Chapitre IV : Allocation optimale des ressources entre secteur public et secteur privé

Ce chapitre étudie uniquement le problème du montant de biens (et services) publics que doit **fournir** l'Etat. Il n'aborde pas la question du **mode de production** de ces biens. En particulier, on ne cherche pas à savoir si l'Etat doit produire lui-même en achetant des facteurs de production ou s'il doit acquérir directement des produits finis (en passant des marchés publics, en concédant au secteur privé la fourniture de certains services, etc.). A quelques exceptions près (la fourniture de sécurité notamment), cette question n'est pas logiquement liée à la nature privé ou publique du bien ou service produit. Elle est en fait plus de ressort de l'économie industrielle que l'économie publique au sens strict. A priori, comme le montre le tableau 4-1 ci-dessous, toutes les combinaisons sont possibles :

	FOURNITURE PRIVEE	**FOURNITURE PUBLIQUE**
PRODUCTION PRIVEE	La majorité des biens et services du marché	Ex : construction de routes, entretien d'équipements collectifs, etc.
PRODUCTION PUBLIQUE	Entreprises nationalisées	La majorité des biens (et services) publics

Tableau 4-1 : Les combinaisons possibles entre production privée/publique et fourniture privée/publique.

Section 1 : L'étendue de la dépense publique

On suppose que la société est composée de deux groupes, A et B, dont les fonctions d'utilité sont : $U_a(x_a,g)$ et $U_b(x_b,g)$. x_a et x_b sont les biens privatifs consommés respectivement par A et B. g est le bien public fourni par l'Etat. Pour simplifier on suppose que les prix des biens, qu'ils soient de type privatif ou de type public, sont tous égaux à 1. x_a et x_b correspondent alors aux montants des dépenses privées de A et de B, et g aux dépenses publiques.

La contrainte générale de ressources s'écrit : $y = x_a+x_b+g$, y étant le montant total des ressources disponibles dans la société.

On obtient un optimum parétien en maximisant U_a pour $U_b = \overline{U}_b$ donné et sous la contrainte générale de ressources.

D'où : $L = U_a(x_{a,g}) + \mu.[U_b(x_{b,g}) - \overline{U}_b] + \lambda.(y-x_a-x_b-g)$

avec :

(1) $\delta L/\delta x_a = \delta U_a/\delta x_a - \lambda = 0$

(2) $\delta L/\delta x_b = \mu.\delta U_b/\delta x_b - \lambda = 0$

(3) $\delta L/\delta g = \delta U_a/\delta g + \mu.\delta U_b/\delta g - \lambda = 0$

En divisant (3) par (1), et sachant que, d'après (1), $\delta U_a/\delta x_a = \lambda$ et que, d'après (2), $\mu = (\delta U_a/\delta x_a)/(\delta U_b/\delta x_b)$, on obtient :

$(\delta U_a/\delta g)/(\delta U_a/\delta x_a) + (\delta U_b/\delta g)/(\delta U_b/\delta x_b) = 1$

Soit :

$TMS_{g/xa} + TMS_{g/xb} = TMT_{g/(xa+xb)}$

Remarque : la linéarité de la contrainte budgétaire et le fait d'avoir normalisé les prix p_x et p_g à 1 ($y = x_a+x_b+g$) expliquent la valeur 1 du $TMT_{g/(xa+xb)}$ (1 de g en plus correspond à 1 de bien privé $[x_a+x_b]$ en moins).

On retrouve les mêmes résultats que dans n'importe quel problème d'optimum à cela près qu'il faut ajouter les TMS de g par rapport au bien privé dans les deux fonctions d'utilité. Cette sommation des TMS reflète le fait que le bien public intervient dans toutes les fonctions d'utilité individuelles. Dans le cas général d'une économie de consommation à n individus, avec p_x prix du bien privé et p_g prix du bien public, la condition de premier ordre pour un optimum s'écrit :

$$\sum_1^n TMS_{g/xi} = \frac{p_g}{p^x}$$

Cette condition d'optimalité parétienne en présence de biens publics est appelée le **théorème de Samuelson** (« théorème » exposé par cet auteur dans un article de 1954).

La condition précédente ($TMS_{g/xa} + TMS_{g/xb} = TMT_{g/(xa+xb)}$) permet de calculer, compte tenu de la contrainte générale de ressources et de la valeur fixée a priori pour U_b, le montant optimal de la dépense publique (et des différentes dépenses privées).

Le choix de $U_b = \overline{U}_b$ permet de sélectionner un optimum particulier sur la frontière d'efficacité sociale. Une valeur $\overline{U}_b$ est plus particulièrement intéressante : celle qui maximise le bien-être collectif. Pour la déterminer, il suffit de résoudre le problème de maximisation suivant :

Max $W(U_a(x_a,g),U_b(x_b,g))$ sous $y = x_a + x_b + g$

On obtient alors les deux conditions d'optimum suivantes :

$(\delta W/\delta U_a).(\delta U_a/\delta x_a) = (\delta W/\delta U_b).(\delta U_b/\delta x_b) = (\delta W/\delta U_a).(\delta U_a/\delta g) + (\delta W/\delta U_b).(\delta U_b/\delta g)$

où $(\delta W/\delta U_a).(\delta U_a/\delta x_a)$ est la *valeur sociale de l'utilité marginale de A* (c'est-à-dire la valeur que la société attribue à la consommation d'une unité supplémentaire de bien x par A), $(\delta W/\delta U_b).(\delta U_b/\delta x_b)$ la *valeur sociale de l'utilité marginale de B* et $(\delta W/\delta U_a).(\delta U_a/\delta g) + (\delta W/\delta U_b).(\delta U_b/\delta g)$ la *valeur sociale de l'utilité marginale de g pour A et B*.

On a trois équations (deux conditions d'optimum et une contrainte générale de ressources), soit une solution unique pour les trois variables (x_a, x_b et g)[37]. Cette solution correspond à « l'optimum optimorum ». On vérifie aisément qu'il s'agit bien d'un optimum de Pareto [c'est-à-dire que $TMS_{g/xa} + TMS_{g/xb} = TMT_{g/(xa+xb)}$]. La valeur obtenue pour $U_b(x_b,g)$ en ce point est celle qu'il aurait fallu choisir comme point de départ pour $\overline{U}_b$ dans le problème précédent pour que l'optimum atteint en maximisant U_a sous $U_b = \overline{U}_b$ corresponde également au maximum de W.

