

## Chapitre 2

# Coûts fixes et offre individuelle de travail au noir

« Pour certaines personnes, la fraude implique des coûts. Ceux-ci se composent d'une partie fixe qui peut s'avérer particulièrement conséquente. De ce fait, les variations marginales des gains pécuniaires pourraient n'avoir qu'un impact négligeable sur l'incitation à la fraude. »

-Jonathan C. Baldry (1987, p.377)-

### 2.1 Introduction

La théorie standard du consommateur nous enseigne que toute situation qui procure à l'individu un revenu supplémentaire est préférée. En ce sens, l'incitation au travail au noir devrait être d'autant plus importante que les revenus issus d'une activité dissimulée ne sont pas soumis à l'imposition. Malgré l'intérêt que peut susciter une telle activité, de nombreux individus choisissent de ne pas participer à l'économie souterraine. Ainsi, les travailleurs au noir ne représentent approximativement que 5% de notre échantillon. Ils soulignent, par ailleurs, leur préférence pour un emploi déclaré (Fortin *et al.*, 1996). Un élément explicatif de ce rejet apparemment irrationnel de revenu pourrait être l'existence d'un coût de participation à l'économie souterraine, *i.e.* d'une désutilité issue de l'entrée sur le marché noir.

Ce coût fixe peut être de nature différente. En premier lieu, il peut s'agir de coûts monétaires, de même type que ceux prévalant sur le marché officiel. Ils sont liés aux dépenses occasionnées par l'activité elle-même (frais de transport ou perte des revenus

de l'activité déclarée, par exemple) ou par les dépenses annexes telles que les frais de garderie. Des coûts d'opportunité sont également susceptibles d'apparaître. Cogan (1980) a montré que ces coûts d'opportunité peuvent fréquemment être transformés en coûts monétaires<sup>1</sup>. En second lieu, il peut exister des coûts de transaction induits par la recherche d'emplois non déclarés. L'accès aux opportunités est *a priori* susceptible de différer selon les individus et selon leur statut sur le marché officiel du travail<sup>2</sup>. Enfin, le coût fixe pourrait s'apparenter à un coût psychologique qui dissuaderait l'individu d'exercer une activité souterraine. En effet, le travail au noir pourrait être connoté négativement, à la fois, pour les participants et les non participants. Il caractériserait ainsi le stigmate social associé au comportement de fraude. Selon le degré de tolérance du travail au noir et son ampleur dans l'économie, un individu identifié en tant que fraudeur pourrait souffrir d'une certaine forme de disgrâce auprès des autres membres de la société. Tout comme les autorités fiscales peuvent imposer des pénalités, la société peut exposer le travailleur au noir à l'opprobre social. Ce risque de rejet de l'individu, qui est habituellement évalué subjectivement (Cowell, 1990), pourrait le dissuader d'entrer sur le marché noir. Dans ce cas, le coût fixe résulterait de l'interaction complexe entre l'évaluation par l'individu des conséquences de sa propre action et la façon dont il pense que son comportement est perçu par les autres, au regard de ce qu'ils font eux-mêmes.

Les coûts de participation à l'économie souterraine peuvent différer d'un individu à l'autre, mais ils sont supportés de façon identique quelle que soit l'intensité du travail. Ils affectent la décision d'entrer sur le marché, mais peuvent rester sans effet sur la détermination du volume horaire de l'activité. Par conséquent, les déterminants de ces deux décisions peuvent être sensiblement différents. L'offre de travail au noir est donc caractérisée par deux décisions distinctes qu'il convient d'examiner.

L'importance de la distinction entre participation et durée du travail a été étayée, sur le marché officiel, par de nombreuses contributions (*e.g.* Blundell 1992). Ces développements théoriques et empiriques ont démontré la présence de coûts fixes sur le marché du travail, en particulier pour les femmes. En revanche, la présence de tels coûts n'a pas, à notre connaissance, été explicitement envisagée dans la littérature théorique du travail au noir. Au niveau empirique, le manque d'applications économétriques s'explique par la difficulté de collecte d'information sur les activités dissimulées et par le caractère non observable des heures de réserve et des coûts de participation. Mais,

<sup>1</sup> Ainsi, un individu peut choisir de prendre un taxi plutôt que d'attendre le bus pour se rendre à son travail.

<sup>2</sup> Un travailleur indépendant a certainement plus de facilités à exercer des heures supplémentaires au noir ou à réaliser des tâches non déclarées en rapport avec son emploi officiel qu'un salarié, par exemple.

nous pouvons raisonnablement supposer l'existence *a priori* de tels coûts sur le marché noir et présumer d'un impact non négligeable sur le choix de l'activité souterraine.

L'enjeu de ce chapitre se veut double. Au niveau théorique d'une part, il consiste à introduire des coûts fixes dans un modèle d'offre de travail au noir et à distinguer la décision de travailler sur ce marché de celle du niveau de l'activité. Une telle démarche nous permet d'appréhender les diverses manifestations possibles du coût de participation à l'économie souterraine. Au niveau empirique d'autre part, notre procédure d'estimation doit permettre de tester la présence éventuelle de coûts à l'entrée sur le marché noir. L'analyse des données nous a révélé l'existence d'un nombre non négligeable d'individus ne travaillant pas sur ce marché. La variable dépendante relative au volume horaire de l'activité est censurée à gauche, de sorte que la fonction d'offre de travail présente une discontinuité. Il importe, dans ces conditions, de distinguer la décision d'exercer une activité non déclarée et le choix du volume horaire de cette activité. Par ailleurs, nous ne pouvons négliger les interactions possibles existantes entre le marché officiel et le marché non officiel. Autrement dit, nous sommes confrontés à une endogénéité potentielle des heures de travail déclarées. La méthode mise en œuvre, dans ce chapitre, repose sur l'approche conditionnelle initialement proposée par Polak (1969, 1971) et plus récemment utilisée par Browning et Méghir (1991), Manser (1993) et Fortin *et al.* (2000). Elle présente l'avantage de prendre en considération non seulement la procédure de sélection qui sous-tend les choix individuels de travailler au noir, mais également l'endogénéité de variables explicatives déterminantes. Celles-ci sont au nombre de trois, *i.e.* le taux de salaire horaire de l'emploi non déclaré, les heures travaillées sur le marché officiel et les revenus disponibles déclarés. Nous procédons aux estimations économétriques de l'offre individuelle de travail au noir selon une procédure en trois étapes. Nos résultats confirment l'existence de coûts importants à l'entrée sur le marché noir. Ils révèlent, en outre, la très grande sensibilité des individus au jugement de l'entourage et de l'opinion publique sur l'exercice d'une activité dissimulée aux autorités fiscales. L'exercice d'une activité souterraine est perçue négativement, de sorte que beaucoup d'individus sont réticents à travailler au noir. En ce sens, les coûts fixes estimés caractérisent l'opprobre jeté sur la dissimulation de la fraude. Nos résultats apportent enfin un éclairage sur l'impact du système fiscal et de répression de la fraude sur les comportements individuels.

La structure du chapitre est la suivante. Une revue de la littérature relative aux coûts de participation est proposée à la *section 2.2*. Le modèle théorique d'offre de travail au noir est exposé en *section 2.3*. Les spécifications retenues dans la procédure d'estimation économétrique sont développées à la *section 2.4*. Enfin, la *section 2.5* commente les principaux résultats.

## 2.2 Littérature relative aux coûts de participation

De nombreux travaux empiriques ont tenté d'expliquer les variations dans la participation et l'intensité au travail<sup>3</sup>. Celles-ci se sont attachées à expliquer l'effet des réformes fiscales et des programmes sociaux sur l'offre de travail des femmes. Les premiers modèles ne tenaient pas explicitement compte du système fiscal<sup>4</sup>. Ils n'étaient donc pas en mesure de fournir des prédictions sur l'effet désincitatif de la fiscalité et des transferts sur les comportements d'offre de travail. Afin de pallier cette limite, divers auteurs ont focalisé leur attention sur l'introduction d'un système fiscal plus réaliste. Initialement proposée par Hausman (1980), la progressivité de l'impôt est l'une des pistes de recherche qui ont été les plus exploitées. Elle a contribué au développement de modèles théoriques et économétriques sophistiqués caractérisant les distributions de variables discrètes et continues pour décrire la participation au marché du travail. Considérés à la pointe de la recherche en ce domaine, ces modèles offrent un mécanisme naturel pour appréhender les caractéristiques institutionnelles telles que le système fiscal et de redistribution.

En revanche, malgré leur influence indéniable sur les comportements individuels d'offre de travail, l'étude des coûts fixes est restée très limitée. La plupart des travaux néglige encore aujourd'hui l'éventualité de leur existence. Pourtant, les contributions visant à tester la présence de tels coûts ont toutes abouti à une réponse positive, bien qu'elles ne soient généralement pas parvenues à en quantifier l'ampleur. L'analyse des coûts de participation au marché du travail constitue dès lors un champ de recherche à explorer. Les modèles, incluant davantage de structure fiscale, offrent un cadre d'analyse particulièrement adapté. Néanmoins, l'introduction de coûts fixes complique singulièrement l'estimation de l'offre de travail au noir.

Cette section dresse un état des lieux des travaux théoriques et des applications économétriques sur l'offre de travail discontinue. Les modèles présentés sont fondés sur l'approche standard et permettent de traiter des variables manquantes ou censurées. Les principes sous-jacents sont très documentés dans la littérature économétrique. Dans les développements qui suivent, nous présentons les modifications du modèle standard permettant de prendre en compte la non linéarité de la contrainte budgétaire induite par le système fiscal, l'existence de coûts fixes et l'absence de salaire observé pour les non participants au marché du travail.

---

<sup>3</sup>Se reporter à Blundell et MaCurdy (1999) pour une présentation des différentes approches de modélisation empirique de l'offre de travail.

