce (1970-83), Pays-Bas, Portugal (1970-77) ;

- système intégré : Danemark, Espagne (1984-93), Finlande, Grèce (1984-93), Irlande, Islande, Italie (1979-93), Portugal (1978-93), Royaume-Uni, Suède.

La formalisation est alors la suivante :

$$santet_{i,t} = a\ \ pibtet_{i,t} + b\ \ prixrel_{i,t} + d\ \ PT_{i,t}$$
$$+ \left(g_1.\mathbf{1}_{remboursement} + g_2.\mathbf{1}_{contrat} + g_3.\mathbf{1}_{\text{int}égré}\right).t + m_i + e_{i,t}$$

L'inclusion de ces tendances linéaires réduit donc les élasticités au revenu et au progrès technique : l'élasticité-revenu vaut alors environ 0,65 (ce qui est cohérent avec la valeur de 0,7 retenue par Gerdtham lorsqu'il prend en compte les spécificités institutionnelles).

[4] Ainsi, en 1974, la part des dépenses de santé dans le PIB est respectivement de 6,3% et 5,3% en France et au Royaume-Uni. En 1997, les chiffres correspondants sont de 9,9% et 6,7%.

Tableau 12 : Impact du mode de rémunération

Variables	*PPA Santé*
PIB	0,66
	(8,1)
Prix	-0,63
	(-10,0)
Progrès Technique	0,05
	(2,7)
γ_1	0,0128
	(5,2)
γ_2	0,0113
	(4,2)
R^2 ajusté	0,9724

Tableau 13 : Impact du mode d'organisation

Variables	*PPA Santé*
PIB	0,64
	(7,9)
Prix	-0,58
	(-9,2)
Progrès Technique	0,07
	(3,4)
γ_1	0,0152
	(6,2)
γ_2	0,0084
	(3,1)
γ_3	0,0093
	(3,5)
R^2 ajusté	0,9735

Le modèle avec prise en compte du mode de rémunération des prestataires de premier recours implique un accroissement annuel autonome du volume de soins de 1,28% lorsque le mode dominant est la paiement à l'acte, contre 1,13% lorsque la capitation ou le salariat sont la règle, soit une différence de seulement 0,15 point. La prise en compte plus globale des spécificités institutionnelles (cf tableau 13) suggère des écarts plus importants. En effet, l'accroissement annuel autonome dans les systèmes de remboursement s'élèverait à 1,52%, soit respectivement 0,59 et 0,68 point de plus que dans les systèmes intégrés et contractuels.

On peut noter que les systèmes intégrés et contractuels comprennent des garde-fous à la croissance des dépenses : dans un système intégré, l'enveloppe budgétaire est allouée ex ante et contraint les producteurs. Dans un système contractuel, la contrainte budgétaire joue également sur les producteurs même si elle résulte d'une négociation entre assureurs et prestataires et n'est pas imposée par l'assureur universel. Bien que fragiles, les estimations tendraient à montrer qu'une certaine forme de responsabilisation financière de l'offre de soins a un impact négatif sur le taux de croissance des dépenses de santé.

Conclusion

L'analyse des déterminants des dépenses de santé, à la suite des travaux de Newhouse (1973), s'est longtemps focalisée prioritairement sur les aspects de demande. Toutefois, il ne semble pas que la dynamique de la consommation de soins soit causée presque exclusivement par l'effet revenu. Ainsi le Danemark et l'Autriche consacrent aujourd'hui une part de leur richesse nationale aux dépenses de santé sensiblement plus faible (6,4% et 7,9% respectivement en 1996) que le Portugal (8,2%), pourtant beaucoup moins riche. Dans ce contexte, la priorité est aujourd'hui à l'analyse des facteurs d'offre (organisation des systèmes de santé, modes d'intégration des nouvelles technologies médicales, impact de l'induction de la demande), qui fait l'objet de nombreuses études sur données microéconomiques.