Section 2 : La répartition du coût de la dépense publique (analyse de Lindahl)

Le respect du **principe des coûts d'opportunité** est une condition *nécessaire* – mais pas suffisante - pour que l'allocation des ressources soit optimale. Il faut pour cela que l'unité marginale de ressource utilisée dans le secteur public soit payée à un prix égal à celui que le secteur privé est disposé à payer pour cette même ressource afin de produire et de fournir des biens privatifs. En d'autres termes, les dispositions marginales à payer pour les ressources de la société doivent être égales dans le secteur public et dans le secteur privé. Un budget déficitaire signifierait que, pour les mêmes ressources, la disposition marginale à payer du secteur public est inférieure à celle du secteur privé. Il y aurait donc trop de biens publics. En sens inverse, un budget excédentaire impliquerait, pour ces mêmes ressources, une disposition marginale à payer du secteur public supérieure à celle du secteur privé. Il y a donc pas assez de dépenses publiques. Dans les deux cas, on pourrait obtenir une amélioration parétienne en réallouant les ressources entre les deux secteurs, par un système d'impôts et de transfert tels que, finalement, le budget soit rééquilibré. On notera que cette règle d'équilibre budgétaire ne s'impose que dans une perspective d'allocation des ressources et dans une économie situé par ailleurs à l'équilibre. Elle ne dit rien quant à l'opportunité de recourir au déséquilibre budgétaire pour stabiliser une économie non équilibrée. De plus, cette règle s'applique de façon intertemporelle. Si l'égalité entre dépenses et recettes doit être respectée à tout moment lorsqu'il s'agit de dépenses publiques courantes (c'est-à-dire concernant uniquement des achats de services ou de biens de consommation), il en va tout autrement lorsque les dépenses correspondent à des investissements. Leur montant doit alors être égal à la somme actualisée des impôts que paieront les contribuables dans le futur, en échange des services que ces investissements permettront de fournir à chaque période. Cela implique à la fois un déficit sur les investissements nouveaux (et donc un financement par emprunt ou sur fonds publics préexistants) et des paiements excédentaires sur les investissements antérieurs, afin d'amortir les dettes contractées à cette occasion, en cas de financement par endettement, ou pour tenir compte du coût d'opportunité des fonds propres utilisés – en d'autres termes du manque à gagner lié au non-placement des fonds employés.

Pour simplifier et sauf précision contraire, on supposera par la suite que les dépenses publiques sont uniquement des dépenses courantes. L'équilibre budgétaire devra donc être respecté à chaque période.

Savoir que les recettes globales doivent être égales aux dépenses n'est pas suffisant pour déterminer une politique d'allocation budgétaire. Il faut pouvoir répondre aux deux questions suivantes :

- 1) quel est le montant souhaitable pour les dépenses publiques?
- 2) quelle est la répartition optimale du coût fiscal de ces dépenses entre les différents bénéficiaires.

L'analyse présentée dans la section 1 a montré comment un planificateur central pouvait répondre à la première question. L'analyse de Lindahl[38], que l'on va exposer maintenant, montre (1) comment il est possible d'apporter une réponse aux deux questions précédentes à la fois, et (2) comment ces réponses peuvent émerger de façon décentralisée (par marchandage entre les groupes sociaux plutôt que par l'intermédiaire d'un planificateur central).

a) le graphique de Lindahl

La société est composée de deux groupes homogènes d'individus, A et B. Leurs ressources avant impôt sont respectivement y_a et y_b (avec $y = y_a + y_b$).

[37] En supposant que la règle de Cramer s'applique (c'est-à-dire que le système se comporte comme un système linéaire).
[38] Economiste suédois (1891-1960), disciple de Knut Wicksell.

Soit g le montant de la dépense publique et t celui des impôts. Le budget doit être équilibré (respect du principe des coûts d'opportunité) : g = t.

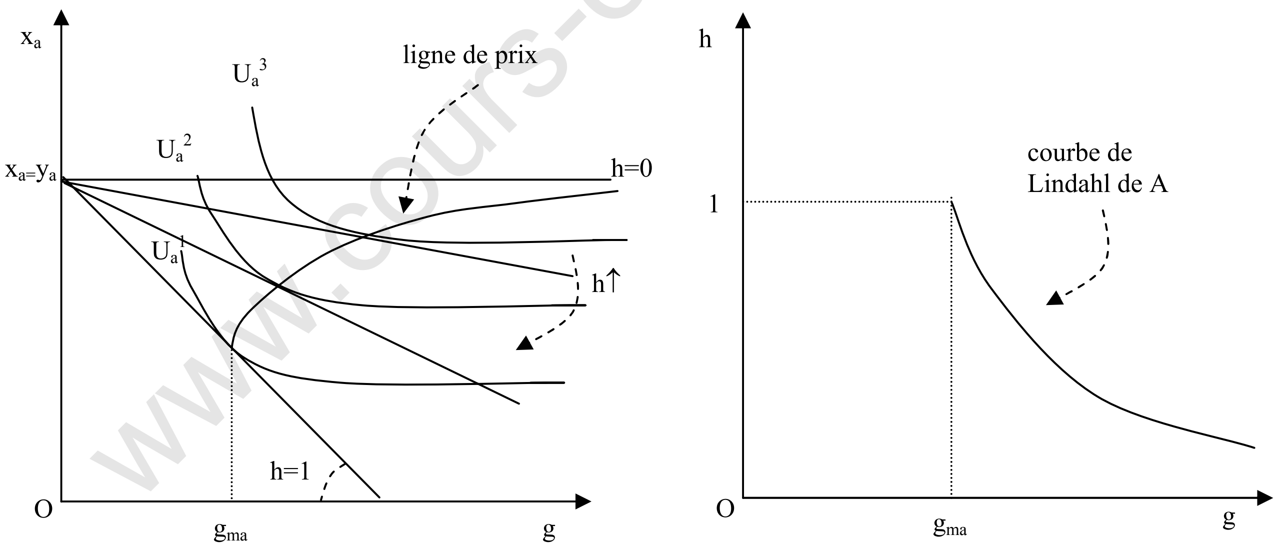
g est payé à h% par A et (1-h%) par B. Les contraintes budgétaires s'écrivent :

$y_a = x_a + h.g$ et $yb = x_b + (1\text{-}h).g$ (avec h compris entre 0 et 1).

x_a et x_b sont les quantités des biens privés consommées respectivement par A et B (on suppose que les agents privés n'épargnent pas).