<sup>4</sup>*e.g.* Cragg (1971), Cogan (1980) ou encore Hanoch (1980).

Essentiellement deux champs de la littérature ont tenté d'appréhender l'existence de stigmatisme ou de coûts de participation. La première démarche tend à négliger la structure de la fiscalité. Dans ce cas, la différence entre un stigmatisme et un coût fixe n'est qu'une question de présentation. Le coût fixe modifie la contrainte budgétaire, générant un coude, tandis que le stigmatisme affecte la fonction d'utilité, créant une discontinuité en zéro. Mais, ces quelques subtilités graphiques et d'écriture du modèle produisent un résultat identique. Cette approche a pour principal avantage d'exposer explicitement l'impact des coûts sur le choix de participation. En ce sens, elle offre un cadre d'analyse intéressant pour l'offre de travail au noir en présence de coûts de participation à l'économie souterraine. Néanmoins, nous pouvons raisonnablement penser que la fiscalité est un des facteurs déterminants de l'exercice d'une activité souterraine. Nous sommes alors amenés à considérer des modèles qui proposent une formalisation précise du système fiscal. La seconde approche envisagée intègre explicitement le système fiscal et tente de rendre compte de la présence de coûts de participation. Les implications du système fiscal n'étant pas les mêmes selon qu'il s'agisse d'un impôt sur les revenus d'activité ou d'une attribution de prestations sociales, la modélisation des coûts diffère considérablement. L'introduction d'une structure fiscale plus réaliste s'effectue, en revanche, au prix d'une formalisation complexe et d'une procédure d'estimation extrêmement coûteuse. Les contributions théoriques et empiriques restent ainsi très limitées dans ce domaine. La majeure partie des travaux, visant à incorporer le système fiscal, s'inscrivent dans une démarche simplifiée reposant notamment sur l'exploitation de variables instrumentales.

Dans les développements qui suivent, nous présentons les différentes manières dont les coûts de participation ont été introduits dans la littérature. Puis, nous envisageons plus spécifiquement l'introduction de coûts du travail dans un modèle où la structure fiscale est formalisée. Nous rappelons alors brièvement les caractéristiques essentielles du modèle d'offre de travail lorsque le système fiscal génère une non linéarité de la contrainte budgétaire.

### 2.2.1 Coûts fixes et intensité de participation

La présence de coûts pour l'achat de biens de consommation ou l'entrée sur certains marchés a pu être révélée par un certain nombre de travaux précurseurs.

Ainsi, Cragg (1971) développe plusieurs modèles à variables dépendantes limitées et identifie des coûts de recherche importants dans l'acquisition de biens durables. Il remarque que la caractéristique essentielle des variables étudiées est leur probabilité non négligeable d'être égale à zéro pour un grand nombre d'observations. Il existe,

en effet, de nombreux consommateurs pour lesquels la probabilité d'acheter certains biens pendant une période donnée est nulle. En revanche, lorsqu'un bien est acheté, le montant dépensé varie très peu d'un consommateur à l'autre. Dans certains cas, l'acquisition peut se produire si l'achat désiré est, en quelque sorte, positif. Mais, il existe de nombreuses situations pour lesquelles des coûts peuvent empêcher de réaliser les achats désirés. Dès lors, les modèles statistiques doivent tenir compte du fait que les variables se répartissent sur un intervalle de valeurs borné inférieurement. Cela signifie que la décision d'achat et le montant des dépenses doivent être distinguées. L'approche proposée par Cragg est une extension des modèles de type tobit. Elle diffère néanmoins du tobit simple en permettant aux déterminants du montant de la variable, lorsqu'elle n'est pas nulle, d'être différents des déterminants de la probabilité d'être nulle. Il compare la capacité de différents modèles à rendre compte de l'existence de coûts fixes. Aucun de ces modèles n'est pleinement satisfaisants dans la mesure où ils ne tiennent pas compte de la possibilité de revenus nuls<sup>5</sup>. Néanmoins, parmi les spécifications retenues, il s'avère que le modèle basé sur l'estimation de la participation par probit suivie de la régression du logarithme du montant des dépenses, ou du montant lui-même, a le pouvoir prédictif le plus élevé. En outre, quelle que soit la spécification retenue, l'hypothèse selon laquelle différentes variables ont le même effet sur les deux composantes du modèle (participation et volume des dépenses) est rejetée. Ces résultats attestent de la présence de coûts fixes affectant le comportement des individus et révèlent les limites des spécifications de type tobit simple dans ce cas.

En ce qui concerne l'offre de travail, une distinction est introduite pour caractériser les coûts fixes. En effet, ils sont considérés, à la fois, dans leur composante monétaire, mais également sous l'aspect du temps dépensé par l'entrée sur le marché du travail (Cogan 1980, Hanoch 1980).

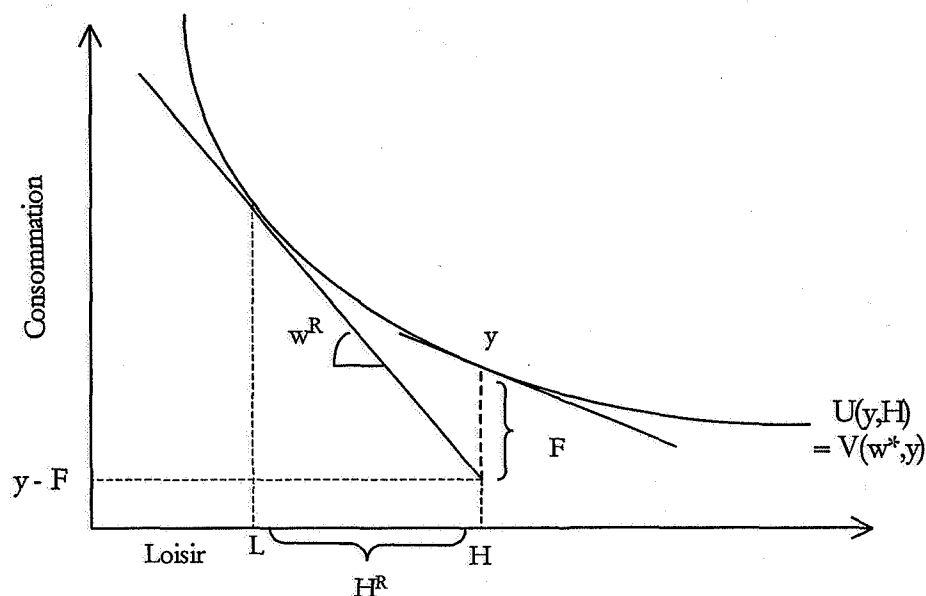
La *figure 2-1* illustre les différents coûts générés par l'exercice d'une activité de travail.  $H$  et  $Y$  représentent respectivement les dotations de l'individu en temps et en revenu hors-travail. Les variables  $F$  et  $t$  désignent le coût monétaire fixe et la perte de temps provoqués par l'entrée sur le marché du travail. Le salaire de réserve  $w^R$  est l'offre de salaire nécessaire pour inciter l'individu à travailler.

Comme cela est illustré à la *figure 2-1*, les individus ne sont pas disposés à travailler en-dessous d'un nombre minimum d'heures, qualifiées d'heures de réserve et notées  $H^R$ .

---

<sup>5</sup>Ce problème a été résolu par simple élimination des observations pour lesquelles le revenu est égal à zéro.

Figure 2-1 :  
Offre de travail au noir en présence de coûts fixes à l'entrée



Ces auteurs soulignent l'effet des coûts sur la contrainte budgétaire de l'individu et examinent comment de telles considérations peuvent être incorporées dans un modèle d'offre de travail. Les coûts sont supportés quelle que soit la durée du travail choisie, mais ils peuvent varier d'un individu à l'autre. Leur prise en considération complique significativement l'analyse dans la mesure où le niveau de travail optimisé sous la contrainte budgétaire, lorsque l'individu travaille, peut ne pas représenter le choix optimal. Dès lors, le choix de ne pas travailler, et donc de ne pas subir les coûts, doit être explicitement examiné. Autrement dit, tant au point de vue théorique qu'empirique, nous devons modéliser séparément la décision d'entrer sur le marché et le choix des heures de travail. En effet, la décision de travailler ou de ne pas travailler n'obéit vraisemblablement pas aux mêmes raisons que la décision d'accroître ou de réduire le nombre d'heures travaillées pour les individus participant au marché du travail. Les individus ne travailleront pas, à moins que le gain soit suffisamment important pour compenser les coûts fixes. Un nombre minimum d'heures est requis, pour tout niveau de coût, créant un intervalle atteignable dans la distribution des heures observables. Pour les participants, en revanche, les coûts fixes ont déjà été supportés et représentent un coût irrécupérable. L'introduction de coûts fixes génère une discontinuité dans la contrainte budgétaire, invalidant les simples procédures de maximisation.

Un certain nombre d'études complémentaires ont incorporé des coûts du travail dans leurs spécifications économétriques. La plupart d'entre elles modélise les coûts

comme un coût fixe à l'entrée sur le marché du travail. Cogan (1981), par exemple, estime par maximum de vraisemblance, un modèle de participation, de salaires et d'heures de travail qui incorpore des coûts fixes. Il contribue à améliorer l'approche développée par Cragg dans la mesure où il spécifie et estime une équation de salaire. Il considère les combinaisons de salaire et heures de travail dans la détermination de l'offre de travail. Une procédure en deux étapes lui permet d'obtenir des estimateurs convergents des paramètres de l'équation de salaire, tout en tenant compte du biais de sélection relevant du choix de l'activité. Il maximise ensuite la fonction de vraisemblance conditionnelle aux valeurs estimées. Les résultats obtenus, concernant les femmes mariées américaines, confirment la présence de coûts fixes importants à l'entrée sur le marché. Ils s'élèvent à 1022 dollars pour les chômeurs et 949 dollars pour les travailleurs. Les gains annuels sont réduits de 28% en moyenne après inclusion des coûts fixes. Cela explique que certaines femmes ne travaillent pas en raison de coûts du travail trop élevés. Par ailleurs, il montre qu'aux Etats-Unis, l'élasticité de l'offre de travail par rapport au salaire des femmes mariées est de 0,5 ou de 2 selon que les coûts fixes sont inclus ou non. De même, l'étude réalisée au Royaume-Uni par Blundell, Ham et Meghir (1987) sur les femmes mariées a montré une baisse significative de leur élasticité salaire après prise en compte des coûts fixes.