Une analyse pays par pays semble indiquer que le poids de la demande induite par les prestataires varierait beaucoup d'un pays à l'autre, et pourrait expliquer une part importante de la variabilité de la dynamique des dépenses entre les pays. Pour confirmer cette conjecture, nous nous sommes efforcés également d'estimer un modèle macroéconomique unique sur plusieurs pays de l'OCDE, afin de distinguer deux types d'effets : les effets de demande (revenu et prix), dont il peut sembler légitime de supposer qu'ils varient peu d'un pays à l'autre, et les effets d'offre, qui résultent en grande partie de l'organisation des systèmes de santé.

Les résultats demeurent fragiles. Il apparaît toutefois que le mode d'organisation du système de santé (contractuel, intégré ou de remboursement) a un impact différencié sur la croissance des dépenses. En particulier, toutes choses égales par ailleurs (revenu, prix et progrès technique), les pays dont le système de santé fonctionne principalement sur le mode du remboursement (notamment en médecine ambulatoire) connaîtraient une croissance annuelle moyenne des dépenses supérieure de 0,6 point à ce que l'on observe dans les pays où le système de santé est de type contractuel ou intégré. Sur la période d'estimation (1975-1993), ceci explique un différentiel d'augmentation supérieur à 11%. L'interprétation de l'impact de ces facteurs institutionnels sur la croissance des dépenses exige toutefois une approche microéconomique complémentaire (Bocognano et al., 1998).

Bibliographie

BLOCH L., RICORDEAU P. (1996), La régulation du système de santé en France, Revue Française d'Economie, vol 11, n°1, pp 87-146.

BLOMQVIST A. (1991), The doctor as double agent : information asymmetry, health insurance and medical care, Journal of Health Economics, vol.10.

BOCOGNANO A., COUFFINHAL A., GRIGNON M., MAHIEU R., POLTON D. (1998) : Mise en concurrence dans le domaine de la santé : théorie et bilan des expériences étrangères, CREDES 1243, novembre.

CUTLER David M., MCCLELLAN Mark, NEWHOUSE Joseph P., REMLER Daniel, Are medical prices declining?, NBER working paper 5750, septembre.

CUTLER David M., SHEINER L. (1997), Managed care and the growth of medical expenditures, NBER working paper 6140, août.

DIONNE G., CONSTANDRIOPOULOS A.P. (1984), Doctors and their workshops : a review article, Département d'administration de la santé, Université de Montréal.

DRANOVE D. (1988), Demand inducement and the physician/patient relationship, Economic Inquiry, vol. 24.

ENGLE R.F., GRANGER C.W.J., (1987), Cointegration and error correction : representation, estimation, and testing, Econometrica, vol 55, n°2, mars, pp 251-276.

FULLER W.A. (1976), Estimation and hypothesis testing for non stationary time series, Ph.D. Thesis, Iowa State University.

GERDTHAM U.G., JÖNSSON B.(1991), Price and quantity in international comparisons of health care expenditure. Applied Economics n°23, pp 1519-1528.

GERDTHAM U.G., SIÖGAARD J., ANDERSSON F.(1992), An econometric analysis of health care expenditures : a cross-section study of the OECD-countries. Journal of Health Economics n°11, pp 63-84.

GERDTHAM U.G., JÖNSSON B., MACFARLAN M., OXLEY H.(1995), Factors affecting health spending : a cross-country econometric analysis. OECD, New directions in health care policies : improving cost control and effectiveness. OECD, Paris.

HAMILTON J.D., Time series analysis, Princeton University Press, 1994.

HOLMSTROM B., MILGROM P. (1991), Multitask Principal-Agent analyses : incentives contracts, asset ownership and job design, Journal of Law Economics and Organization, vol.7, pp 24-51.

HOURRIEZ J.M.(1992), Les évolutions de la demande des ménages : facteurs économiques et démographiques, INSEE, Division Conditions de vie des ménages, mai.