La figure 4-1a montre les choix d'un individu rationnel représentatif du groupe A entre bien public et bien privé. La répartition optimale entre x_a et g pour A se situe au point de tangence entre la contrainte budgétaire pour h donné et la plus haute courbe d'indifférence. Selon la valeur de h, la pente de la contrainte budgétaire est modifiée, et donc le point d'équilibre. Quand h = 1 (A paie 100% de la dépense publique), on se situe au point g_{ma} : la contrainte budgétaire, $x_a = -h.g + y_a$, est de pente -1 et l'ordonnée à l'origine est égale à y_a. Lorsque h diminue, la contrainte budgétaire remonte vers l'horizontale en tournant autour du point fixe y_a. Quand h tend vers 0, la demande de g par A tend vers l'infini (puisque g ne coûte plus rien à A - et sous réserve de non-saturation).

Le lieu des points d'équilibre pour les différentes valeurs de h correspond à une ligne de prix, dont on peut déduire la courbe de demande par A de g, ou « courbe de Lindahl de A », g = $g_a(h,y_a)$ (figure 4-1b).



4-1a 4-1b

Figure 4-1 : Demande de bien public et courbe de Lindhal de A

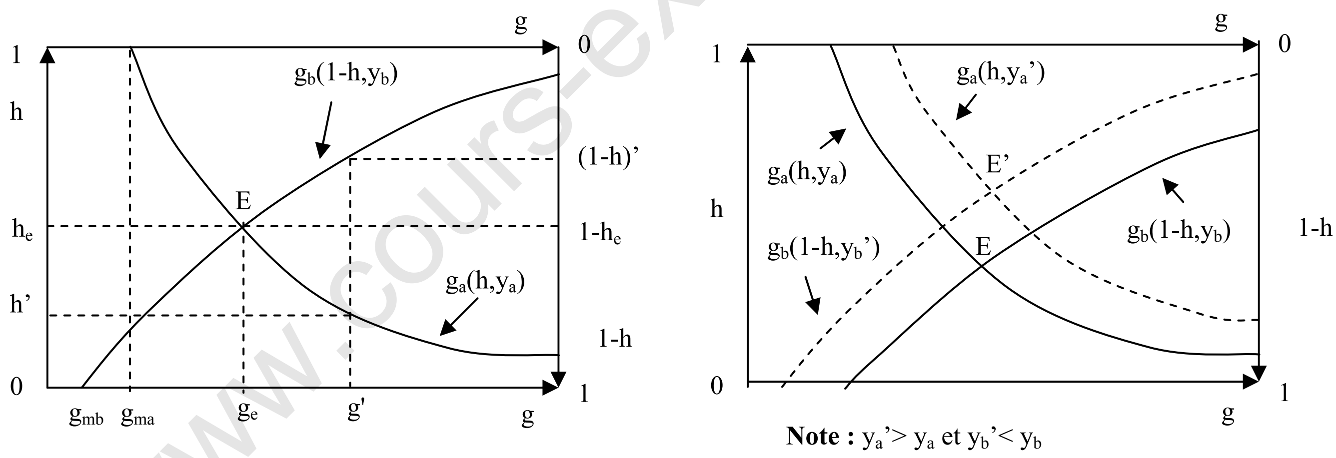
On peut établir de la même manière la courbe de Lindahl de B, g = $g_b(1\text{-}h,y_b)$. Si l'on représente ces deux courbes sur une même figure (figure 4-2a, dite « graphique de Lindahl »), il est possible de déterminer le point d'équilibre E. En ce point, appelé « solution de Lindahl », les deux groupes s'accordent à la fois sur le montant de la dépense publique (g_e) et sur la répartition de son financement (h_e par A et 1-h_e par B).

On notera que :

- les axes verticaux sont orientés en sens inverse (toute augmentation de la part payée par A correspond à une diminution équivalente de la part payée par B);

- les axes horizontaux sont orientés dans le même sens, puisque la quantité de g consommée par A est la même que celle consommée par B. Cela provient de ce que, à la différence de ce qui se passe pour un bien privé, les individus ne sont pas rivaux pour la consommation de g. Ils le sont seulement pour son financement. Cette non-rivalité implique toutefois un accord sur le montant fourni (avec $g = g_a(.) = g_b(.)$);

- les courbes de Lindahl dépendent des ressources dont disposent les groupes concernés. Si, par exemple, y_a augmente, la courbe de Lindahl pour A se déplace vers la droite, sous réserve que le bien g ne soit pas un bien inférieur pour A. Selon la même logique, si y_b diminue, y_b se déplace vers la gauche (B demande moins de g pour une même valeur de 1-h, toujours sous réserve que g ne soit pas un bien inférieur pour lui - figure 4-2b).



4-2-a

4-2-b

Figure 4-2 : Graphique de Lindahl

Si l'on suppose que les deux groupes possèdent une information parfaite sur leurs courbes de demande respectives, on peut penser que le point d'équilibre sera atteint par marchandage. Soit une situation déséquilibrée : sur la figure 2-14a, une valeur possible de la dépense publique est annoncée. A accepte de payer h' et paie (1-h)'. Le problème est que la somme de h et de (1-h)' est inférieure à 1. A et B ne sont pas prêts à financer la totalité d'un montant g'. Une valeur inférieure sera alors annoncée, afin de rapprocher la somme des dispositions à financer de l'unité. Après un certain nombre de marchandages de ce type, et sous réserve que le processus soit convergent, le point d'équilibre g_e sera atteint.

