

Départ à la retraite et taxation optimale¹

Helmuth Cremer², Jean-Marie Lozachmeur³ et Pierre Pestieau⁴.

Avril 2004

¹Les auteurs tiennent à remercier un rapporteur anonyme pour ses remarques pertinentes.

²IDEI et GREMAQ, Université de Toulouse.

³GREMAQ, Université de Toulouse.

⁴CREPP, Université de Liège, CORE et DELTA.

Résumé

Pour expliquer les départs précoces à la retraite, les économistes utilisent le concept de taxation implicite sur tout prolongement de l'activité. De là à proposer que l'on élimine cette taxe, il n'y a qu'un pas à franchir. Cette contribution montre que si le système de retraite se veut redistributif et que le gouvernement n'observe pas certaines caractéristiques associées aux individus, une certaine taxe implicite sur l'âge de la retraite est inévitable. Notre analyse étudie aussi les moyens de limiter cette taxe implicite.

To

Classification *JEL* : E62, H23, H55

Mots clefs : retraite, invalidité, audit.

1 Introduction

Dans presque tous les pays européens, on observe des départs à la retraite “précoces”. Le taux d’activité des personnes âgées de 55 à 64 ans varie de 24% en Belgique à 88% en Irlande, la majorité des pays étant plus proches de la Belgique que de l’Irlande. En soi, une retraite avancée est une chance et non pas un problème à la condition que cela n’entraîne pas une charge insupportable. Or on sait qu’une partie de la facture sera payée par les générations futures comme l’indique la comptabilité générationnelle.¹

Des études comparatives récentes montrent clairement que les pays où l’âge de la retraite effectif est le plus bas sont aussi ceux où tout prolongement de l’activité est pénalisé lourdement. Une année de travail supplémentaire est sanctionnée par les prélèvements standards : cotisations sociales et impôts sur le revenu. En outre, dans certains pays à tout le moins, elle n’entraîne pas une augmentation mais une diminution des “droits à la retraite”, mesurés par la valeur actualisée des prestations nettes futures. Cette double pénalité est souvent désignée par le terme de “taxe implicite”.

La Figure 1 met en relation la taxe implicite et l’âge effectif de la retraite. Les données sont individuelles et concernent la Belgique. Les unités de comparaison sont les déciles de revenus. Les classes de revenus les plus élevées travaillent plus longtemps que les classes de revenus les plus basses et il font face à une moindre incitation à partir à la retraite de façon précoce. On peut observer que les individus à bas revenus sont incités à prendre une retraite anticipée non pas uniquement parce que la taxe implicite marginale est élevée (effet de substitution), mais aussi parce qu’ils bénéficient de la redistribution inhérente au système de retraite belge (effet revenu).

Pourrait-on faire l’économie de ces taxes implicites et de ce fait encourager les travailleurs âgés à cesser leur activité plus tard ? Oui, sans doute mais en partie seulement. En effet, ces taxes implicites ne sont pas nécessairement le reflet d’une mauvaise conception du système de retraite ; elles peuvent également résulter de l’objectif de redistribution que les pouvoirs publics lui assignent. C’est cet argument que nous soutenons dans cet article. A partir du moment où il y a redistribution et que les autorités publiques ne disposent que d’une information partielle sur les caractéristiques des individus, les distorsions sont inévitables. Les taux d’activité seront donc plus bas que ceux induits par une politique de premier rang malheureusement impraticable.

Nous présentons d’abord le modèle de base. Les individus sont caractérisés par une capacité physique ou psychologique à travailler plus longtemps

¹Voir le numero spécial d’*Economie et Prévision*, #154-3, 2002 consacré à la comptabilité générationnelle.