Blau et Robins (1988) se sont intéressés, quant à eux, à l'offre de travail des ménages. Ils incorporent des coûts liés aux soins des enfants dans les contraintes budgétaires des femmes mariées. Ils estiment un modèle logit multinomial et trouvent que ces coûts affectent significativement le comportement sur le marché du travail. De même, Ribar (1992), en proposant une extension de ce dernier modèle, montre que les coûts de soin aux enfants influence la décision de participation des femmes.

Enfin, dans un modèle de choix discret d'offre de travail, Hoynes (1996) analyse les effets de programmes sociaux (AFDC). Cet auteur modélise la contrainte budgétaire d'une famille pour diverses combinaisons d'emploi et d'heures de travail pour les femmes et leurs maris. Elle ajoute des coûts fixes à l'entrée sur le marché du travail et montre leur significativité.

Comme nous venons de le voir, les coûts du travail sont généralement assimilés à des coûts fixes à l'entrée sur le marché. Il convient de noter également que la plupart des études empiriques considère un horizon temporel annuel, voire mensuel. Or, selon l'intervalle de temps couvert par la base de données, des coûts fixes constituent ou non une approximation raisonnable des coûts réels supportés par l'individu. Cogan (1981) souligne, en effet, que la modélisation de coûts fixes implique que les données correspondent à la fréquence à laquelle ces coûts apparaissent. Autrement dit, la dimension



temporelle de l'offre de travail et celle durant laquelle les coûts sont supportés doivent être semblables. Si les coûts sont journaliers, la mesure idéale de l'offre de travail, en présence de coût, est basée sur les heures journalières. Cette spécification est appropriée dans la mesure où les coûts quotidiennement subis (les déplacements domicile-travail, par exemple) sont certainement les mêmes, que l'individu travaille une ou huit heures. Ce type de stratégie a été adoptée par Blank (1988) qui introduit des coûts horaires et hebdomadaires dans un modèle d'offre de travail de forme réduite. Elle montre que le nombre de semaines travaillées par an est choisi différemment du nombre d'heures par semaine. Elle explore ainsi la nature différente du choix des semaines et des heures de travail. Bien que ces décisions soient prises conjointement, elles dépendent de facteurs différents. Elle insiste, par ailleurs, sur la nécessité de distinguer ces décisions de celle de travailler ou non. Les résultats des estimations révèlent, en outre, la présence de coûts fixes significatifs qui affectent considérablement l'offre de travail des femmes monoparentales américaines.

Les modèles, présentés dans ce paragraphe, reposent sur le principe que le salaire de réserve détermine la participation au marché du travail. Cette théorie prétend qu'il existe un salaire donné, pour tout individu, tel que les offres du marché supérieures à ce salaire l'incitent à travailler. Cependant, ces modèles ne tiennent pas compte du système fiscal auquel fait face l'individu, ni des non convexités qu'il est susceptible de générer. Cette théorie, essentiellement locale, étudie ce qu'il se passe au voisinage des heures nulles. Mais, la présence de non convexités créées par des coûts fixes et par des taux d'imposition décroissants impliquent qu'il n'existe pas forcément un salaire de réserve unique. Dans ce cas, les modèles s'avèrent incomplets. Lewis (1956), initialement, a montré que le concept de salaire de réserve unique n'est pas valide lorsque la contrainte budgétaire est non convexe. De même, Gronau (1973) a souligné que l'existence de coûts fixes peut invalider cette théorie. Par conséquent, le modèle de participation au marché du travail doit être étendu pour considérer conjointement les combinaisons d'heures et de salaires. Cela implique que la théorie du salaire de réserve doit être remplacée par une théorie fondée sur la comparaison des niveaux d'utilité.

### 2.2.2 Stigmate social et décision de participation

Les modèles reposant sur une comparaison du niveau d'utilité atteignable dans les deux états possibles (participation ou non) tendent à formaliser les coûts de participation sous la forme d'un stigmate social. Toutefois, très peu d'études se sont attachées à considérer la présence éventuelle de ce stigmate dans la détermination des comportements individuels. Quelques exceptions sont cependant à noter parmi lesquelles Moffit (1983), Cowell (1990), Scott et Garen (1994) et Fortin *et al.* (2000).

Si Cowell est le premier à évoquer la présence de stigmatisme associé à la participation au marché noir, il ne propose pas de modélisation formelle. Il précise, en revanche, la nature des interactions individuelles à l'origine de ce stigmatisme. L'offre de travail au noir résulte, selon lui, d'une interaction complexe entre l'évaluation par l'individu des conséquences de ses actions et de la perception qu'en ont les autres membres de la société. Une manière de rendre compte des interactions entre les individus est de formuler l'hypothèse qu'une offre de travail au noir très répandue peut générer une externalité de consommation qui affecte le bien-être de chacun de ses membres. Cette proposition, également mentionnée par Gordon (1989), a été reprise par Fortin *et al.* (2000).

Une telle approche est animée par la notion de stigmatisme et par l'observation que la propension individuelle à frauder semble être fortement influencée par le nombre de personnes qui se comportent ainsi. L'analyse suppose que les préférences individuelles sont influencées par le comportement des autres. Supposons que  $Z$  représente les caractéristiques individuelles exogènes, incluant par exemple l'aversion au risque ou le goût pour les biens publics. Ces attributs individuels influencent la décision d'offrir ses services sur le marché noir. L'évaluation subjective de la perception du travail au noir, dans la société, est introduite dans la fonction objectif de l'individu. Ainsi :

$$V(e, E) = EU(C, e, Z, E)$$

où  $e$  caractérise l'offre de travail au noir et  $E$  représente la taille de l'économie souterraine ou le montant agrégé de la fraude, ou bien encore la proportion de travailleurs au noir dans l'économie. Cette fonction comporte désormais un quatrième argument qui rend compte de l'ampleur de l'activité souterraine dans la société. L'utilité espérée individuelle est exprimée en fonction de l'offre de travail au noir de l'individu ( $e$ ) et de la proportion de travailleurs au noir dans l'économie ( $E$ ). Celle-ci nous permet de capter le phénomène d'interaction sociale. L'offre de travail au noir est néanmoins conditionnée par les paramètres fiscaux et par le revenu hors-travail. Mais, ces éléments ont été supprimés par simple commodité de notation.

L'introduction du terme d'interaction  $E$  dans la fonction d'utilité répond à plusieurs justifications. La première raison résulte de l'existence d'un stigmatisme qui peut affliger l'individu si son comportement frauduleux est découvert. L'opprobre sera d'autant plus fort que l'individu se perçoit lui-même comme faisant partie d'une minorité. En revanche, si le travail au noir est extrêmement répandu, l'individu se jugera moins sévèrement. D'autres mécanismes sociaux peuvent produire un effet similaire. Par exemple, l'apprentissage d'une activité souterraine par un proche ou un parent

peut être un élément décisif dans le comportement de fraude.<sup>6</sup> De la même manière, les coûts de recherche d'un travail au noir se réduisent à mesure que la taille de l'économie souterraine s'accroît. Cette diminution des coûts a pour conséquence un niveau d'utilité plus élevé pour les travailleurs au noir.<sup>7</sup>

Le signe des dérivées premières et secondes de  $V(e, E)$  dépend de la nature de l'externalité dans le modèle. Ainsi, lorsque l'individu s'expose à la réprobation de la fraude par la société, l'utilité marginale de son activité non déclarée est diminuée :

$$V_e(e, 0) < 0$$

où  $V_e$  est la dérivée première de  $V$  par rapport à  $e$ .

En revanche, le signe de la dérivée de  $V$  par rapport à la taille de l'économie souterraine,  $V_E$ , est plus délicat à déterminer. En effet, l'accroissement du nombre de participants au marché noir accroît l'utilité marginale de l'individu, si celui-ci participe à ce marché, de sorte que

$$V_E(h, E) > 0.$$

Mais, s'il ne travaille pas au noir, l'individu peut être scandalisé du développement de la fraude. Ainsi,

$$V_E(0, E) < 0.$$

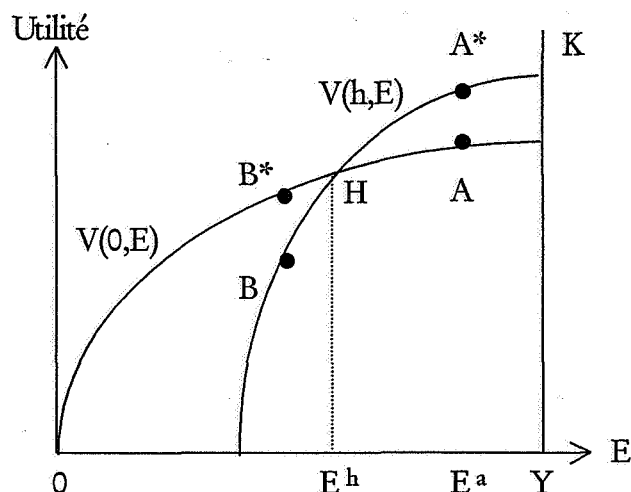
Les points importants du modèle sont illustrés à la figure 1. Le nombre de participants au marché noir est représenté sur l'axe des abscisses et le niveau de l'utilité sur l'axe des ordonnées.