JACOBZONE S., ROCHAIX L. (1997), L'hypothèse de demande induite : un bilan économique. Economie et Prévision n°129-130.

JOBERT T. (1993), Tests de stationnarité, cointégration et modélisation multivariée, La persistance du chômage, P.Y. Hénin, Economica, pp 259-289.

LEU R.E. (1986), The public-private mix and international health care costs. A.J. Culyer et B. Jönsson, Public and private health services. Basil Blackwell, Oxford.

L'HORTY Y., QUINET A., RUPPRECHT F.(1997), Expliquer la croissance des dépenses de santé : le rôle du niveau de vie et du progrès technique. Economie et Prévision n°129-130.

MACKINNON J. (1991), Critical values for cointegration tests, in R.F. Engle and C.W.J. Granger : long-run economic relationships, Oxford University Press, pp267-276.

MOUGEOT M. (1994), Systèmes de santé et concurrence, Economica.

MURILLO C., PIATECKI C., SAEZ M.(1993), Health care expenditure and income in Europe. Health Economics, vol. 2, pp 127-128.

NEWHOUSE J.-P. (1977), Medical care expenditure : a cross-national survey. Journal of Human Resources, vol. 12, pp 115-125.

PARKIN D., MCGUIRE A., YULE B. (1987), Aggregate health care expenditures and national income : is health care a luxury good?.Journal of Health Economics n°6, pp109-126.

OCDE (1995), Nouvelles orientations dans la politique de santé, Etudes de politique de santé n°7, Paris.

PHILLIPS P.C.B., LORETAN M. (1991), Estimating long-run economic equilibria, Review of Economic Studies n°58, pp 407-436.

ROCHAIX L. (1993), Financial incentives for physicians : the Quebec experience, Health Economics, vol. 2, pp 163-176.

ROCHAIX L. (1995), La demande induite : pourra-t-on un jour en mesurer l'envergure?, Dossier Santé de la Commission des comptes et des budgets économiques de la nation.

ROCHAIX L. (1997), Asymétries d'information et incertitude en santé : les apports de la théorie des contrats, Economie et Prévision n°129-130.

Annexe 1 : Tests de stationnarité des séries

Notations

lsa désigne le logarithme du volume des dépenses de santé par habitant ;

lpm désigne le logarithme de la densité médicale ;

lpo désigne le logarithme du volume des dépenses en appareils thérapeutiques par médecin ;

lpx désigne le logarithme du prix relatif des dépenses de santé ;

d(y) désigne la différence première de la série y.

Méthodologie

Nous appliquons une stratégie de test séquentiel exposée par Jobert (1993). Le nombre k de retards est fixé de manière à blanchir les résidus. Pour ce faire, nous nous appuyons essentiellement sur la statistique de Ljung-Box, qui permet de tester les autocorrélations à un ordre élevé (guère plus de cinq en l'occurrence vu la faible taille des données). Les cinq modèles considérés sont les suivants :

Modèle 1 $$\Delta y_t = a_1 + b_1 t + y_1 y_{t-1} + \sum_{i=1}^{k} g_i \Delta y_{t-i} + e_t$$

Modèle 2 $$\Delta y_t = a_2 + b_2 t + \sum_{i=1}^{k} g_i \Delta y_{t-i} + e_t$$

Modèle 3 $$\Delta y_t = a_3 + y_3 y_{t-1} + \sum_{i=1}^{k} g_i \Delta y_{t-i} + e_t$$

Modèle 4 $$\Delta y_t = a_4 + \sum_{i=1}^{k} g_i \Delta y_{t-i} + e_t$$

Modèle 5 $$\Delta y_t = y_5 y_{t-1} + \sum_{i=1}^{k} g_i \Delta y_{t-i} + e_t$$

Les cases des tableaux de résultats donnent les statistiques de Student ; toutefois, les valeurs critiques à utiliser diffèrent de celles de la table standard et ont été tabulées par Fuller (1976).