Il reste à montrer que cette solution de Lindahl correspond à un optimum de Pareto, ce que l'on va faire algébriquement.

b) La solution de Lindahl : analyse algébrique.

Les courbes de demande de g par A et B sont obtenues en maximisant les utilités de A et B sous leurs contraintes budgétaires respectives.

Pour A : Max $U_a(x_a,g)$ sous $y_a = x_a + h.g$

$L = U_a(x_a,g) + \alpha.(y_a - x_a - h.g)$

avec : $\delta L/\delta x_a = \delta U_a/\delta x_a - \alpha = 0$

$\delta L/\delta g = \delta U_a/\delta g - \alpha.h = 0$

d'où : $\delta U_a/\delta g = h\ \delta U_a/\delta x_a$ (1)

A atteint son maximum d'utilité quand l'utilité marginale de g, dont il ne paie que h%, est égale au $h^{ème}$ de l'utilité marginale du bien privé, dont il paie la totalité. En d'autres termes, l'impôt optimal pour A doit être égal au taux marginal de substitution entre le bien public et le bien privé pour l'individu (ou le groupe) A $((\delta U_a/\delta g)/(\delta U_a/\delta x_a))$.

La fonction de demande $g = g_a(h,y_a)$ est obtenue en combinant l'équation (1) et la contrainte budgétaire de A pour éliminer x_a. Dans le cas général, $\delta g_a/\delta h < 0$ et $\delta g_a/\delta y_a > 0$.

Pour B, Max $U_b(x_b, g)$ sous $y_b = x_b + (1-h).g$ donnerait de même une fonction de demande $g = g_b(1-h,y_b)$ avec, dans le cas général, $\delta g_b/\delta(1-h) < 0$ (soit $\delta g_b/\delta h > 0$) et $\delta g_b/\delta y_b > 0$.

Les valeurs d'équilibre de g et de h sont obtenues en égalisant g_a et g_b. Ce point est un optimum de Pareto, car, pour y_a et y_b donnés, toute autre valeur de g ou de h réduirait soit l'utilité de A soit celle de B.

Analytiquement, on a :

$TMS_{g/xa} = (\delta U_a/\delta g)/(\delta U_a/\delta x_a) = h$ et $TMS_{g/xb} = (\delta U_b/\delta g)/(\delta U_b/\delta x_b) = 1-h$

D'où : $TMS_{g/xa} + TMS_{g/xb} = h + 1-h = 1$ (soit le $TMT_{g/(xa+xb)}$)

c) Solution de Lindahl et maximum de bien-être collectif

A chaque distribution des revenus (à chaque valeur de $y_a°$ et de $y_b° = y - y_a°$), la solution de Lindahl fait correspondre un optimum parétien différent. On peut alors calculer les valeurs y_a* et $y_b*= y - y_a*$ qui permettent, par un marchandage de Lindahl, d'obtenir l'optimum parétien qui maximise la fonction de bien-être collectif, W. Pour passer de la distribution initiale à cette distribution optimale, un transfert de ressources entre A et B est nécessaire. Soit tr le montant de ce transfert (de cette redistribution) :

$tr = y_a° - y_a* = - (y_b° - y_b*)$. Si $tr = y_a° - y_a*$ est positif, il s'agira, pour un revenu global constant, y, de transférer des ressources de B vers A (et inversement, de A vers B si $tr = y_a° - y_a*$ est négatif).

La politique idéale de maximisation de W comporte alors deux étapes :

- elle effectue d'abord une redistribution des ressources entre les groupes;
- elle laisse ces groupes déterminer g et h par marchandage volontaire à la Lindahl.

Pour une fonction de bien-être collectif de type individualiste par exemple, $W=W[U_a(x_a,g), U_b(x_b,g)]$, la maximisation de W sous la contrainte globale de ressources ($y_a+y_b = x_a + x_b + g$) permet de déterminer les valeurs optimales $g*$, x_a* et x_b*.

Les maximisations de $U_a(x_a,g)$ sous $y_a + tr = x_a + h.g$ et de $U_b(x_b,g)$ sous $y_b\text{-}Tr = x_b + (1\text{-}h).g$ permettent d'obtenir les courbes de Lindahl $g_a(h,y_a+tr)$ et $g_b(1\text{-}h,y_b\text{-}tr)$. Pour chaque valeur de tr on obtient un point d'intersection différent (une solution de Lindahl différente). Il suffit alors de choisir la valeur de tr qui correspond à g*. En d'autres termes, les variations de tr permettent de sélectionner parmi tous les optima parétiens obtenus par marchandage volontaire celui qui correspond au maximum de bien-être collectif.

On notera enfin que l'analyse de Lindahl repose sur plusieurs hypothèses restrictives. Tout d'abord, les effets revenu de la fiscalité allocative sont considéré comme négligeables. Ensuite, chaque groupe considère que la part fiscale de l'autre est donnée et ajuste la quantité qu'il demande en conséquence, en fonction de sa courbe de demande propre. Les deux groupes se comportent donc en *price-takers*, c'est-à-dire comme s'ils étaient en concurrence. Dans la réalité, le résultat risque plutôt de dépendre, comme dans toute situation de marchandage bilatéral, des pouvoirs de négociation respectifs des joueurs et des conditions du jeu (coopératif ou non coopératif, avec des comportements symétriques de type Nash ou avec un leader et un suiveur – à la Stackelberg).

Section 3 : Les conséquences de la désagrégation des dépenses publiques

On suppose maintenant qu'il existe deux types de dépenses publiques, g_1 et g_2, avec :

- $U_a = U_a(x_a, g1, g_2)$ et $y_a = x_a + h_1.g_1 + h_2.g_2$
- $U_b = U_b(x_b, g_1, g_2)$ et $y_b = x_b + (1-h_1).g_1 + (1-h_2).g_2$

h_1 et h_2 sont les parts respectives des dépenses g_1 et g_2 prises en charge par A.