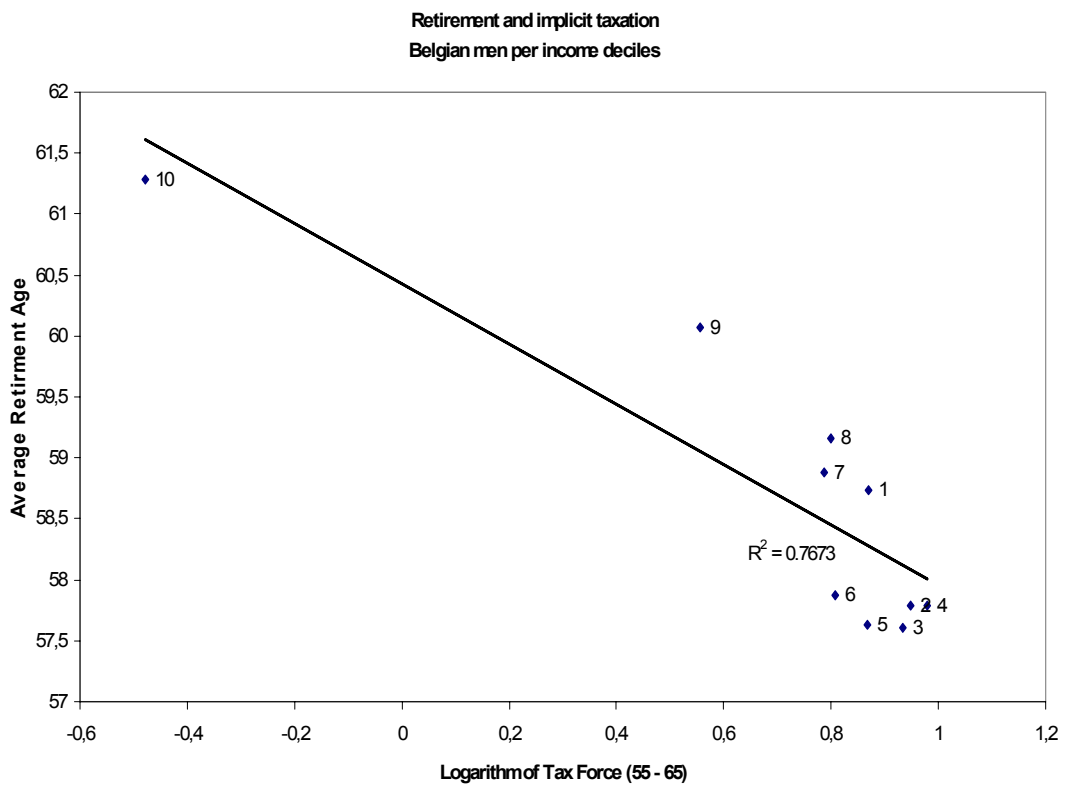


FIG. 1 – Taxation implicite et âge de la retraite en Belgique.

qui varie et que n'observe pas l'autorité publique. Chaque individu, étant donné cette caractéristique et le système de retraite, choisit son niveau de consommation et la durée de son activité. Nous déterminons le système de retraite qui maximise le bien-être utilitariste. Cette analyse peut donc s'interpréter comme de la redistribution ou de l'assurance sociale derrière le voile de l'ignorance. Dans le cas où les caractéristiques individuelles ne sont pas observables, il apparaîtra que le système de retraite optimal conduit à une taxation implicite de l'âge de la retraite. Nous montrons ensuite que l'utilisation de procédures d'audit ou de contrôle des aptitudes physiques peut permettre de réduire cette taxation implicite et améliorer la redistribution.

Enfin, nous envisageons deux extensions du modèle de base. La première consiste à prendre en compte le fait que l'état de santé des individus n'est pas purement exogène, mais est aussi influencé par des dépenses de santé. La subvention des dépenses de santé, ou la fourniture directe (en nature) de soins de santé par les pouvoirs publics constituent alors des instruments de redistribution supplémentaires. Alors que le système de retraite permet de redistribuer des personnes aptes vers les inaptes, l'intervention au niveau des soins de santé permet de redistribuer *ex ante*, ce qui réduit d'autant la nécessité d'effectuer des transferts *ex post*.

Dans une seconde extension, nous distinguons entre les vrais et les faux "invalides". Nous introduisons l'idée qu'une faible capacité à travailler longtemps puisse être due tant à un mauvais état de santé qu'à une forte préférence pour le loisir. En d'autres termes, tant les invalides que les "épicuriens" auront des préférences qui favorisent une retraite précoce et il peut s'avérer difficile de les distinguer. Cependant, il est possible que la société ne soit pas prête à traiter ces deux catégories de la même façon. En d'autres termes s'il peut paraître légitime d'aider les invalides qui quittent le marché du travail de façon précoce, des transferts vers les épicuriens peuvent ne pas être perçus de la même façon. Nous étudions l'impact de ce phénomène sur la conception du système de retraite (et d'invalidité). En particulier nous réexaminons l'utilisation de procédures des contrôles (portant sur l'état de santé) dans ce contexte.

Avant de poursuivre deux remarques de clarification s'imposent. Premièrement, notre approche s'inspire de la littérature sur la taxation optimale des revenus. Le système de taxation et de transferts envisagé regroupe la sécurité sociale et la fiscalité des travailleurs âgés et des retraités. Tout particulièrement, il inclut l'assurance invalidité qui permet aux travailleurs âgés de cesser leur activité avant l'âge où ils peuvent toucher leurs prestations de retraite. Deuxièmement, quand nous parlons d'âge de la retraite nous parlons de l'âge effectif pour chaque individu et non pas de l'âge "légal". Dans notre modèle comme dans la réalité, il existe une dispersion assez large d'âges de

départ à la retraite qui dépendent des paramètres du système de retraite (âge légal, taux de remplacement, durée de la carrière normale, etc.) et des caractéristiques propres aux individus.