Un élément essentiel est que, quelle que soit l'exacte spécification, la courbe  $V(h, E)$  coupe la courbe  $V(0, E)$ . Pour simplifier l'analyse, supposons que l'offre de travail au noir soit une variable dichotomique, prenant pour uniques valeurs *zéro* ou  $h$ . La fonction  $V(e, E)$  est représentée par deux courbes,  $V(0, E)$  et  $V(h, E)$ , correspondant à chacune de ces deux valeurs.

<sup>6</sup>Lea *et al.* (1987) ont montré l'importance, pour les parieurs, de l'expérience acquise avec un proche.

<sup>7</sup>Le nombre de fraudeurs que l'individu connaît personnellement a fréquemment été cité comme facteur prédisposant à la fraude (Spicer et Lundstedt, 1976, Vogel, 1974).

Figure 2-2 :  
Offre de travail au noir en présence de stigmatisme social



En accord avec les hypothèses développées précédemment, nous observons qu'au départ,  $V(0, E)$  se situe au-dessus de  $V(h, E)$ , puis le rapport s'inverse. La position des deux courbes dépend des préférences de l'individu : lorsque celui-ci dispose de revenus plus élevés, la courbe  $V(h, E)$  se déplace vers le haut par rapport à  $V(0, E)$  ; lorsqu'il est relativement averse au risque, la pente de  $V(h, E)$  est plus forte.

Afin d'examiner la question du poids de la société sur les choix individuels, considérons le cas d'un individu complètement honnête dans une économie largement dominée par le travail au noir. Il se situe sur la courbe  $V(0, E)$ , en un point tel que le point  $A$  de la figure 1. Ce point n'est manifestement pas un point d'équilibre puisque pour cette valeur de  $E$ , il pourrait atteindre un niveau d'utilité plus élevé :  $V^a(h, E) > V^a(0, E)$ . S'il entre sur le marché noir, il passe du point  $A$  au point  $A^*$ , de sorte que son utilité augmente. Un argument similaire, mais de sens opposé, s'applique à un travailleur au noir vivant dans une économie majoritairement honnête. Un gain d'utilité est réalisé en se déplaçant de  $B$  (travail au noir) à  $B^*$  (abandon d'activité dissimulée). La courbe enveloppe définit le comportement optimal de l'individu au regard du comportement des membres de la société.

Ce modèle nous fournit un cadre théorique intéressant pour l'analyse des comportements d'offre de travail au noir en présence de stigmatisme social. Il doit néanmoins être enrichi pour permettre à l'individu de déterminer l'intensité de son offre, sur un intervalle de valeurs comprises entre une participation nulle et une participation totale

$[0, T]$ , où  $T$  est le temps dont il dispose. Il est, en effet, assez simpliste de considérer uniquement ces deux valeurs extrêmes. La décision d'offre de travail au noir est un processus à deux étapes. En premier lieu, l'individu choisit de participer ou non à l'économie souterraine. En second lieu, s'il a décidé d'entrer sur le marché noir, il détermine combien de temps il consacra à l'activité non déclarée. La première étape implique une possibilité de perte de réputation dans la société qui va au-delà des éventuelles pertes monétaires. Mais, une fois préparé au risque d'opprobre social en cas de détection, l'individu procède à l'évaluation du risque qu'il est prêt à prendre face aux autorités fiscales. L'intérêt de cette analyse repose sur l'explicitation des interactions individuelles propres au développement de l'activité souterraine. Or, la base de données dont nous disposons nous offre la possibilité de saisir l'impact de l'entourage sur l'offre de travail au noir. Nous reprendrons donc cette proposition en introduisant, dans notre modèle, une variable subjective d'évaluation du nombre de participants à l'économie souterraine dans l'entourage de l'individu.

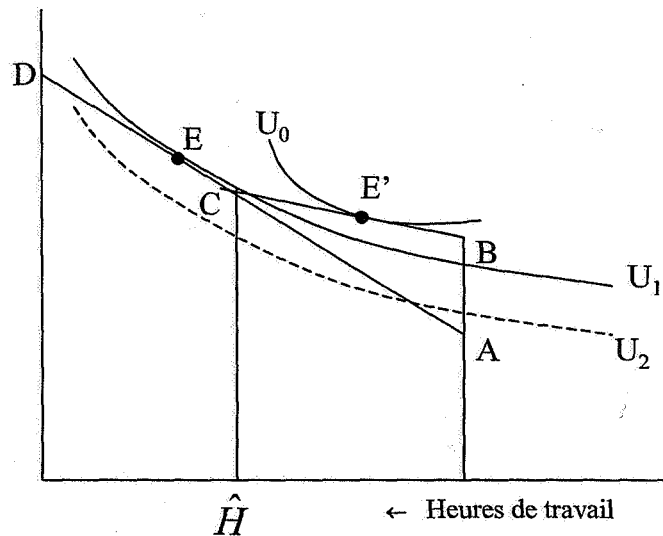
Comme nous l'avons mentionné plus haut, des coûts de participation à l'économie souterraine ont également été envisagés par Fortin *et al.* (2000). Ces auteurs testent la présence de coûts fixes associés à la consommation de biens sur le marché noir. Ils admettent la possibilité de solution de coins et de perte d'utilité induite par les achats de produits non déclarés. Un modèle de fraude fiscale avec revenus endogènes est développé pour tenir compte de la nature spécifique de l'activité souterraine. Basée sur une approche conditionnelle, le modèle de forme quasi-réduite est estimé selon la méthode du maximum de vraisemblance à information complète. Différentes spécifications du modèle sont considérées. La confrontation des estimations de type tobit simple et tobit type 2 montrent que le choix d'acheter sur le marché noir et le montant des dépenses engagées subissent une influence différente des variables explicatives. En revanche, le test de Chow proposé par Scott et Garen (1994) ne permet pas d'accepter l'hypothèse de l'existence de coûts fixes au niveau de l'achat de biens de consommation sur le marché noir.

Moffit (1983), quant à lui, s'intéresse au stigmatisme associé à l'attribution de minima sociaux. Il développe un modèle d'offre de travail pour expliquer que de nombreux individus n'effectuent pas les démarches nécessaires pour bénéficier de prestations sociales, en dépit des sommes substantielles dont ils pourraient bénéficier et alors que leur situation financière le leur permettrait. Le modèle standard d'offre de travail suppose, en effet, que tous les individus éligibles aux prestations sociales perçoivent effectivement ces transferts. Les individus qui travaillent en-dessous du seuil minimal requis pour bénéficier d'allocations mais qui ne les reçoivent pas, sont en-dessous de leur contrainte budgétaire, ce qui est impossible dans le cadre de l'analyse standard.

Pourtant, de nombreux individus sont dans ce cas. Une telle situation est expliquée par l'hypothèse d'un stigmate, ou perte d'utilité, associé au bénéfice de prestations sociales.

La situation peut être illustrée par la *figure 2-3*.

Figure 2-3 :  
Offre de travail et stigmate lié aux programmes sociaux



La droite  $ACD$  représente la contrainte budgétaire d'un individu en dehors des programmes sociaux et la droite  $BC$  est celle qui correspond à l'octroi de prestations. L'utilité hors assistance est maximisée au point  $E$ , c'est-à-dire au-delà des heures d'éligibilité  $\hat{H}$ . En l'absence de stigmate, l'individu choisira de diminuer ses heures en-dessous de  $\hat{H}$  et pourra ainsi prétendre aux prestations sociales. L'utilité maximale, dans ce cas, sera atteinte au point  $E'$ . Mais, la présence de stigmate implique que le niveau d'utilité associé aux programmes sociaux est inférieur, comme l'illustre la courbe d'indifférence en pointillé ( $U_2$ ). Si  $U_2$  se trouve en-dessous de la courbe  $U_1$ , l'individu ne réduira pas ses heures en-dessous de  $\hat{H}$  et restera en  $E$ . Il ne cherchera donc pas à percevoir de prestations sociales. Si, en revanche,  $U_2$  est supérieure à  $U_1$ , l'individu choisira de restreindre ses heures à un niveau inférieur à  $\hat{H}$  et bénéficiera de revenus de transfert. Par conséquent, un individu initialement inéligible aux programmes sociaux ne limitera son activité que si l'utilité additionnelle, issue d'un niveau de loisir plus élevé, compense non seulement la perte de revenu, mais également le stigmate provoqué par le versement de prestations sociales.

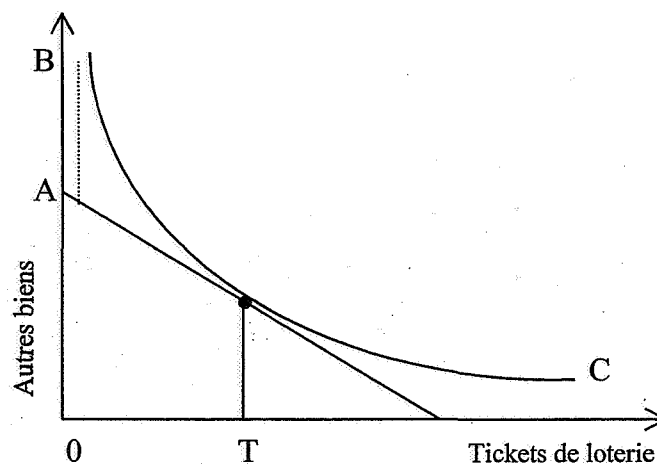
L'estimation du modèle nécessite de spécifier une forme fonctionnelle à l'équation d'offre de travail et aux fonctions d'utilités directe et indirecte. La décision de parti-

ciper aux programmes sociaux repose sur la comparaison du niveau d'utilité atteint par l'individu dans chaque état (participation ou non). La spécification du système d'équations, retenue par ces auteurs, est linéaire et les termes d'erreurs sont additifs. Puisque les paramètres inconnus associés au stigmate sont linéaires, cela revient à supposer une hétérogénéité de goût pour le travail ou de rejet pour la dépendance aux programmes sociaux dans la population. Ces paramètres sont alors considérés comme des fonctions de variables observées telles que les caractéristiques individuelles. Le modèle constitue un système d'équations simultanées avec deux variables endogènes, *i.e.* les heures de travail et la décision de participation. Puisque l'une de ces deux variables est dichotomique, le système est non linéaire, ce qui implique que les techniques simultanées conventionnelles ne peuvent être appliquées. Une procédure d'estimation telle que proposée par Heckman (1979) est mise en œuvre. Elle consiste à maximiser la fonction de vraisemblance non linéaire, en utilisant les probabilités jointes d'observer la participation aux programmes sociaux et un nombre d'heures de travail.