La maximisation de U_a sous la contrainte de revenu de A conduit à :

$$\delta U_a/\delta g_1 = h_1.\delta U_a/\delta x_a \quad (1)$$

et $$\delta U_a/\delta g_2 = h_2.\delta U_a/\delta x_a. \quad (2)$$

Celle de U_b sous la contrainte de revenu de B conduit à :

$$\delta U_b/\delta g_1 = (1-h_1).\delta U_b/\delta x_b \ (3)$$

et $$\delta U_b/\delta g_2 = (1-h_2).\delta U_b/\delta x_b \ (4)$$

On a au total six équations (deux contraintes budgétaires plus les quatre équations ci-dessus) pour déterminer six variables (x_a, x_b, g_1, g_2, h_1, h_2). Le système, s'il se comporte comme un système linéaire, est donc juste déterminé. A une répartition des ressources y_a et y_b donnée, correspond un seul équilibre de Lindahl, avec détermination des montants optimaux g_1 et g_2 et répartition optimale de leurs coûts respectifs (h_1 et h_2). Comme dans le cas précédent, un système de transfert permet, en modifiant la distribution des ressources, de se déplacer sur la frontière d'efficacité sociale (c'est-à-dire sur le lieu des optima parétiens), et donc de sélectionner l'optimum optimorum.

Supposons maintenant que, pour des raisons de simplification administrative par exemple, le gouvernement décide d'effectuer une répartition globale des coûts correspondant à la totalité des dépenses publiques (g_1+g_2).

Cela revient à poser une équation supplémentaire telle que $h_1 = h_2 = h$. En effet, c'est seulement dans ce cas que les contraintes budgétaires peuvent s'écrire :

$$y_a = x_a + h.(g_1+g_2) \text{ et } y_b = x_b + (1-h).(g_1+g_2).$$

Mais cette équation supplémentaire fait que le modèle devient surdéterminé. Il n'y a plus d'accord optimal possible entre les groupes A et B si l'on essaie de répartir globalement la charge de la dépense totale, g_1+g_2. Au moins l'un des groupes n'est pas sur l'une de ses courbes de demande (pour g_1 et/ou pour g_2) et la collectivité doit supporter une perte sociale sèche. La répartition fiscale ne correspondra au mieux qu'à un « optimum de second rang ».

Le résultat précédent laisse à penser qu'il est préférable de déterminer systématiquement les impôts à un niveau désagrégé et d'affecter directement la fiscalité à chaque type de dépense. Cette conclusion contredit donc la règle traditionnelle de non-imputation des ressources aux dépenses, c'est-à-dire un principe de base du droit budgétaire qui existe dans plusieurs pays (dont la France). Cette contradiction peut s'expliquer par la prise en compte de facteurs autres que ceux introduits dans l'analyse théorique simplifiée qui précède :

- économiquement, la non-imputation permet de réduire divers coûts de gestion administrative (coûts d'information, coûts de calcul d'impôts trop différenciés, coûts de perception d'une fiscalité complexe);

- politiquement, la non-imputation permet à l'Etat de ne pas être obligé de suivre à tout moment les volontés individuelles et donc de conserver une marge de manœuvre importante dans l'utilisation des recettes budgétaires.

On observe toutefois que, depuis plusieurs années, le champ des ressources affectées a tendance à s'étendre de façon notable. Certains auteurs y voient le signe des réticences croissantes des populations face à l'accroissement de la pression fiscale : spécifier l'utilisation des impôts supplémentaire permettrait au gouvernement de les rendre plus facilement acceptables.

Section 4 : La neutralité fiscale

L'Etat doit percevoir des impôts pour financer tout à la fois les biens publics et ses interventions de correction de l'économie de marché. Cette contrepartie financière de la politique d'allocation des ressources est imposée par la nécessité de respecter le principe des coûts d'opportunité. Idéalement elle doit être « neutre »[39]. Dans le cas contraire, on risque de perdre en partie d'une main ce que l'on donne de l'autre : des actions visant à rapprocher l'économie de l'optimum seraient partiellement obérées par leur mode de financement (qui, du fait de sa non-neutralité, provoquerait des distorsions éloignant de l'optimum).

Il est facile de respecter la neutralité fiscale s'il existe des **impôts forfaitaires**, c'est-à-dire des impôts dont le prélèvement modifie seulement les dotations initiales en ressources des individus, sans changer la façon dont ils emploient ces ressources. S'il s'agissait d'effectuer un prélèvement une fois pour toutes, la plupart des impôts pourraient être considérés comme

[39] Par « fiscalité neutre », sans autre précision, on entend en économie une *fiscalité allocativement neutre*, c'est-à-dire une fiscalité qui permet de rester sur la frontière d'efficacité sociale – compte tenu du prélèvement fiscal global. En revanche, cette fiscalité peut parfaitement être *distributivement neutre*, c'est-à-dire correspondre à une distribution des ressources très différente de la distribution initiale (et donc à un point très différent sur la frontière d'efficacité sociale, avec des gagnants et des perdants relatifs, notamment lorsque la fiscalité est non proportionnelle aux ressources). Si la neutralité allocative est privilégiée dans la sphère économique, la neutralité distributive joue au contraire un rôle centrale dans la sphère politique.

forfaitaires. En revanche, dès qu'il s'agit de mettre en place un prélèvement permanent, des agents rationnels vont rapidement découvrir la logique de définition des différents impôts et réagir en conséquence. Le nombre des impôts potentiellement forfaitaires se restreint alors considérablement :

- même la capitation (impôt par tête), souvent citée comme exemple-type d'impôt forfaitaire, peut avoir des effets en longue période, sur la natalité notamment;

- la plupart des impôts cités comme forfaitaires sont inapplicables, parce qu'ils présentent trop d'inconvénients par ailleurs (notamment du fait de leurs effets redistributifs, souvent très régressifs).

Le problème est toutefois différent selon que l'on se place dans une économie de consommation ou dans une économie de production, où les marchés de facteurs sont explicitement pris en compte.

Dans une économie de pure consommation, les revenus des individus sont donnés. Tout impôt « direct », c'est-à-dire dont le montant est soit exogène, soit uniquement fonction de ces revenus données, sera de type forfaitaire. Les égalités marginales correspondant à l'optimum seront indépendantes de la valeur des revenus disponibles individuels (revenu disponible = revenu − impôts directs). La fiscalité amputera les ressources des individus mais par un simple déplacement de leur contrainte budgétaire parallèlement à elle-même mais sans modification de sa pente).