2 Le modèle de base

Le modèle de base s'inspire de Cremer et al. (2004a). Pour analyser le choix de l'âge de la retraite, nous envisageons un simple modèle de cycle de vie. L'individu considéré traverse plusieurs phases de son cycle de vie et choisit l'âge de retraite z et la consommation annuelle c en tenant compte d'une seule contrainte budgétaire globale. Cette contrainte s'écrit :

$$\int_0^h c(t) dt = \int_0^z w(1 - \tau) dt + \int_z^h p(z) dt, \quad (1)$$

où h est la durée de la vie, w le salaire, $w\tau$ la cotisation sociale et $p(z)$ le montant instantané des retraites. Nous faisons abstraction de toute contrainte de liquidité. L'épargne (à tout instant) peut être positive ou négative et le taux d'intérêt est constant (petite économie ouverte) et normalisé à zéro. Le consommateur maximise son utilité de cycle de vie :

$$U = \int_0^h u(c(t)) dt - \int_0^z r(t) dt, \quad (2)$$

où $u(\cdot)$ est une fonction d'utilité croissante et concave, sous la contrainte (1). Le terme $r(t)$ reflète la désutilité instantanée du travail ; il croît avec t , ce qui veut dire que la pénibilité du travail augmente avec l'âge. On notera que le taux de préférence pour le temps est nul (égal au taux d'intérêt). Étant donné ces hypothèses, il est possible de simplifier la contrainte budgétaire et l'utilité sur le cycle de vie. Nous obtenons :

$$hc = wz - T(z) \quad (3)$$

et

$$U = h u(c) - R(z). \quad (4)$$

où $R(z) = \int_0^z r(t) dt$.

On a ainsi introduit une fonction de "taxation"

$$T(z) \equiv \int_0^z w\tau dt - \int_z^h p(z) dt = z\tau w - (h - z)p(z)$$

qui représente le système de taxation et de retraite sous une forme réduite. Dans T on trouve l'ensemble des taxes et cotisations dont sont soustraites

les prestations de retraite, d'invalidité et de chômage. La taxe implicite sur le prolongement de la vie active qu'ont estimée Gruber et Wise (1999) n'est autre que

$$T'(z) = \tau w + p(z) - (h - z)p'(z)$$

En d'autres termes, une année de travail supplémentaire peut entraîner un double coût : la cotisation sociale, τw , et la perte du droit instantané à la retraite $p(z)$ éventuellement compensée par une augmentation du montant des retraites si $p'(z) > 0$.

2.1 L'arbitrage entre revenus et loisirs sur le cycle de vie

Chaque agent choisit la valeur de z qui maximise :

$$U = hu \left(\frac{wz - T(z)}{h} \right) - R(z).$$

La condition du premier ordre s'écrit :

$$u'(c) (w - T'(z)) - R'(z) = 0. \quad (5)$$

Afin d'interpréter ces conditions, nous introduisons le concept de taux marginal de substitution

$$TMS_{cz} = \left. \frac{dc}{dz} \right|_U = \frac{R'(z)}{u'(c)},$$

entre c et z . Cela nous permet de réécrire la condition du premier ordre :

$$TMS_{cz} = w - T'(z). \quad (6)$$

La taxation marginale peut influencer le choix optimal de l'âge de la retraite : si $T'(z)$ est positif, alors le choix de z par rapport à c est altéré à la baisse.

2.2 L'optimum social

Nous considérons une société où les individus se différencient selon leur préférence pour la retraite R_j . Par souci de simplicité, nous supposons qu'il y a deux types de préférence $R_1 > R_2$. Les individus de type 1 sont donc les individus en mauvaise santé alors que les individus de type 2 sont en bonne santé. Soit f_i la fréquence du type i ($i = 1, 2$). Le taux de salaire w est le même pour tous et normalisé à 1. Cette économie peut être représentée par

un modèle à générations imbriquées avec population constante saisi à l'état stationnaire.²

2.2.1 L'optimum de premier rang

Considérons maintenant un planificateur social qui souhaite maximiser la somme des utilités d'une cohorte³ :

$$W = \sum_i f_i U_i.$$

Supposons dans un premier temps que le planificateur observe les caractéristiques individuelles R_i . Son problème se ramène à maximiser le Lagrangien :

$$\mathcal{L} = \sum_i f_i [hu(c_i) - R_i(z_i)] - \gamma \sum_i f_i (hc_i - wz_i),$$

où γ est le multiplicateur associé à la contrainte de ressources de l'économie. Les conditions du premier ordre impliquent :

$$\begin{aligned} u'(c_1) &= u'(c_2), \\ R'_1(z_1) &= R'_2(z_2). \end{aligned}$$

Il y a de la redistribution dans la consommation : tous les c_i sont égalisés. De plus, il n'y a pas de distorsion induite du choix de l'âge de la retraite : $T'_1(z_1) = T'_2(z_2) = 0$. Les désutilités marginales du travail étant égalisées entre les deux individus, l'individu 1 bénéficiera d'un âge de la retraite plus faible étant donné son mauvais état de santé.

Peut-on décentraliser un tel optimum ? Oui, si l'on dispose d'un système de transferts forfaitaires individualisés T_i . Cependant, ce système de transferts forfaitaires n'est réalisable que si les caractéristiques individuelles sont observables. En général, on ne pense pas que les R_i soient parfaitement observables. Dans ce cas, il est nécessaire d'avoir recours à des schémas de second rang. En effet, si les allocations optimales de premier rang décrites ci dessus étaient proposées, l'individu 2 aurait intérêt à se faire passer pour un individu 1 et ainsi consommer un même montant en travaillant moins. Que peut faire

² Il y a identité entre une coupe transversale avec les deux types d'agents âgés de 0 à h et une cohorte de quatre types d'agents saisis sur la totalité de leur cycle de vie. Nous étudierons le système de retraite qui est optimal du point de vue longitudinal ; il l'est aussi du point de vue transversal

³Dans Cremer et.al (2004a), nous envisageons une fonction de bien être concave (section 5). Les résultats exposés ici à savoir la forme de la taxation implicite sont similaires dans ce cas. Seul le degré de redistribution change.

un gouvernement qui ne dispose donc que de transferts T dépendant de la seule variable observable z_i et non plus du type de l'individu dénoté par R_i ? C'est à cette question que s'adresse le calcul de l'optimum de second rang.