Les résultats obtenus à partir de cette méthode montrent que le stigmate apparaît au niveau de la décision de dépendre de programmes d'assistance, mais ne varie pas avec le montant des prestations reçues. En revanche, la probabilité de participer à ces programmes diffère considérablement avec la taille des revenus escomptés.

L'analyse développée par Moffit a été reprise par Scott et Garen (1994) pour étudier la participation à des jeux tels que les loteries. Ils remarquent que de nombreuses personnes, dans l'échantillon, choisissent de ne pas participer à ces jeux. Ils considèrent dès lors que l'achat de tickets de loterie peut véhiculer un stigmate ou un coût fixe. Ces derniers affectent la fonction d'utilité et génèrent une discontinuité en zéro.

Figure 2.4 :  
Participation aux jeux de loterie en présence de stigmate



La *figure 2-4* illustre la courbe d'indifférence dans son ensemble, soit  $ABC$ . Lorsque l'individu décide de participer à une loterie, la courbe a la forme habituelle. Cependant, lorsque le niveau d'achat de tickets de loterie est nul, la courbe d'utilité est caractérisée par le segment vertical  $AB$ . Cette distance verticale représente le stigmat. Elle indique que si  $T = 0$ , l'individu n'a pas à supporter la désutilité provoquée par la participation au jeu de loterie. Il est indifférent entre recevoir des revenus pour un montant  $OA$  et recevoir des revenus  $OB$  en participant à la loterie pour un montant infiniment petit. Il convient de noter qu'en l'absence de stigmat, la courbe d'indifférence est identique à celle du modèle standard. En ce sens, l'approche standard est un cas particulier du modèle proposé par Scott et Garen. Ils généralisent l'analyse des solutions de coin, dans la mesure où le prix de réserve est habituellement défini par la pente de la courbe d'indifférence au point où  $T = 0$  et les dépenses de réserve sont, par définition, nulles. Or, en présence de stigmat, ces dernières sont positives. Cela implique, par conséquent, de distinguer la décision de participer et le montant des dépenses engagées dans les jeux.

La solution du problème est obtenue selon une procédure en deux étapes. La première consiste à déterminer les valeurs de  $T$  et de  $X$  (autres biens ou niveau de consommation) qui maximisent la fonction d'utilité, conditionnelles à  $T > 0$ . Cela revient à traiter le stigmat comme fixe et à l'ignorer. Les choix de  $T$  et  $X$  ne sont pas affectés par le stigmat. La seconde étape implique une comparaison du niveau d'utilité dans le cas où l'individu participe au jeu et dans le cas où il ne participe pas.

Pour l'estimation, une modélisation de type tobit paraît *a priori* logique en raison de la censure en zéro de certaines observations. Cependant, cette spécification impose une restriction contraignante. En effet, les variables explicatives sont supposées avoir le même effet sur la décision de participation et sur le niveau de jeu des participants. Ils développent alors une spécification qui n'impose pas cette restriction. L'examen des résultats du tobit simple laisse penser que certaines caractéristiques influencent fortement le niveau de jeu. Mais, au regard du tobit généralisé, il s'avère que certains facteurs affectent seulement le choix de participer et restent sans effet sur le montant des dépenses. En particulier, les considérations religieuses affectent considérablement la décision de jouer. Les individus de confession catholique ont, par exemple, tendance à jouer davantage que les fondamentalistes. Mais, ces caractéristiques ne sont pas significatives pour le niveau de jeux des participants. Ces résultats confirment que les loteries peuvent véhiculer un stigmat tel qu'il dissuade certains individus de participer à la loterie, sans qu'il n'ait aucun effet sur le niveau de jeu parmi ceux qui participent. La comparaison des estimations atteste de la présence de coûts fixes de participation dont la simple estimation du modèle tobit ne permet pas de rendre compte. Enfin,



un test de Chow appliqué aux estimations contraintes et non contraintes apporte un appui supplémentaire à l'existence de coûts fixes.

Les contributions, développées dans ce paragraphe, offrent un cadre conceptuel plus approprié que celui de la théorie du salaire de réserve dans la mesure où elles permettent la comparaison du niveau d'utilité atteint dans chaque état (travail ou non). Cependant, elles ne modélisent pas explicitement le système fiscal. Or, nous pouvons raisonnablement penser que la fiscalité influence fortement la décision de travailler ou non au noir. C'est pourquoi, il importe désormais de préciser les différentes manières de formaliser la fiscalité dans un modèle d'offre de travail en présence de coûts fixes. Avant de développer les implications de ces coûts dans une structure fiscale plus réaliste, il est utile de rappeler brièvement les caractéristiques essentielles du modèle d'offre de travail lorsque le système fiscal génère une non linéarité de la contrainte budgétaire.

### 2.2.3 Coûts fixes et structure fiscale

Dans le modèle d'offre de travail statique standard, les individus déterminent leurs heures de travail et leur consommation en maximisant une fonction d'utilité  $U(C, h)$  par rapport à la contrainte budgétaire :

$$C = Wh + Y - T(I) \quad (2.1)$$

où  $C$  représente la consommation,  $W$  le salaire horaire brut,  $h$  le nombre d'heures de travail sur le marché,  $Y$  est le revenu hors-travail,  $T$  est l'impôt déterminé par la fonction  $T(\cdot)$ ,  $I$  est le revenu imposable annuel avec  $I = Wh + Y - D$  et  $D$  caractérise les déductions fiscales annuelles.

Les complexités introduites par le système fiscal et de redistribution créent une distorsion de la contrainte budgétaire. En effet, les nombreux taux d'imposition marginaux associés aux différentes tranches d'imposition et l'existence d'un revenu hors-travail impliquent que la contrainte budgétaire devient non linéaire. Les autorités fiscales calculent l'imposition sur la base du revenu, ce qui induit une contrainte budgétaire linéaire par segment. Les programmes sociaux, quant à eux, s'appliquent uniquement pour un niveau de revenus prédéfini. Ils génèrent donc également différentes tranches.

En l'absence de taxes, la maximisation de la fonction d'utilité par rapport à la contrainte budgétaire définit la fonction d'offre de travail suivante :

$$h = f(W, Y, X, v) \quad (2.2)$$

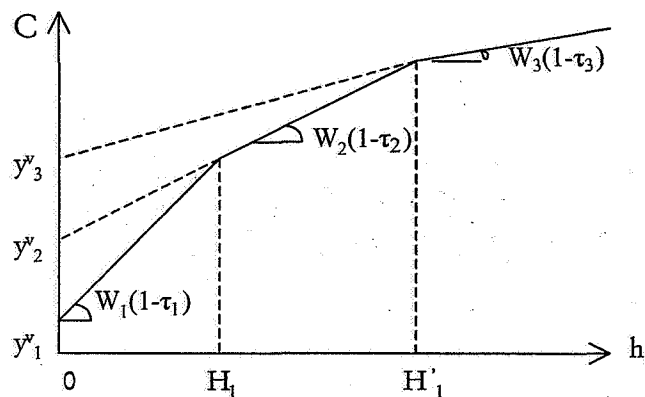
où  $X$  est un vecteur de variables influençant les choix de l'individu et  $v$  est un terme d'erreur reflétant la contribution de facteurs pertinents pour l'agent économique, mais

inobservable par l'économètre. Si l'on réinterprète  $W$  et  $Y$  comme des mesures «nettes d'impôt», la fonction  $f$  permet de décrire le comportement d'offre de travail en tenant compte du système fiscal. Cela crée des non linéarités complexes qui à leur tour affectent la contrainte budgétaire.

Les considérations fiscales occupent une large place dans la littérature sur l'offre de travail. La modélisation de la fiscalité est perçue comme essentielle à l'analyse de l'offre de travail afin de capter les opportunités disponibles pour les individus. La littérature a recours à deux approches pour modéliser ces non linéarités induites par le système fiscal. La première approche examine, étape par étape, chacune des tranches d'imposition inhérentes à la structure du système fiscal. La seconde approche consiste à linéariser les relations qui caractérisent ces tranches d'imposition et les taux marginaux correspondants. Ces deux perspectives permettent d'estimer les paramètres de la fonction d'offre de travail par le recours, soit à une procédure qualifiée de «méthode d'Hausman» ou de changement de régime, soit à la méthode des variables instrumentales.

L'approche des fonctions linéaires par segment peut être schématisée par la *figure 2-5*. Cette caractérisation est issue des travaux sur la fiscalité et l'offre de travail de Hall (1973). La *figure 2-5* illustre la solution optimale pour les heures de travail. Elle décrit la contrainte budgétaire d'un individu confronté à une imposition progressive de son revenu. Sur ce graphique, l'axe des ordonnées représente le revenu net d'impôt ou le niveau de consommation de biens marchands. Les segments correspondent aux différents taux d'imposition marginaux. En particulier, l'individu fait face au taux marginal  $\tau_1$  entre 0 et  $H_1$  heures de travail (segment 1), et respectivement au taux marginal  $\tau_2$  entre  $H_1$  et  $H'_1$  (segment 2) et au taux marginal  $\tau_3$  au-delà de  $H'_1$  (segment 3).