Dans une économie de production, où l'on tient compte explicitement des marchés de facteurs, le problème est différent (sauf si l'on suppose que les offres de facteurs de production sont parfaitement inélastiques). Les quantités de facteurs offertes dépendront de la rémunération nette de chacun d'entre eux, et donc du niveau de la fiscalité qui les frappe. La fiscalité directe n'est plus forfaitaire : par exemple, une hausse des impôts sur les revenus du travail modifie le salaire net (après impôts), et donc le prix relatif du travail par rapport au loisir, et donc la quantité finalement offerte.

a) Neutralité fiscale dans une économie de consommation

Par définition, la fiscalité directe est neutre dans ce type d'économie. Elle correspond à une simple modification exogène du revenu dont disposent les individus. Le seul problème devient alors celui de la fiscalité indirecte, c'est-à-dire de la fiscalité qui frappe chaque unité de bien au moment de sa vente.

1 - Conditions générales de neutralité de la fiscalité indirecte

On suppose que l'économie comprend deux biens, X et Y, produits selon la fonction de transformation $f(X,Y) = 0$ et utilisés dans des quantités $x°$ et $y°$ par le secteur public et x et y par le secteur privé. Par conséquent :

$X = x + x°$ et $Y = y + y°$

Pour simplifier, on fait l'hypothèse que les montants $x°$ et $y°$ sont fixés de façon exogène et que ce que fait l'Etat avec ces biens n'a pas d'influence sur la fonction d'utilité du consommateur représentatif. Cette fonction s'écrit donc :

$U = U(x,y)$

L'optimum de Pareto est obtenu en maximisant U sous f (en déterminant ce que ferait un planificateur central bienveillant). Cette maximisation implique que :

$$\frac{\delta U/\delta x}{\delta U/\delta y} = \frac{\delta f/\delta x}{\delta f/\delta y} = \frac{\delta f/\delta X}{\delta f/\delta Y} \qquad (1)$$

La seconde égalité résulte de l'hypothèse d'exogénéité de $x°$ et $y°$. Si les biens x et y sont maintenant vendus sur un marché parfait à des prix hors impôts p et q, les producteurs maximisant leur profit vont égaliser TMT et prix relatifs, soit :

$$\frac{\delta f/\delta X}{\delta f/\delta Y} = \frac{p}{q} \qquad (2)$$

Pour financer ses dépenses en $x°$ et $y°$, l'Etat a le choix entre la fiscalité directe (T_{dir}) ou la fiscalité indirecte (T_{ind}), qui consiste à prélever sur chaque unité vendue d'un bien au secteur privé un montant fixe d'impôt. Dans ce dernier cas, les prix payés par les consommateurs (P et Q) sont différents de ceux reçus par les producteurs (p et q), avec :

$P = p + t_x$ et $Q = q + t_y$

La contrainte budgétaire du secteur public s'écrit :

$p.x° + q.y° = T_{dir} + T_{ind} = T_{dir} + t_x.x + t_y.y$

et le problème pour l'Etat est de choisir des combinaisons de T_{dir}, t_x et t_y qui soient allocativement neutres.

Les consommateurs vont maximiser leur utilité sous contrainte de leur revenu (R), compte tenu des impôts directs (T_{dir}) qu'ils doivent payer et des prix impôts indirects inclus, P et Q :

Leur contrainte budgétaire devient : $R-T_{dir} = P.x + Q.y$. A l'équilibre, la maximisation de U sous cette contrainte implique que :

$$\frac{\delta U/\delta x}{\delta U/\delta y} = \frac{P}{Q} \qquad (3)$$

L'équilibre de marché ne respectera les conditions d'optimalité (équation 1) que si P/Q = p/q, du fait des équations 2 et 3. En d'autres termes, la fiscalité indirecte ne doit pas provoquer de distorsions dans les prix relatifs (préserver leur valeur).

En remplaçant P et Q par leur valeur, on obtient :

$P/Q = (p+ t_x)/(q+t_y) = p/q \,.\, (1+ t_x/p)/(1+t_y/q)$

P/Q ne sera égal à p/q que si :

- soit $t_x=t_y=0$ (on n'utilise que la fiscalité directe) ;
- soit $t_x/p = t_y/q$, ce qui correspond à un impôt proportionnel à taux unique sur tous les biens.

Une TVA à taux unique respecte cette condition de neutralité de la fiscalité, et c'est là un de ses principaux avantages. En revanche, cette propriété de neutralité disparaît lorsque la TVA comprend plusieurs taux[40].

[40] Le fait de percevoir la TVA *à chaque stade de production*, comme l'implique le système actuel, est intéressant du point de vue du contrôle des vendeurs par les acheteurs (car ces derniers veulent pouvoir déduire la TVA payée aux stades antérieurs de production). En revanche, cela n'a aucune effet spécifique du point de vue de la neutralité. Une taxe proportionnelle unique au stade de détail présenterait les mêmes avantages de neutralité. Elle serait plus simple à percevoir (puisqu'elle serait collectée en une seule fois), mais aussi plus facile à frauder.