2.2.2 L'optimum de second rang

Formellement, le problème de second rang avec deux types est identique au problème de premier rang auquel on ajoute une contrainte d'auto-sélection. Celle-ci implique que l'individu "le mieux nanti" ne se fasse pas passer pour l'autre individu. Si l'individu en bonne santé est 2, cette contrainte s'écrit :

$$hu(c_2) - R_2(z_2) - hu(c_1) + R_2(z_1) \geq 0.$$

Des conditions du premier ordre, on tire les résultats suivants :⁴

$$\begin{aligned} c_1 &< c_2 \\ T'_1(z_1) &> 0 \\ T'_2(z_2) &= 0 \end{aligned}$$

La contrainte d'incitation ne permet pas de redistribuer autant que l'on voudrait. De plus, un départ à la retraite précoce est désirable pour l'individu en mauvaise santé : une taxation à la marge du prolongement de l'activité est optimale. Intuitivement, cette propriété vient du fait que les agents de type 1 ont des courbes d'indifférence plus verticales que les individus de type 2 en tout points du plan (c, z) . En d'autres termes, pour accepter de travailler plus longtemps, les agents de type 1 doivent être plus généreusement compensés que les agents de type 2 voulant les imiter. Inversement, la réduction de revenus que les individus de type 1 sont prêt à accepter pour travailler une année de moins est plus grande que celle qu'accepteraient les individus de type 2. Cela implique que, en partant de l'arbitrage de premier rang, une variation $dz_1 < 0$ avec $dc_1 = -(MRS_{cz}^1)dz_1$ n'a pas d'effets de premier ordre sur l'utilité de l'agent de type 1 et baisse l'utilité de l'agent de type 2 imitant l'agent de type 1. En conséquence, la distorsion à la baisse de z_1 est une manière de relâcher la contrainte d'incitation.

2.3 Comment réduire cette taxe implicite : l'audit sur l'état de santé

Comme on l'a vu précédemment, c'est l'existence même de l'information asymétrique sur les caractéristiques des individus qui implique une taxation

⁴Voir Cremer et al. (2004a).

à la marge sur l'âge de la retraite. Cette section présente les instruments permettant de rendre l'absence d'information moins contraignante.

2.3.1 L'audit parfait (Cremer et al., 2004b))

Jusqu'ici, nous avons supposé que l'état de santé n'était pas publiquement observable. Cette hypothèse paraît bien sûr extrême dans la mesure où souvent un simple contrôle médical permet de diagnostiquer l'état de santé d'un individu. Ce contrôle est d'ailleurs une composante des programmes d'assurance invalidité. Formellement, nous introduisons dans le modèle une probabilité π d'audit sur les individus se réclamant du type 1. Cet audit est supposé parfait dans la mesure où il permet de déterminer sans erreur si l'individu concerné est en bonne ou mauvaise santé. Cet audit a un coût marginal k et le coût total associé à une probabilité π d'audit est donc donné par $f_1 k \pi$. Nous supposons enfin, que les individus 2 étant pris à tricher se voient attribuer une utilité minimale \bar{u} . Comme en l'absence de contrôle, la politique optimale doit être conçue de telle façon qu'aucun individu n'aura intérêt à mentir sur son état de santé. La possibilité d'audit affecte cependant les contraintes de ressource et d'auto-sélection qui sont maintenant données par :

$$\sum_i f_i (wz_i - hc_i - f_1 k \pi) = 0,$$

$$u(c_2) - R_2(z_2) - (1 - \pi)(u(c_1) - R_2(z_1)) - \pi \bar{u} = 0.$$

Ces deux équations montrent qu'augmenter la probabilité d'audit a deux effets : le premier est d'engendrer une hausse des dépenses et le deuxième de relâcher la contrainte d'incitation en baissant le niveau d'utilité de l'individu 2 imitant l'individu 1. A l'optimum, un arbitrage s'opérera donc entre ce coût et ce gain marginaux.⁵

La solution est donc simple. Pour un coût d'audit nul, l'Etat aura intérêt à utiliser l'audit avec une forte probabilité π de façon à rendre la contrainte d'incitation non liante. Dans ce cas, la règle d'allocation du premier rang est réalisable et aucune taxation à la marge de la retraite n'est nécessaire. A mesure que le coût augmente, il est optimal d'utiliser un audit sur une fraction des individus se réclamant invalides. La contrainte d'incitation est "moins liante" par rapport à une situation sans audit et donc la taxation marginale de l'âge de la retraite est toujours présente mais plus faible. Quand

⁵Si $\pi = 1$, la contrainte d'auto-sélection n'est plus pertinente. Avec une probabilité de contrôle (et de pénalité) de 100%, tricher devient une stratégie dominée. On revient donc à un contexte d'information complète, mais au prix d'un coût d'audit qui peut être considérable.

ce coût devient très élevé, l'audit n'est plus utilisé et l'optimum de second rang est le même que celui présenté dans la section précédente.