France

Série	**Résultat**	Retards	Modèle 1			Modèle 2	Modèle 3		Modèle 4	Modèle 5
			a_1	b_1	y_1	b_2	a_3	y_3	a_4	y_5
lsa	**I(1)+T²**	2	2,33	1,13	-1,57	-2,87				
lpo	**I(0)+T+c**	3	-3,95	3,62	-4,01					
lpm	**I(1)+T²**	3	-4,24	0,30	-0,23	4,33				
lpx	**I(1)+T**	1	2,62	-2,70	-2,91	0,72	-0,64	-1,19	-2,17	
d(lsa)	**I(0)+T+c**	1	4,43	-2,87	-4,89					
d(lpo)	**I(0)+c**	1	1,36	-0,57	-3,11	0,26	1,82	-3,13		
d(lpm)	**I(0)+T+c**	2	-4,91	4,33	-4,63					
d(lpx)	**I(0)+c**	3	-3,80	1,99	-6,55		-4,20	-5,70		

Allemagne

Série	**Résultat**	Retards	Modèle 1			Modèle 2	Modèle 3		Modèle 4	Modèle 5
			a_1	b_1	y_1	b_2	a_3	y_3	a_4	y_5
lsa	**I(1)**	1	2,90	1,12	-2,25	-1,77	3,00	-2,76	1,48	1,07
lpo	**I(1)+T²**	1	-2,03	0,36	-2,45	-3,06				
lpm	**I(1)+T**	1	-2,71	-2,43	-2,41	-0,42	-2,60	0,32	-4,50	
lpx	**I(1)+T**	1	-2,50	-2,69	-2,62	0,50	1,21	0,01	2,61	
d(lsa)	**I(1)[a]**	1	1,83	-1,59	-2,46	0,05	0,92	-1,79	-0,68	-1,70
d(lpo)	**I(1)[a]**	1	2,33	-2,19	-2,92	-0,21	0,72	-1,74	-0,80	-1,82
d(lpm)	**I(0)+c**	1	-2,87	-0,26	-3,16	-0,22	-3,21	-3,26		
d(lpx)	**I(0)+c**	1	0,90	0,79	-4,73		3,32	-4,86		

Pays-Bas

Série	**Résultat**	Retards	Modèle 1			Modèle 2	Modèle 3		Modèle 4	Modèle 5
			a_1	b_1	y_1	b_2	a_3	y_3	a_4	y_5
lsa	**I(1)+T**	1	0,93	1,21	-0,89	1,64	-0,86	1,38	3,45	
lpo	**I(1)+T**	1	-1,17	1,65	-1,16	1,39	0,88	0,74	2,11	
lpm	**I(1)+T²**	1	-4,17	-1,21	-1,88	3,78				
lpx	**I(0)**	1	-1,40	1,34	-3,80		-0,40	-3,67		-3,83
d(lsa)	**I(0)+c**	1	0,50	2,00	-4,22		2,93	-3,43		
d(lpo)	**I(1)[a]**	1	-0,57	1,63	-3,09	-0,18	1,85	-2,52	0,14	-1,61
d(lpm)	**I(0)+T+c**	1	-3,21	2,81	-3,51					
d(lpx)	**I(0)**	3	-0,85	0,97	-2,33	1,53	0,43	-2,77	-0,72	-2,96

Danemark

Série	**Résultat**	Retards	Modèle 1			Modèle 2	Modèle 3		Modèle 4	Modèle 5
			a_1	b_1	y_1	b_2	a_3	y_3	a_4	y_5
lsa	**I(1)+T**	1	3,18	2,58	-3,05	-1,30	2,20	-1,98	2,96	
lpo	**I(1)**	1	-3,80	2,56	-3,87		-2,56	-2,60	0,72	-0,84
lpm	**I(1)+T²**	1	-2,49	-0,77	-1,66	2,03				
lpx	**I(0)**	1	0,00	0,02	-1,14	1,62	0,09	-2,04	-0,81	-2,27
d(lsa)	**I(0)+c**	1	3,22	-2,02	-5,63		3,02	-4,90		
d(lpo)	**I(0)**	1	1,14	-0,75	-5,49		1,11	-5,52		-5,38
d(lpm)	**I(0)+T+c**	3	-3,27	3,08	-3,63					
d(lpx)	**I(0)**	1	-1,59	1,39	-4,70		-0,77	-4,50		-4,48