Figure 2-5 :  
Impôt progressif sur les revenus de travail



Les salaires nets associés à chaque segment sont les suivants :  $w_1 = (1 - \tau_1)W_1$  pour le segment 1,  $w_2 = (1 - \tau_2)W_2$  pour le segment 2 et  $w_3 = (1 - \tau_3)W_3$  pour le segment 3. Le revenu virtuel, *i.e.* le revenu correspondant à l'extrapolation linéaire de la contrainte budgétaire, est calculée par :  $y_1^v = Y - \tau(Y, 0)$ ,  $y_2^v = y_1^v + (w_1 - w_2)H_1$  et  $y_3^v = y_2^v + (w_2 - w_3)H_2$ .

Les changements de tranches d'imposition génèrent des coudes. Dans notre représentation graphique, ils sont au nombre de trois. En réalité, un système fiscal comporte de nombreux taux d'imposition<sup>8</sup>. La translation de ce système dans l'espace heures-consommation implique alors un grand nombre de coudes.

La seconde approche consiste à approximer le système fiscal par une fonction différentiable. Elle conduit au développement d'un modèle empirique d'offre de travail qui reconnaît l'influence des taxes. Une manière cohérente de construire cette fonction est d'approximer la fonction, en pallier, des taux d'impôt marginaux par une fonction différentiable. Cette approximation doit, en elle-même, être facilement intégrable de façon à obtenir une forme proche de la structure fiscale linéaire par segment.

Une fonction de ce type a été proposée par MaCurdy *et al.* (1990). Elle se présente ainsi :

$$\begin{aligned} \tau(I(h)) &= \tau_1 && \text{entre } I(H_1) \text{ et } I(H_2) \\ &= \tau_2 && \text{entre } I(H_2) \text{ et } I(H_3) \\ &= \tau_3 && \text{au-delà de } I(H_3) \end{aligned} \quad (2.3)$$

où  $\tau[I(h)]$  est le taux d'impôt marginal,  $I(h)$  le revenu imposable à  $h$  heures de travail et  $\tau_i$  est le taux d'impôt marginal (avec  $i = 1, 2, 3$ ). Pour simplifier l'exposé, supposons que  $\tau_1 = 0$ . Considérons l'approximation suivante du système fiscal, qui mobilise trois droites aux niveaux  $\tau_1 (= 0)$ ,  $\tau_2$  et  $\tau_3$ , et deux fonctions pondérées paramétrées pour basculer d'un segment à l'autre :

$$\hat{\tau}(I(h)) = \tau_2[\Phi_1(I(h)) - \Phi_2(I(h))] + \tau_3[\Phi_2(I(h))] \quad (2.4)$$

où les fonctions pondérées sont définies par la fonction de distribution cumulative  $\Phi_i(I(h))$  de moyenne  $\mu_i$  et de variance  $\sigma_i^2$  (avec  $i = 1, 2$ ). Le segment du milieu correspond à  $\tau_2$  et s'applique à un revenu imposable compris entre  $I(H_2)$  et  $I(H_3)$ . Afin de capter cet élément, nous devons paramétrer  $\Phi_1(\cdot)$  et  $\Phi_2(\cdot)$  avec les moyennes  $\mu_1 = I(H_1)$  et  $\mu_2 = I(H_2)$  respectivement, et avec des variances  $\sigma_i^2$  faibles. La première

<sup>8</sup>Le Québec compte ainsi 20 tranches d'imposition provinciales et 13 fédérales. Celles-ci se chevauchent et se composent de taux d'impôt différents (Gouvernement du Québec, 1999).

fonction de distribution,  $\Phi_1(\cdot)$  prend une valeur proche de zéro pour les niveaux de revenu inférieurs à  $H_1$  et donc bascule pour prendre une valeur égale à l'unité pour les revenus plus élevés. De manière similaire,  $\Phi_2(\cdot)$  prend une valeur de zéro jusqu'à  $I(H_2)$  et une valeur de un ensuite. La différence entre les deux vaut zéro jusqu'à  $I(H_1)$ , est égale à l'unité entre  $I(H_1)$  et  $I(H_2)$  et vaut zéro au-delà. Par conséquent, la différence prend une valeur égale à l'unité uniquement sur l'intervalle où  $\tau_2$  est pertinente<sup>9</sup>. En règle générale, l'ajustement des paramètres de moyennes et de variances permettent d'appréhender presque parfaitement chaque segment du système fiscal et de basculer relativement rapidement, tout en maintenant la différentiabilité aux points de basculement.

Une généralisation de cette approximation prend la forme suivante (Blundell et MaCurdy 1999) :

$$T'(I(h)) = \sum_{i=1}^k [\Phi_i(I(h)) - \Phi_{i+1}(I(h))] b_i(I(h)) \quad (2.5)$$

où les fonctions  $b_i(I(h))$  sont des polynômes de revenus. L'équation (2.5) fournit une bonne approximation de la contrainte budgétaire linéaire par segment, excepté que les coudes sont arrondis.

Le principal avantage de l'approche linéaire par segment est de tenir compte de l'endogénéité des taux d'impôts marginaux dans les estimations. Elle reconnaît ainsi l'influence du système fiscal sur l'offre de travail. Elle permet, en outre, de prendre en considération les éventuels coûts de participation au marché du travail. En ce sens, elle fournit un cadre d'analyse privilégié pour notre recherche.

Certaines applications de l'approche linéaire par segment incorporent donc des coûts fixes du travail. Hausman (1980) a tout d'abord étendu un modèle d'offre de travail pour considérer la possible non linéarité, voire même non convexité, de la contrainte budgétaire. Il étudie la participation des femmes au marché du travail en développant un modèle structurel qui tient compte de ces éléments. Il est alors contraint à envisager l'ensemble budgétaire complet. La décision de participer au marché du travail est déterminée par la comparaison de l'utilité atteinte dans chaque état (travail ou non). Il choisit de spécifier, en premier lieu, une fonction d'offre de travail linéaire.

<sup>9</sup> Il convient de noter qu'en ajustant les valeurs de  $\mu_1$  et  $\mu_2$ , nous pouvons contrôler le moment où la valeur un débute et celui où elle se termine. De même, en ajustant les variances des fonctions de distribution cumulative, nous pouvons contrôler la rapidité de basculement d'une tranche d'imposition à l'autre. Dans ce dernier cas, l'arbitrage repose sur le choix entre une transition graduelle plus brutale ou une transition plus précise.

Puis, l'identité de Roy et le principe de dualité lui permettent de déterminer les fonctions d'utilité directe et indirecte correspondantes. Les résultats des estimations par maximum de vraisemblance attestent de la présence de coûts fixes importants sur le marché du travail et expliquent que de nombreuses femmes choisissent de ne pas travailler. L'analyse, initialement concentrée sur le choix de participer ou non au marché du travail, a été prolongée par la suite pour permettre la détermination du volume horaire (Hausman 1985). Cette procédure, qualifiée de « méthode d'Hausman », a donné lieu à plusieurs travaux et s'est imposée dans de nombreuses estimations de l'offre de travail en présence de contrainte budgétaire linéaire par segment. Cette méthode présente de nombreux avantages et mérite que l'on s'y attarde quelque peu.

Envisageons le cas d'un individu disposant d'un revenu hors-travail noté  $Y$ . Celui-ci doit payer des coûts fixes monétaires,  $F$ , associés au travail. De part la présence de coûts fixes, le revenu hors-travail  $Y$  doit être remplacé par :

$$\begin{cases} Y - F & \text{si } h > 0 \\ Y & \text{si } h = 0 \end{cases} \quad (2.6)$$

$F$  étant partiellement observable, il est modélisé par un élément stochastique, variant selon les individus. Par conséquent, la contrainte budgétaire enregistre un saut discontinu vers le bas, correspondant à  $F$ , lorsque l'individu choisit de travailler.

Afin de résoudre l'optimum face à cette contrainte budgétaire, deux régimes doivent être considérés explicitement : *travailler* et *ne pas travailler*. L'estimation consiste à trouver l'utilité maximale dans chaque régime et à les comparer pour déterminer l'option qui est choisie.

Dans chaque régime, la fonction d'utilité  $U(C, h, v)$  -où un composant inobservé  $v$  est explicitement introduit- est maximisée par rapport aux équations (2.1) et (2.6). Ainsi la contrainte devient :

$$\begin{cases} C = Wh + Y - T(I) & \text{si } h = 0 \\ C = Wh + Y - F - T(I) & \text{si } h > 0 \end{cases} \quad (2.7)$$

Dans le régime 1 (l'individu ne travaille pas), la solution est relativement simple. L'utilité est donnée par :  $U(Y - T(Y - D), 0, v)$  dans la mesure où  $h = 0$ .

Dans le régime 2 (l'individu travaille), la solution est déterminée par une procédure en trois étapes. Selon une première étape, le choix du nombre d'heures de travail est conditionnel à la localisation de l'individu sur un segment particulier. Pour localiser

les coudes et les pentes de la contrainte budgétaire, il est nécessaire de connaître le niveau de revenu hors-travail, le taux de salaire brut, les heures de travail et la structure du système fiscal. Les heures associées aux coudes sont données par  $H_j = (I_j - Y - F + D)/W$  où  $Y$  et  $D$  représentent respectivement le revenu hors-travail imposable et les déductions fiscales, et  $I_j$  est le revenu imposable maximum pour le segment  $j$ . La pente de chaque segment est donnée par le taux d'impôt marginal pour ce segment :  $w_j = (1 - \tau_j)W$  où  $j$  est le segment,  $\tau_j$  représente le taux d'impôt marginal et  $W$  est le taux de salaire brut. Enfin, le revenu hors-travail correspondant à zéro heure de travail (l'intersection de l'axe vertical) est  $y_1 - F = V + Y - F - T(Y - D)$  où  $T(\cdot)$  est la fonction d'imposition évaluée au revenu imposable de l'individu. Les revenus virtuels associés à chaque segment successif sont calculés sur la base de l'application répétée de la formule suivante :  $y_j = y_{j-1} + (w_{j-1} - w_j)H_{j-1}$ .