Notons enfin qu'une "prime" devrait être accordée aux individus 1 contrôlés lors de l'audit. En effet, cette prime ne modifie pas la contrainte d'auto-sélection. La partie de la population dont l'état de santé révélé par l'audit est faible peuvent ainsi bénéficier d'une meilleure prestation sans que les individus 2 puissent y prétendre.

On voit ainsi que l'introduction de l'audit permet de réduire la taxation marginale de l'âge de la retraite des individus 1 et ce, d'autant plus que le coût de ce contrôle est faible.

2.3.2 L'audit imparfait

L'audit sur la santé est loin d'être parfait pour certains types de maladie. Pour cette raison, il est raisonnable de penser que l'audit ne révèle pas toujours correctement l'état de santé des individus. Diamond et Sheshinski (1995) ont introduit ce type d'erreur dans un modèle similaire à celui que nous venons de présenter à ceci près que la décision de travailler est binaire⁶. Formellement, cela revient à dire que l'audit révèle correctement l'état de santé d'un individu avec une certaine probabilité. Les auteurs supposent de plus que l'audit est gratuit et éliminent donc l'arbitrage entre coût financier de l'audit et gain d'efficacité de celui-ci. Le contrôle est ainsi opéré sur tout les agents dans ce modèle.

Dans ce cadre, les auteurs montrent qu'à l'optimum, la population peut être divisée en trois groupes : (i) les travailleurs valides en activité, (ii) les travailleurs valides et les travailleurs invalides en inactivité percevant des allocations d'invalidité et enfin (iii) les travailleurs invalides percevant des prestations plus faibles que les allocations d'invalidité. Il est donc optimal de créer un régime d'invalidité dans lequel on retrouve une certaine proportion de travailleurs valides (égale à la probabilité d'erreur de l'audit). Cette propriété implique donc une certaine taxation implicite du travail. Les travailleurs invalides jugés valides suite à une erreur d'audit percevront des allocations plus faibles de sorte qu'aucun travailleur valide ne souhaite recevoir ces allocations. La seule différence par rapport à l'audit parfait est donc la présence de personnes en bonne santé percevant des allocations d'invalidité.

Pour résumer, les résultats de ces deux modèles d'audit impliquent donc la coexistence de deux régimes séparés :

⁶Les travailleurs décident donc de travailler une unité de temps (par exemple un an) ou bien cesser leur activité. On aura donc un seuil de désutilité du travail déterminé par le système d'assurance au delà duquel les individus décideront de ne plus travailler.

- Un régime de pré-retraite spécifique aux travailleurs âgés comprenant les individus en mauvaise santé non contrôlés (audit parfait) ou non reconnus comme tels par l’audit (audit imparfait). Les prestations de ce régime seraient telles que les travailleurs valides soient indifférents entre prolonger leur activité et bénéficier de ce système.
- Un régime d’assurance invalidité comprenant les travailleurs invalides ayant subi un contrôle (audit parfait) ou des travailleurs valides et invalides reconnus comme invalides par le contrôle (audit imparfait). Les prestations de ce second régime seraient donc supérieures à celles du premier.

3 Extensions

Nous envisageons deux extensions du modèle de base. La première consiste à prendre en compte le fait que l’état de santé des individus n’est pas purement exogène, mais est aussi influencé par des dépenses de santé. La subvention des dépenses de santé, ou la fourniture direct (en nature) de soins de santé par les pouvoirs publics constituent alors des instruments de redistribution supplémentaire.

Dans une seconde extension, nous distinguons les vrais et les faux “invalides”. Nous introduisons l’idée qu’une faible capacité à travailler longtemps puisse être due soit à un mauvais état de santé, soit à une forte préférence pour le loisir. En d’autres termes, les invalides et les “épicuriens” auront des préférences qui favorisent une retraite précoce et il peut s’avérer difficile de les distinguer⁷. Cependant, nous supposons que ces deux types d’individus ne sont pas traités de la même façon dans la fonction de bien-être social. Nous étudions l’impact de ce phénomène sur le système de retraite (et d’invalidité). En particulier nous réexaminons l’utilisation de procédures de contrôle (portant sur l’état de santé) dans ce contexte.

3.1 Dépenses de santé et transferts en nature (Cremer et al. (2004c))

Il est généralement reconnu que la fourniture d’un bien en nature peut être désirable dans un contexte de second rang. Un candidat naturel pour ce type de transferts est la santé.