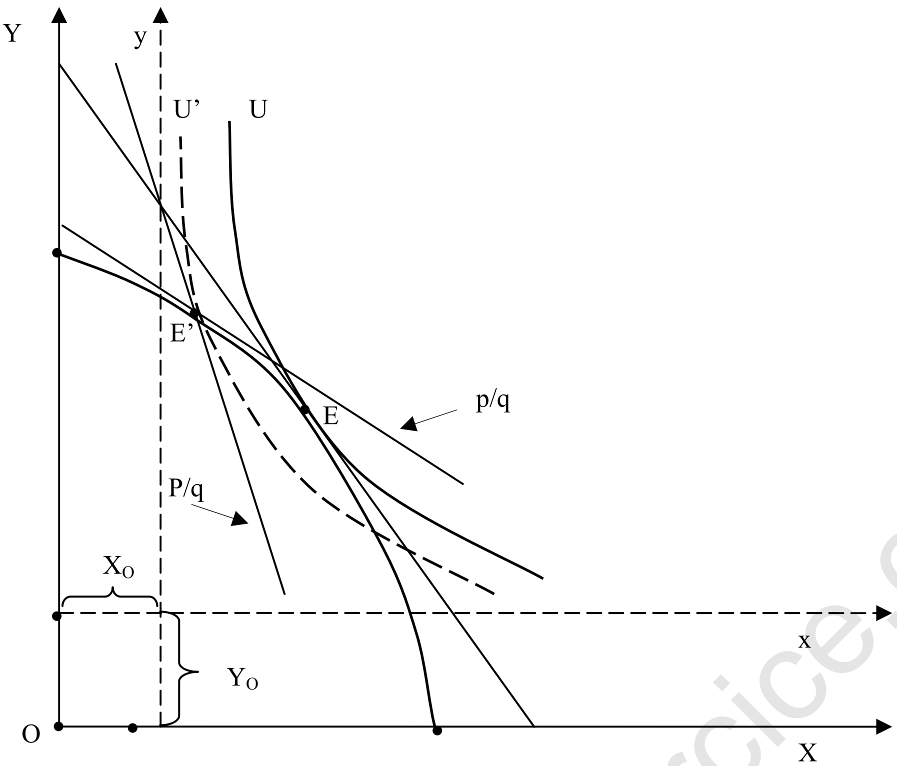


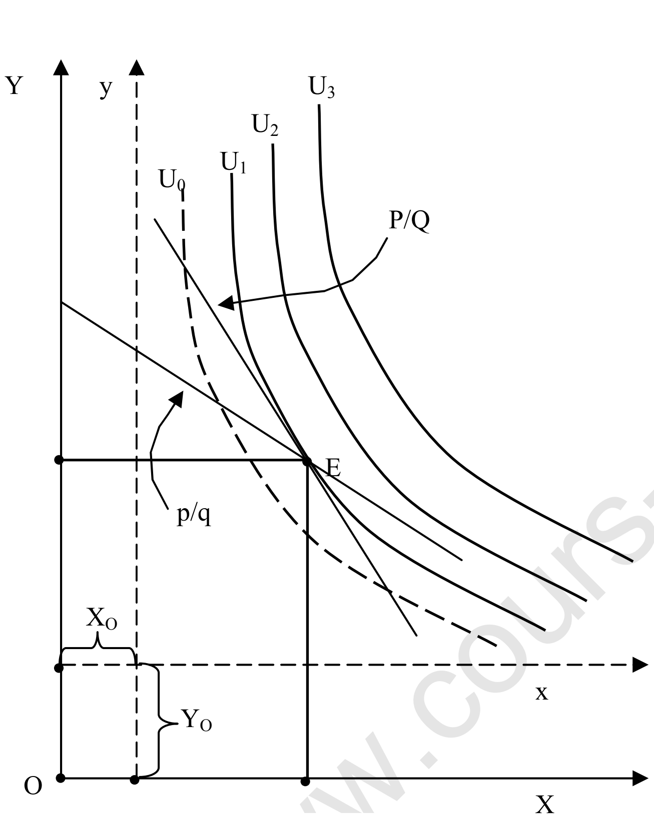
Figure 4-3 : Effets d'une fiscalité non neutre

La figure 4-3 montre comment les distorsions fiscales entraînent des pertes de bien-être. Une augmentation des dépenses publiques se représente par des déplacements des axes correspondant à x et y. La frontière des possibilités de production reste en revanche inchangée. Au point E les dépenses publiques sont financées soit par la fiscalité directe, soit par des impôts indirects neutres. La contrainte budgétaire se déplace parallèlement à elle-même (en fait ce sont les axes xoy qui se déplacent en sens inverse, mais cela revient au même). Le point d'optimum reste le même, quelle que soit la répartition de X et Y entre le secteur public (x°, y°) et le secteur privé (x,y).

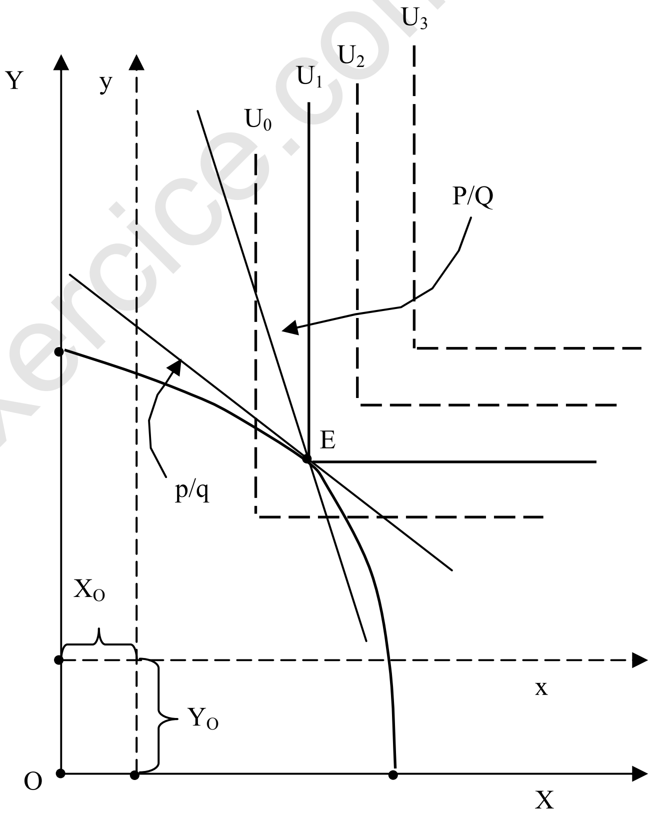
Supposons en revanche que, partant d'une fiscalité uniquement directe, on applique une fiscalité indirecte non neutre, touchant uniquement le bien x. ($t_y = 0$, $P=p+t_x$ et $Q=q$). La contrainte budgétaire conserve la même ordonnée à l'origine dans le plan xoy (si le consommateur ne consomme que y, il échappe à l'impôt). En revanche, sa pente ($P/q = -(p+ t_x)/q$) est plus forte en valeur absolue. Le nouveau point d'équilibre E' correspond à un niveau d'utilité inférieur pour un même montant d'achats publics x° et y° (avec des prix relatifs payés par les consommateurs - pente de la contrainte budgétaire - différents de ceux reçus par les producteurs - pente de la tangente à la courbe de transformation au point E').