Etant donnée la convexité de la contrainte budgétaire, le problème de maximisation de l'individu consiste à maximiser  $U(C, h, v)$  par rapport à :

$$\begin{aligned} C &= y_1 - F && \text{si } h = 0 \\ &= w_1 h + y_1 - F && \text{si } H_0 < h \leq H_1 \\ &= w_2 h + y_2 - F && \text{si } H_1 < h \leq H_2 \\ &= w_3 h + y_3 - F && \text{si } H_2 < h \leq \bar{H} \\ &= w_3 \bar{H} + y_3 - F && \text{si } h = \bar{H} \end{aligned}$$

La fonction d'offre de travail  $f(w, y, v)$  procure ainsi la solution de  $h$  donnée par l'expression suivante :

$$\begin{aligned} h &= 0 && \text{si } h = 0 && \text{(limite inférieure)} \\ &= f(w_1, y_1, v) && \text{si } 0 < h < H_1 && \text{(segment 1)} \\ &= H_1 && \text{si } h = H_1 && \text{(coude 1)} \\ &= f(w_2, y_2, v) && \text{si } H_1 < h < H_2 && \text{(segment 2)} \\ &= H_2 && \text{si } h = H_2 && \text{(coude 2)} \\ &= f(w_3, y_3, v) && \text{si } H_2 < h < H_3 && \text{(segment 3)} \\ &= \bar{H} && \text{si } h = H_2 && \text{(limite supérieure)} \end{aligned} \tag{2.8}$$

Une seconde étape est requise pour déterminer l'utilité maximale dans ce régime. Il convient d'associer un niveau d'utilité à chaque choix possible d'heures. Le long de chaque segment  $i$ , l'utilité est donnée par la fonction d'utilité indirecte  $V(w_j, y_j, v)$ . En revanche, au niveau des coudes, la fonction d'utilité directe est nécessaire et correspond à  $U(w_j H_j + y_j, H_j, v)$  pour le coude  $j$ . Posons :

$$f_j = f(w_j, y_j, v) \equiv \frac{V_w(w_j, y_j, v)}{V_y(w_j, y_j, v)}$$

où  $V_w$  et  $V_y$  représentent les dérivées partielles de  $V$ . Cette relation correspond à l'identité de Roy définissant la fonction d'offre de travail  $f$  évaluée aux niveaux de salaire  $w_j$  et  $y_j$ . Par conséquent, l'utilité maximale, lorsque l'individu travaille, est définie de la manière suivante :

$$\begin{array}{ll}
 -\infty & f \leq 0 \\
 V(w_1, y_1, v) & 0 < f_1 < H_1 \\
 U(w_1 H_1 + y_1, H_1, v) & f_2 < H_1 \leq f_1 \\
 V^*(w, y, v) = V(w_2, y_2, v) & H_1 < f_2 < H_2 \\
 U(w_2 H_2 + y_2, H_2, v) & f_3 < H_2 \leq f_2 \\
 V(w_3, y_3, v) & H_2 < f_3 < H_3 \\
 U(w_3 \bar{H} + y_3, \bar{H}, v) & f_3 \geq \bar{H}
 \end{array} \quad (2.9)$$

L'usage de  $-\infty$  pour  $h = 0$  indique simplement que l'option de ne pas travailler n'est pas incluse dans ce régime. Sélectionner cet élément revient à dire que le régime 1 (hors-travail) est préféré.

Etant donnée l'utilité maximale dans chaque régime, la dernière étape consiste à comparer les deux régimes. Un individu choisira de travailler aux heures spécifiées par l'équation (2.8) si

$$V^*(w, y, v) \geq U(Y - T(Y - D), 0, v) \quad (2.10)$$

et il choisira de ne pas travailler dans le cas contraire. En posant l'égalité des deux termes de l'équation (2.10), nous pouvons définir, quel que soit  $v$ , un seuil critique de coûts fixes  $F^*(v)$  au-dessus duquel l'individu préfère ne pas travailler.  $F$  entre dans cette relation par l'intermédiaire de la variable de revenu virtuel  $y$ . Dans la mesure où les heures de travail désirées augmentent avec  $v$ , cette valeur critique sera généralement croissante avec  $v$ . Une plus grande propension à travailler implique, en effet, que des coûts fixes plus élevés sont nécessaires pour dissuader l'individu de travailler. Si des restrictions sont placées sur  $F$ , de sorte que  $f > \underline{F}$ , certaines valeurs de  $v$  seront suffisamment faibles pour écarter le régime 2. Dans ce cas, il peut exister un important fossé dans le bas de la distribution de  $h$ .

Un autre élément à prendre en considération est l'absence d'observation de salaire pour les individus ayant choisi de ne pas travailler. Le salaire brut  $W$  ne peut être observé et, par conséquent, la contrainte budgétaire ne peut être dérivée. Cela implique que  $W$  doit être endogénéisé. La forme fonctionnelle la plus communément utilisée est la suivante :

$$\log W = \alpha' Z + \eta \quad (2.11)$$

où  $Z$  inclut toutes les variables observables déterminant  $W$  et  $\eta$  est un composant inobservable. Il est, en effet, courant, dans la littérature sur le capital humain de spécifier le salaire sous forme semi-logarithmique. Mincer (1974) a apporté une justification théorique à cette pratique et plusieurs auteurs, dont Heckman (1979) et d'autres par la suite, en ont fourni un appui empirique.

Afin d'explicitier la fonction de vraisemblance, nous devons considérer, en premier lieu, la contribution à la vraisemblance d'un individu qui choisit de ne pas travailler. Supposons que cette décision de «non travail» puisse être observée, de sorte qu'il n'y ait pas d'erreur de mesure. Dans ce cas, deux raisons expliquent la non participation au marché du travail :

- (i) les coûts fixes sont excessivement élevés, *i.e.*  $F > F^* \equiv F^*(v, \eta)$  quels que soient  $v$  et  $\eta$ .
- (ii) si les coûts fixes sont inférieurs à la valeur seuil (*i.e.*  $F^* \leq F$ ), alors les heures désirées sont faibles, de sorte que  $v < v^* \equiv v^*(\eta)$  quel que soit  $\eta$ <sup>10</sup>.

La probabilité de cet évènement est :

$$l_0 = \int_{-\infty}^{\infty} \int_{-\infty}^{v^*} \int_{F^*}^{\infty} g_{v\eta F}(v, \eta, F) dF dv d\eta \quad (2.12)$$

où  $g_{v\eta F}$  est la densité jointe de  $(v, \eta, F)$ .

Pour le régime 2, la contribution à la vraisemblance d'un individu qui choisit de travailler est la suivante :

$$\begin{aligned} l_1 = & \sum_{j=1}^3 \int_{v_j}^{\bar{v}_j} \int_0^{F^*} g_2[h^* - f_j, v, W - W(Z), F] dF dv \\ & + \sum_{j=1}^2 \int_{\bar{v}_j}^{v_{j+1}} \int_0^{F^*} g_1[h^* - H_j, v, W - W(Z), F] dF dv \\ & + \int_{\bar{v}_3}^{\infty} \int_0^{F^*} g_1[h^* - \bar{H}, v, W - W(Z), F] dF dv \end{aligned}$$

où  $g_1$  et  $g_2$  correspondent aux distributions de  $(\varepsilon, v, \eta, F)$  et  $(\varepsilon + v, v, \eta, F)$ ,  $v_j$  est la solution de l'équation  $f(w_j, y_j, v_j) = H_{j-1}$ , et où  $\bar{v}_j$  est la solution de l'équation  $f(w_j, y_j, \bar{v}_j) = H_j$ .

<sup>10</sup>La valeur critique  $v^*$  est obtenue en posant l'égalité des termes de l'équation (2.10) et en évaluant le revenu virtuel à  $F$ .



Définissons  $P_E = 1$  si l'individu travaille et  $P_E = 0$  autrement. La fonction de vraisemblance pour un individu est donc :

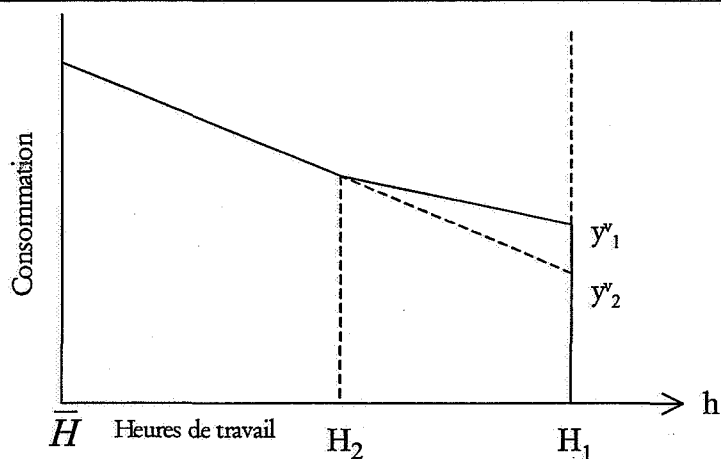
$$l = (l_1)^{P_E} (l_0)^{1-P_E} \quad (2.13)$$

La procédure d'estimation consiste à maximiser la somme des log-vraisemblances individuelles, comme habituellement. Néanmoins, la procédure dans ce cas est particulièrement complexe puisqu'elle nécessite de connaître, à la fois, l'utilité directe  $U$  et l'utilité indirecte  $V$ . Il est également nécessaire de comparer les régimes pour tous les individus et toutes les valeurs des paramètres.