Pour introduire ces dépenses dans notre modèle, réécrivons la désutilité du travail d’un individu $R_i(z_i, s_i)$ où s_i représente le niveau de soins de santé

⁷Dans ce cas le mécanisme d’autosélection est inopérant.

consommé par l'individu i . Nous faisons en outre l'hypothèse que $R_1(z, s) > R_2(z, s)$ et $dR_1/ds < dR_2/ds < 0$. En d'autres termes, nous supposons que la consommation d'une unité supplémentaire de soins de santé a un effet plus fort pour l'individu en mauvaise santé que pour l'individu en bonne santé. Dans un optimum de premier rang, les consommations sont égalisées et normalement, le travailleur de type 2 consomme moins de santé et travaille plus longtemps que l'individu de type 1. Ce qui nous intéresse est ce qui se passe lorsque l'information n'est pas parfaite.

Deux configurations peuvent émerger selon l'information dont disposent les pouvoirs publics. S'ils observent les dépenses de santé des individus, alors Cremer et al. (2004c) montrent qu'il est optimal de fournir un niveau de soin supérieure à celle désirée par les individus de type 1. Ceci peut se faire via une fourniture directe ou via une subvention personnalisée du prix de la santé. En effet, les courbes d'indifférence des individus de type 1 sont plus horizontales que celles des individus de type 2 dans le plan (c, s) puisque la santé réduit marginalement plus l'effort pour les individus en mauvaise santé. Un raisonnement analogue à celui présenté dans la section 2.2.2 démontre donc qu'il faut altérer à la hausse le choix de s_1 afin de relâcher la contrainte d'incitation.

Toutefois, lorsque la consommation de soins de santé des individus n'est pas observable, il n'est plus possible d'offrir une allocation personnalisée. La politique optimale dans ce cas consiste à subventionner uniformément le prix de la santé et en plus, de fournir (gratuitement) à tous les individus une certaine quantité de soins de santé. Dans la mesure où ce bien réduit plus à la marge la désutilité du travail de l'individu 1, la quantité désirée s_1 du bien pour ce dernier est supérieure à celle désirée \bar{s}_2 par un individu de type 2 imitant l'agent de type 1. En d'autres termes, à durée de carrière égale, les individus en mauvaise santé demandent plus de soins de santé que les individus en bonne santé. De plus, à carrière plus longue est associée une demande du bien supérieure. En conséquence, l'individu de type 2, s'il ne triche pas, désire une quantité s_2 supérieure à \bar{s}_2 . Un transfert en nature uniforme s tel que $\bar{s}_2 < s < \min(s_1, s_2)$ n'aura donc pas d'effet de premier ordre sur le bien être des deux types d'individus mais contraindra les choix de l'individu 2 imitant l'individu de type 1. Ainsi donc, en fournissant de façon adéquate une quantité uniforme de soins de santé aux deux types d'individus, l'Etat fait en sorte qu'il soit plus difficile pour le type 2 de se faire passer pour le type 1 et la redistribution en devient plus efficace.

3.2 L’audit pour séparer les vrais des faux invalides (Cremer et al.(2004d))

L’assurance invalidité est une voie fréquemment utilisée pour sortir du marché du travail. Le tableau 1 présente le nombre de bénéficiaires de l’allocation d’invalidité en Belgique (B), Allemagne (G), Suisse (S) et aux Pays Bas (NL) pour l’année 1998. Dans ce dernier pays, il est frappant de constater que 75% des assurés sociaux soient bénéficiaires de l’assurance invalidité chez les 55-64 ans. Le score moins élevé dans d’autres pays atteste au moins d’un certain “abus” de l’assurance invalidité dans ce pays. Ce résultat peut provenir d’une volonté⁸ de ne pas distinguer les individus “vraiment” invalides de ceux qui ont une préférence pour le loisir relativement forte.

Years	Recipients ¹				New recipients ²				Completions ³			
	B	G	S	NL	B	G	S	NL	B	G	S	NL
<25	3	1	9	4	2	0	2	4	325	118	-	392
25-34	14	7	16	29	4	1	2	12	167	79	-	201
35-44	40	22	42	73	8	4	4	15	106	71	-	103
45-54	96	61	94	184	13	11	8	22	64	79	-	56
55-64	202	317	240	754	12	29	16	31	153	44	-	135
Total	54	62	76	112	7	8	6	15	115	53	-	110

Legends: 1 per 1.000 insured
2 per 1.000 insured
3 per 1.000 recipients

Tableau 1 : Bénéficiaires de l’allocation d’invalidité selon l’âge (source : Beljaars et Prins (2000)).

Pour étudier cette hypothèse nous nous inspirons de Cremer et al. (2004d) et nous considérons une économie composée de trois types d’individus. Comme dans le modèle de base, il y a des individus en bonne santé et des individus en mauvaise santé. Les premiers ont une désutilité pour le travail égale à R_H et les autres R_D . A ces types d’individus s’ajoute une troisième catégorie L , composé d’individus qui ont également une forte préférence pour le loisir ($R_L = R_D$) mais pour des raisons autres que la santé. Nous continuons de supposer que tous les individus ont la même productivité.

Le gouvernement n’accorde pas le même poids aux individus invalides par rapport aux épicuriens de la catégorie L . Plus précisément, l’objectif du gouvernement s’écrit :

$$\max \sum_i f_i u(c_i) - f_H R_H(z_H) - f_L R_H(z_L) - f_D R_L(z_D).$$

⁸Cette volonté peut s’expliquer par des motivations macroéconomiques. Elle peut aussi venir d’une réticence à procéder à des contrôles vexatoires.