2 - *Conditions particulières de neutralité*

Dans deux cas extrêmes, la distorsion des prix relatifs provoquée par un impôt indirect n'aura pas d'effet négatif sur le bien-être collectif. Il s'agit du cas de productions strictement jointes (ou liées) du côté de l'offre et de complémentarité stricte entre les deux biens du côté de la demande.



(4-4a)



(4-4a)

Figure 4-4 : Cas particuliers de neutralité fiscale

La figure 4-4a montre le cas de produits strictement joints, avec des courbes de transformation à angle droit. Quels que soient les prix relatifs obtenus par les producteurs (et donc que les impôts indirects soient ou non proportionnels), on reste au point E, les prix relatifs payés par les consommateurs demeurent les mêmes à l'équilibre. Le poids d'une fiscalité distorsive est donc supportée en totalité par les producteurs.

La figure 4-4b correspond au cas symétrique. La complémentarité concerne la demande : les courbes d'indifférence sont à angle droit. Quelles que soient les distorsions introduites, on reste au point E. Les prix relatifs obtenus par les producteurs sont les mêmes et le poids de la fiscalité distorsive est supporté en totalité par les consommateurs.

L'analyse précédente rappelle la règle traditionnelle selon laquelle il est préférable d'imposer les biens qui ont soit une offre soit une demande inélastique. Dans le cas précédent, il s'agit d'offres et de demandes relatives (d'inélasticité aux prix relatifs), mais on retrouve a fortiori les mêmes résultats en cas d'offres et de demandes absolues.

Considérons trois biens, X, Y et Z, avec $X=x+x°$, $Y=y+y°$, $Z=z+z°$ et une fonction de transformation $f(X,Y,Z)=0$.

Supposons que le bien Z soit un bien indispensable, dont une quantité définie z' doit impérativement être consommée (sinon l'utilité baisse considérablement), et tel que l'utilité n'augmente pratiquement pas quand l'on dépasse le seuil de consommation z’. L'exemple historique habituel est celui du sel, produit indispensable de conservation sous l'Ancien régime. Dans ce cas, seule la quantité z=z' est compatible avec la maximisation de l'utilité. La fonction U s'écrit alors :

$U = U(x,y)|_{z=z'}$

L'optimum implique que z=z' et que :

$$\frac{\delta U/\delta x}{\delta U/\delta y} = \frac{\delta f/\delta x}{\delta f/\delta y} = \frac{\delta f/\delta X}{\delta f/\delta Y}$$

Sans impôts indirects, x, y et Z sont vendus à des prix p, q et v.

$R\text{-}T_{dir} = p.x+q.y+v.z$ et $T_{dir} = p.x°+q.y°+v.z°$

L'équilibre du producteur implique que :

$$\frac{\delta f/\delta x}{p} = \frac{\delta f/\delta y}{q} = \frac{\delta f/\delta Z}{v}$$

Si l'on instaure des impôts indirects t_x, t_y, t_z, la contrainte budgétaire des consommateurs devient :

$P.x + Q.y + V.z = R\text{-}T_{dir}$

La maximisation de l'utilité par les consommateurs implique que :

$$\frac{\delta U/\delta x}{P} = \frac{\delta U/\delta y}{Q} \quad \text{et } z = z'$$

L'optimum n'est réalisé que si P/Q = p/q, mais le prix V du bien z après impôt est sans importance. En conséquence :

- si l'on a des biens dont la demande est élastique, les droits doivent être proportionnels. On ne peut avoir de droits spéciaux sur ces biens;
- en revanche, si la demande d'un bien est inélastique, on peut avoir n'importe quel droit spécial sur ce bien (comme le consommateur ne modifie pas sa demande, le droit est équivalent à une amputation de revenu, c'est-à-dire à un impôt direct).

On obtient un résultat analogue en cas d'inélasticité de l'offre, lorsqu'un montant défini du bien est produit ou disponible (Z=Z') et lorsqu'il n'existe aucune possibilité de substitution avec d'autres biens.

La fonction de transformation devient : f(X,Y) = 0 et Z = Z'

A l'équilibre de marché on a :

$$\frac{\delta U/\delta x}{P} = \frac{\delta U/\delta y}{Q} = \frac{\delta U/\delta z}{V} \quad \text{pour les consommateurs}$$

et

$$\frac{\delta f/\delta x}{p} = \frac{\delta f/\delta y}{q} \quad \text{et } \ Z = Z' \quad \text{pour les producteurs}$$

Là encore, l'optimum implique que P/Q = p/q mais il est indépendant de V (et de t_z).

L'inélasticité de l'offre a notamment été avancée pour justifier une imposition spéciale de la terre. Cependant, quand on décide d'instaurer un droit spécial sur un bien, d'autres considérations entrent en ligne de compte. Dans le cas de la terre, on peut penser que les objectifs allocatifs rencontrent les objectifs redistributifs (imposer la terre est considéré comme allocativement neutre et généralement considéré comme équitable). En revanche, pour les biens de consommation, neutralité allocative et équité semblent aller en sens inverse : les biens dont la demande est inélastique occupent en général une part plus importante dans le budget des ménages à revenu faible.

Ce qui a été dit pour les impôts indirects peut évidemment s'étendre au cas des subventions, c'est-à-dire à des valeurs négatives de t_x, t_y et t_z. Si l'on veut subventionner certains biens, pour des motifs de redistribution par exemple, il est préférable de choisir des biens dont la demande est inélastique, à condition que cela favorise effectivement le groupe que l'on souhaite avantager (cas des biens de première nécessité). En revanche, si la subvention concerne des biens avec une élasticité de demande non nulle, des transferts monétaires directs (qui sont équivalents à des impôts indirects négatifs) sont préférables en termes de neutralité allocative.