Etats-Unis

Série	**Résultat**	Retards	Modèle 1			Modèle 2	Modèle 3		Modèle 4	Modèle 5
			a_1	b_1	y_1	b_2	a_3	y_3	a_4	y_5
lsa	**I(0)+T+c**	1	4,86	3,38	-3,82					
lpo	**I(1)**	1	-1,55	0,72	-1,56	0,05	-1,39	-1,40	0,70	-0,73
lpm	**I(1)+T**	1	-2,77	-2,40	-2,47	0,50	-2,05	-0,71	-5,17	
lpx	**I(1)+T²**	2	-2,48	3,24	-2,52	2,83				
d(lsa)	**I(0)+c**	1	3,30	-1,73	-4,35		3,16	-3,81		
d(lpo)	**I(0)**	1	-0,28	0,52	-3,89		0,53	-3,97		-4,02
d(lpm)	**I(0)+c**	4	-3,86	2,09	-4,74		-3,63	-3,95		
d(lpx)	**I(0)+c**	2	0,65	2,05	-4,08		3,30	-3,38		

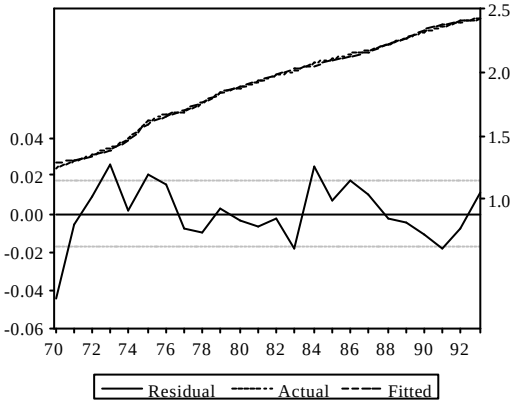
Italie

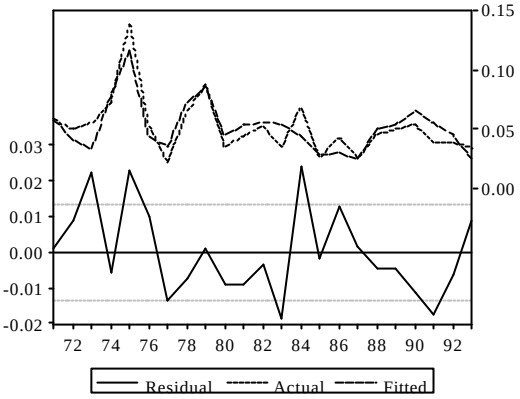
Série	**Résultat**	Retards	Modèle 1			Modèle 2	Modèle 3		Modèle 4	Modèle 5
			a_1	b_1	y_1	b_2	a_3	y_3	a_4	y_5
lsa	**I(1)+T²**	1	2,25	1,78	-2,17	-2,48				
lpo	**I(1)**	2	0,50	-0,53	-0,26	-1,73	2,20	-2,06	1,81	1,64
lpm	**I(1)**	1	1,44	-2,47	-2,83	0,68	-2,22	-1,41	-1,68	-0,15
lpx	**I(1)**	2	-3,24	3,13	-3,05	1,19	-0,78	-1,05	0,46	-0,85
d(lsa)	**I(0)+c**	1	3,68	-1,99	-5,89		3,72	-5,16		
d(lpo)	**I(0)**	2	2,68	-2,23	-4,28		1,58	-3,30		-2,79
d(lpm)	**I(0)**	2	-1,25	0,16	-3,24	0,21	-2,16	-3,35		-2,39
d(lpx)	**I(0)**	1	-0,96	1,29	-4,14		0,57	-3,87		-3,90