Le gouvernement ne prend donc pas en compte la désutilité du travail de la catégorie L telle qu'elle est perçue par ces individus. Il assigne à ces derniers une désutilité *sociale* du travail égale à celle des individus en bonne santé.

L'optimum de premier rang consiste à offrir le même niveau de consommation à tous les individus. L'âge de retraite est le même pour les individus H et L mais supérieur à celui des individus de type D : $z_H = z_L > z_D$. Comme dans le modèle de base, l'âge de la retraite n'est pas taxé à la marge pour les individus H et D . En revanche, il y a une taxe marginale négative (une subvention) sur z_L . Ce résultat provient du fait qu'il existe une différence entre la désutilité sociale du travail R_H et la désutilité "privée" du travail R_L . Le gouvernement doit ainsi inciter les travailleurs L à travailler plus que la normale, ce qui est possible grâce à une subvention "pigouvienne".

Si le gouvernement n'observe pas l'état de santé des individus, l'optimum (de second rang) est simple : on ne peut séparer les individus L et D .⁹ Comme dans la section 2.2.2, une taxe marginale positive sur z_D (et donc z_L) est désirable afin de rendre moins attractif pour l'individu H les allocations offertes à D et L . Ce cas illustre très bien la réalité où l'assurance invalidité est utilisée par un trop grand nombre de travailleurs pour quitter le marché du travail de manière précoce.¹⁰ Il est clair que la non observabilité de l'état de santé est coûteuse en terme de bien être. Cette situation peut être améliorée grâce aux contrôles de l'état de santé par un audit.

Nous adoptons la même technologie d'audit que celle utilisée à la section 2.3.1. L'audit révèle parfaitement l'état de santé des individus et il permet donc de distinguer les individus D de ceux des autres catégories (et en particulier des individus L). Cremer et al. (2004d) montrent que l'optimum de second rang peut conduire à différents régimes selon le coût de l'audit. Nous présentons dans le tableau 2 un exemple numérique afin d'illustrer les résultats. Ce tableau présente pour des paramètres donnés l'optimum de premier rang, puis ceux de second rang sans et avec audit. L'audit 1 est plus coûteux que l'audit 2 et conduit à une probabilité de contrôle plus faible. Il apparaît clairement que l'audit améliore le bien être des invalides au détriment de celui des individus de type L . Avec l'audit 1, une taxe marginale positive est désirable pour les individus L et D car les contraintes d'incitation HL et HD sont liantes. On peut montrer aussi que pour ce régime, les individus de type L seront toujours amenés à travailler moins que les invalides. Ces derniers reçoivent néanmoins une plus grande prestation d'invalidité. Quand l'audit devient moins coûteux (audit 2), HD n'est plus liant. Une taxe marginale

⁹Sous certaines hypothèses, la séparation peut être possible et souhaitable même en l'absence d'audit. Cependant, nous nous concentrons sur le cas le plus simple ici ; voir Cremer et al. (2004d).

¹⁰L'absence d'audit représente sans doute ce qui se passe aux Pays Bas.

sur z_D devient donc inutile. Notons enfin que dans ce régime, les invalides travaillent moins que les individus de type L .

Types	Premier rang			Deuxième rang sans audit		
	D	L	H	D	L	H
c	86.74	86.74	86.74	57.72	57.72	94.04
z	0.28	1.15	1.15	0.39	0.39	1.06
U	4.29	1.80	3.79	3.74	3.74	3.97
$T'(z)$	0	-3	0	0.09	0.09	0

Types	Audit 1			Audit 2		
	$\pi = 0.04$, Régime 1			$\pi = 0.08$, Régime 2		
	D	L	H	D	L	H
c	65.35	54.05	91.55	68.06	53.70	90.59
z	0.38	0.37	1.09	0.36	0.41	1.10
U	3.88	3.71	3.92	3.95	3.63	3.89
$T'(z)$	0.003	0.19	0	0	0.11	0

Tableau 2 : Exemple numérique avec et sans audit

4 Conclusion

On sait que dans certains pays les travailleurs partent à la retraite beaucoup plus tôt que dans d'autres ; la raison majeure étant que dans ces pays il existe une série de possibilités de cesser toute activité bien avant l'âge légal de départ à la retraite. Pour prendre un exemple, aux Pays Bas, très longtemps et encore aujourd'hui, l'assurance invalidité était considérée comme une porte de sortie du marché du travail privilégiée par les travailleurs, leurs employeurs et l'Etat. La possibilité de recourir à ces programmes, que ce soit l'assurance invalidité, l'assurance chômage, les préretraites ou dans certaines branches les régimes spéciaux, encourage les travailleurs à partir à la retraite prématurément. Ils rendent coûteuse pour certains toute prolongation d'activité. Ce coût du prolongement d'activité porte le nom de taxe implicite estimée par Gruber et Wise (1999) et par Blondal et Scarpetta (1998a,b) pour les pays de l'OCDE. La France, les Pays Bas et la Belgique dont l'âge effectif de départ à la retraite tourne autour de 58 ans ont ainsi une taxe implicite plus élevée que la Suède ou le Japon où les travailleurs partent à la retraite beaucoup plus tard.

Le vieillissement démographique, la longévité croissante, la crise des finances publiques et la prise de conscience que décourager l'emploi des travailleurs âgés n'a pas d'effet sur l'emploi des jeunes conduisent les organisations internationales et de nombreux économistes à proposer de réduire, voire

de supprimer, toute taxation imposée au prolongement d'activité avant l'âge raisonnable de départ à la retraite. L'objet de cet article était de montrer que même si la suppression de toute taxation implicite était souhaitable dans un monde d'information et de contrôle parfaits, ce n'est pas le cas dans le monde réel où l'Etat ne dispose que d'information et d'instruments imparfaits. Nous montrons qu'une certaine taxation implicite est inévitable et donc que les départs à la retraite ne peuvent être aussi tardifs qu'il ne serait souhaitable. Cependant, en faisant le meilleur usage possible des instruments disponibles, il y a un moyen de réduire cette taxation implicite à son minimum.

Le premier instrument disponible est un système de transferts (incluant prestations, taxes et cotisations) qui amènent des travailleurs de type différents à se segmenter. Les caractéristiques classiques dans ce type de problème sont la productivité et l'état de santé (la capacité à travailler longtemps). Nous avons mis l'accent sur cette seconde caractéristique pour des raisons de clarté. Mais une extension à deux caractéristiques est possible mais implique des résultats plus complexes. En utilisant donc une contrainte d'autosélection, le système de transferts permet de traiter différemment les travailleurs en bonne et mauvaise santé.

L'autosélection limite les capacités redistributives d'un Etat utilitariste ; nous faisons en effet l'hypothèse que l'objectif de l'Etat est utilitariste. De nouveau une extension à un objectif rawlsien, par exemple, est possible. Si l'état de santé des individus peut être connu moyennant des tests qui ne sont pas trop coûteux, nous montrons que la redistribution et le bien-être social peuvent être améliorés.

Il existe des situations où la technique d'autosélection est inopérante. Ce sont des situations où des agents différents ont des préférences qui paraissent identiques ; en d'autres termes, des agents différents ne peuvent être distingués. Nous prenons l'exemple où une retraite anticipée peut être voulue pour deux raisons : une raison de santé et une raison de préférence pour le loisir. Dans ce cas, seule une technique d'audit peut permettre de distinguer les deux types d'agents. On comprendra que l'Etat ne traite pas ces deux types de manière identique ; il souhaite aider davantage les travailleurs souffrant d'invalidité que les travailleurs aimant le loisir. Pour représenter cette préférence, nous adoptons un critère non-welfariste ou plutôt paternaliste, qui implique que l'Etat veut faire travailler les épicuriens plus longtemps que les invalides.

Enfin, si des dépenses de santé peuvent améliorer la santé de tous mais tout particulièrement celle des invalides, nous montrons que, sous certaines conditions, il faut recourir à cet instrument même en situation d'information asymétrique. Un investissement public dans ce domaine permet une meilleure redistribution parce qu'elle rend moins facile pour le travailleur bien portant

de se faire passer pour un travailleur invalide.

Références

- [1] Beljaars P. et R. Prins, (2000), Disability programme reforms and labour market participation in the Netherlands (1990-2000) : Principles, measures and outcomes in a decade of combating high disability rates, mimeo.
- [2] Blondal, S. and S. Scarpetta, (1998a), Falling participation rates among older workers in the OECD countries, OECD, Paris.
- [3] Blondal, S. and S. Scarpetta, (1998b), The retirement decision in OECD countries, OECD-EDWP 202.
- [4] Cremer, H., J.-M. Lozachmeur et P. Pestieau, (2004a), Social security, retirement age and optimal income taxation, à paraître dans *Journal of Public Economics*.
- [5] Cremer, H., J.-M. Lozachmeur et P. Pestieau, (2004b), Optimal Retirement and disability benefits with audit, *Finanzarchiv* 60, 278-295.
- [6] Cremer, H., J.-M. Lozachmeur et P. Pestieau, (2004c), Optimal Retirement and disability benefits with health expenditures, mimeo.
- [7] Cremer, H., J.-M. Lozachmeur et P. Pestieau, (2004d), Disability testing and retirement, mimeo.
- [8] Dellis, A., R. Desmet, A. Jousten et S. Perelman, Micro-modelling of retirement in Belgium, in Gruber, J. and Wise D., (éd.), *Micro-modelling of Retirement Incentives around the World*, University of Chicago Press et NBER, Chicago, à paraître.
- [9] Diamond, P. and E. Sheshinski, (1995), Economic aspects of optimal disability benefits, *Journal of Public Economics*, 57, 1-23.
- [10] Fabel, D., (1994), *The Economics of Pension and Variable Retirement Schemes*, J. Wiley, New York.
- [11] Gruber, G. and D. Wise, (1999), *Social Security and Retirement around the World*, The Chicago University Press, Chicago.