Annexe 2 : Relation de long terme et modèle à correction d'erreur

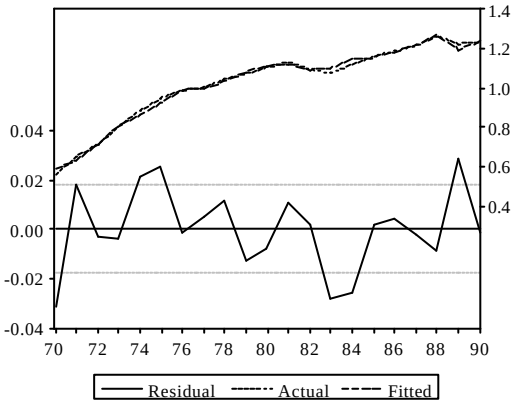
Pour chaque pays, le graphique de gauche donne l'ajustement de la relation de long terme par rapport à la variable observée et les résidus de la relation de long terme ; le graphique de droite donne le modèle à correction d'erreur (ECM) : série expliquée en différences premières, ajustement et résidus issus de l'ECM.

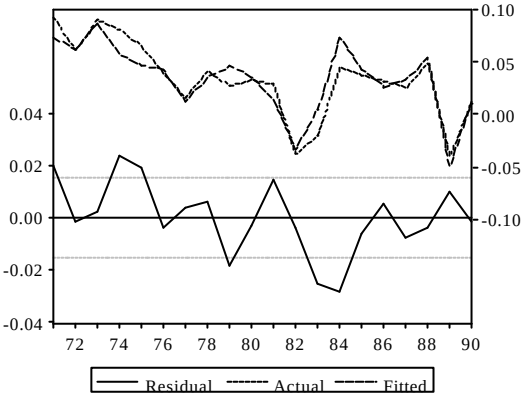
France



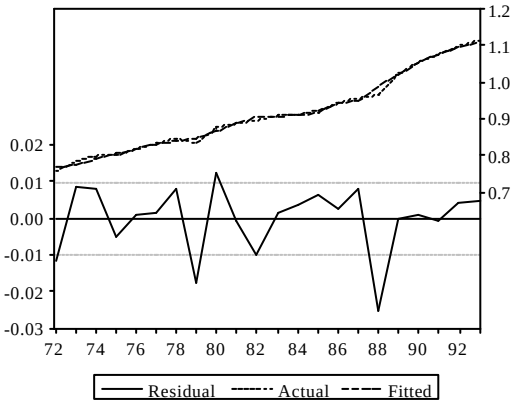


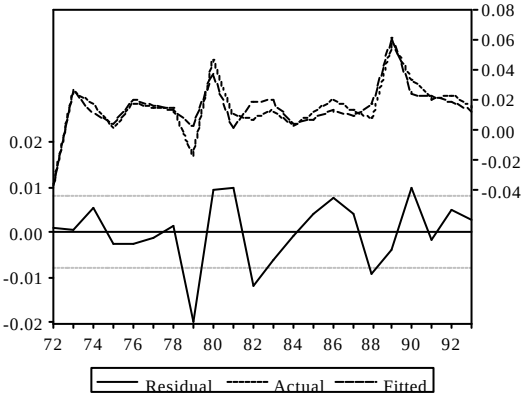
Allemagne



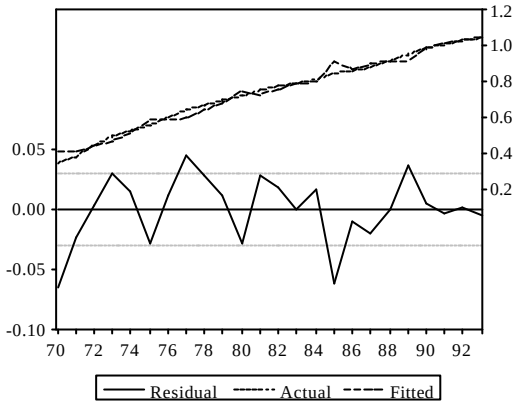


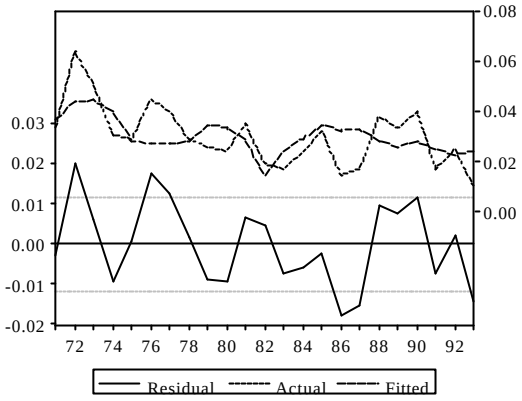
Pays-Bas



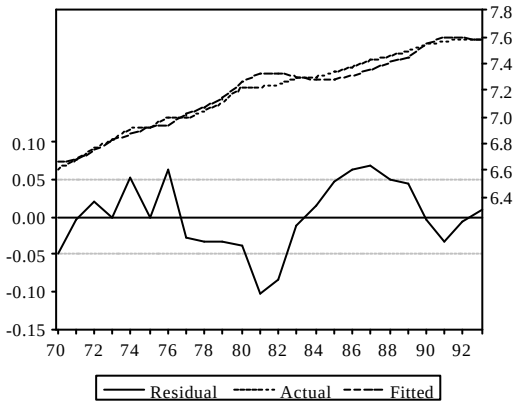


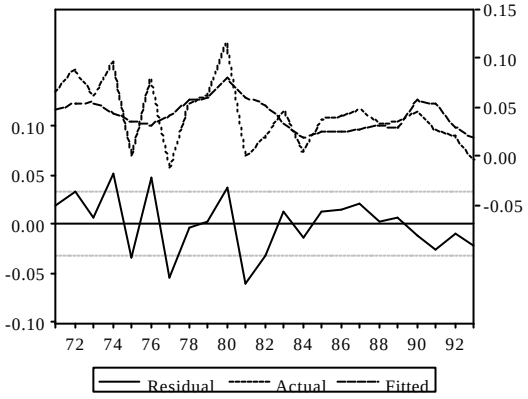
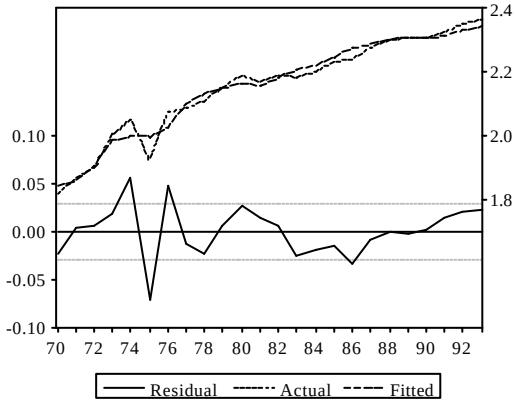
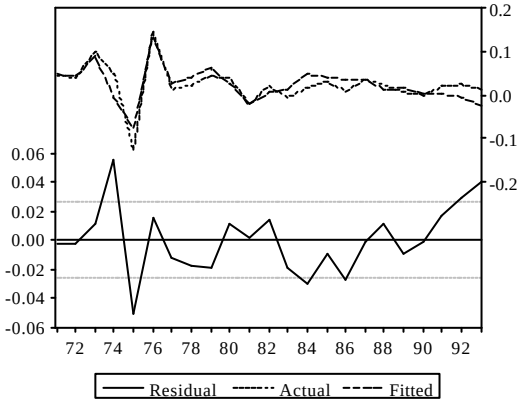
Etats-Unis





Italie





Danemark