Une analyse similaire a été proposée par Blundell et MaCurdy (1999) concernant l'éventualité d'un stigmate associé à la participation à des programmes sociaux. Afin de simplifier l'exposé, supposons que les seules sources d'imposition, auxquelles l'individu est soumis, sont les réductions des prestations sociales induites par l'accroissement de ses revenus d'activité. L'individu est confronté à des taux d'imposition effectifs très élevés dès lors qu'il travaille. En revanche, lorsque le montant des prestations devient nul, le taux d'impôt diminue.

Figure 2-6 :

Contrainte budgétaire en présence de programmes sociaux



Ces éléments impliquent une contrainte budgétaire coudée non convexe, comme illustrée par la *figure 2.6*. En raison de cette non convexité, la solution du problème de maximisation de la fonction d'utilité ne peut être caractérisée par le point de tangence avec la contrainte budgétaire. De nombreux points de tangence sont, en effet, possibles dans ce cadre et nous devons les comparer afin de déterminer l'optimum. Une procédure de changement de régime doit être mobilisée (Blundell et MaCurdy 1999). En

conséquence, la fonction d'utilité doit être modifiée pour rendre compte de la présence éventuelle d'un coût de participation. Celle-ci s'écrit désormais :

$$U = U(C, h, v) - P_B \xi \quad (2.14)$$

où  $P_B = 1$  si l'individu perçoit des allocations, 0 sinon et  $\xi$  définit le niveau de stigmatisme, strictement positif, qui varie selon les individus. L'individu maximise son utilité par rapport à la contrainte budgétaire suivante :

$$C = Wh + Y + B[I(h)] \quad (2.15)$$

où les prestations sociales  $B[I(h)]$  sont allouées selon le simple principe suivant :

$$B[I(h)] = \begin{cases} G - bWh & \text{si } G - bWh > 0 \\ 0 & \text{sinon} \end{cases}$$

Le montant minimum garanti  $G$  diminue à un taux  $b$  lorsque les revenus d'activité  $Wh$  augmentent. Cette spécification génère une non convexité dans la contrainte budgétaire et la formation d'un coude, noté  $H_2$ , au point  $H_2 = G/bW$ . Ce coude caractérise la situation où l'individu ne reçoit aucune prestation sociale ( $B[I(h)] = 0$ ), de sorte que l'individu est confronté à une contrainte composée de deux segments :

**Segment 1** :  $h < H_2$ ,

le salaire net est  $w_1 = (1 - b)W$  et le revenu virtuel est  $y_1 = Y + G$

**Segment 2** :  $h > H_2$

le salaire net est  $w_2 = W$  et le revenu virtuel est  $y_2 = Y$

L'introduction d'un stigmatisme implique, à nouveau, de distinguer deux régimes selon que l'individu participe ou non aux programmes sociaux. Considérons le régime pour lequel l'individu dispose de revenus d'activité suffisamment faibles pour lui permettre de prétendre aux prestations sociales, *i.e.*  $h < H_2$ . Le stigmatisme n'affecte pas les décisions marginales de l'individu, sachant que celui-ci est indemnisé. Par conséquent, le programme de maximisation conditionnel au salaire et au revenu hors-travail effectifs peut être résolu grâce à la fonction d'offre de travail définie par  $f(w_1, y_1, v)$ . Soit  $v_1$ , la valeur de  $v$  pour laquelle  $f(w_1, y_1, v) = 0$ . Le choix des heures optimales pour le premier segment est donné par :

$$\begin{aligned} h &= f(w_1, y_1, v) & \text{si } v > v_1 \\ h &= 0 & \text{si } v \leq v_1 \end{aligned}$$

L'utilité optimale est déterminée de la manière suivante :

$$V_1^*(w_1, y_1, v) = \begin{cases} V(w_1, y_1, v) - \xi & \text{si } 0 < f_1 \leq H_2 \\ U(y_1, 0, v) - \xi & \text{si } f_1 \leq 0 \\ -\infty & \text{si } f_1 > H_2 \end{cases} \quad (2.16)$$

où  $f_1$  est défini par  $f_j = f(w_j, y_j, v) \equiv \frac{V_w(w_j, y_j, v)}{V_y(w_j, y_j, v)}$ .

L'analyse du second régime diffère quelque peu dans la mesure où il est possible d'observer un individu avec  $h < H_2$ , mais ne bénéficiant pas de prestations sociales ( $P_B = 0$ ). La présence de stigmata entraîne que certains ayants droit n'effectuent pas les démarches pour bénéficier des revenus de transfert. Le régime 2 est uniquement défini par  $P_B = 0$ . Les heures de travail optimales, conditionnelles à  $w_2$  et  $y_2$ , sont données par :  $f(w_2, y_2, v)$ . En définissant  $v_2$  la valeur de  $v$  pour laquelle  $f(w_2, y_2, v) = 0$ , les heures de travail dans ce régime sont données par :

$$\begin{aligned} h &= f(w_2, y_2, v) & \text{si } v > v_2 \\ h &= 0 & \text{si } v \leq v_2 \end{aligned}$$

L'utilité optimale est désormais :

$$V_2^*(w_2, y_2, v) = \begin{cases} V(w_2, y_2, v) & \text{si } f_2 > 0 \\ U(y_2, 0, v) & \text{si } f_2 \leq 0 \end{cases}$$

Le choix du régime repose sur la comparaison de  $V_1^*$  et  $V_2^*$ . L'individu choisira le régime 1, avec indemnisation, si  $V_1^* > V_2^*$  et il choisira le régime 2 dans le cas contraire. Puisque la propension à travailler augmente avec  $v$ , ce choix peut être caractérisé par un seuil  $v^*$  défini par  $V_1^*(w_1, y_1, v^*) = V_2^*(w_2, y_2, v^*)$ . Pour les valeurs de  $v$  au-dessus de  $v^*$ , le régime 2 est sélectionné et pour les valeurs inférieures à  $v^*$ , le régime 1 est choisi.

Trois ensembles  $\Omega_0$ ,  $\Omega_1$  et  $\Omega_2$  sont définis tels que, pour  $v \in \Omega_0$ , l'individu choisit de ne pas travailler, pour  $v \in \Omega_1$  l'individu se localise sur le segment 1 avec un nombre positif d'heures et enfin, pour  $v \in \Omega_2$ , l'individu se situe sur le segment 2. Deux cas doivent alors être considérés pour définir ces ensembles exactement. En premier lieu, si  $v^* > v_1$ , nous obtenons :

$$\begin{aligned} \Omega_0 &= \{v | v \leq v_1\} \\ \Omega_1 &= \{v | v_1 < v \leq v^*\} \\ \Omega_2 &= \{v | v > v^*\} \end{aligned}$$

Alternativement, si  $v^* \leq v_0$ , le changement au régime 2 se produit avant que les heures dans le régime 1 soient effectives, *i.e.*

$$\begin{aligned}\Omega_0 &= \{v|v \leq v^*\} \\ \Omega_1 &= \emptyset \\ \Omega_2 &= \{v|v > v^*\}\end{aligned}$$

Pour tout  $v$  défini dans les ensembles  $\Omega_0$ ,  $\Omega_1$  et  $\Omega_2$ , il existe un niveau critique de  $\xi^* = \xi^*(v)$ , qui dépend de  $v$ , tel que le régime 2 est choisi lorsque  $\xi > \xi^*$ . Le régime 1 est choisi dans le cas contraire.

Enfin, la fonction de vraisemblance pour chaque combinaison de  $P_B$  et  $P_E$  peut être déterminée à partir de cette caractérisation et des densités jointes  $g_{v\xi\eta}(v, \xi, \eta)$  et  $g_{v\eta}(v, \eta)$  :

$$\begin{aligned}P_B = 1, \quad P_E = 0 \quad l_{10} &= \int_{-\infty}^{\infty} \int_{\Omega_0} \int_0^{\xi^*} g_{v\xi\eta}(v, \xi, \eta) d\xi dv d\eta \\ P_B = 0, \quad P_E = 0 \quad l_{00} &= \int_{-\infty}^{\infty} \int_{-\infty}^{v^+} \int_0^{\xi^*} g_{v\xi\eta}(v, \xi, \eta) d\xi dv d\eta\end{aligned}$$

La procédure d'estimation consiste à sélectionner la composante appropriée de la vraisemblance pour chaque individu et à maximiser la somme des log-vraisemblances. La fonction de vraisemblance est, une nouvelle fois, excessivement complexe à définir. En effet, pour chaque ensemble possible de valeurs des paramètres et pour chaque régime, un maximum doit être calculé. Ces maxima sont ensuite comparés afin de déterminer  $\xi^*$ .

Une transposition du modèle au marché noir pourrait être envisagée. Cependant, nous devons tenir compte de la présence d'un second marché du travail -non officiel- dans le choix de l'activité et l'intensité du travail. Dès lors, la complexité du modèle ne nous permet plus d'obtenir de prédictions testables empiriquement. A notre connaissance, aucune contribution n'a d'ailleurs pu être réalisée en ce sens. La complexité de la procédure a conduit les économistes à adopter une méthodologie simplifiée fondée essentiellement sur les variables instrumentales. Ainsi, l'approche alternative qui consiste à linéariser la contrainte budgétaire a été largement exploitée et présente de nombreux avantages pour les questions d'offre de travail qui nous préoccupent.

En particulier, Bourguignon et Magnac (1990) utilisent cette méthode et confirment la présence de coûts fixes significatifs. Ils estiment séparément les fonctions d'offre de travail pour les hommes et les femmes à partir d'un échantillon d'individus mariés et

travaillant en France. La contrainte budgétaire est linéaire par segment et supposée convexe. Les décisions d'offre de travail sont, par ailleurs, supposées séquentielles. L'homme détermine son offre sous l'hypothèse qu'aucun autre revenu de travail n'est perçu par la famille. Ensuite, la femme choisit sa propre offre de travail, étant donnée celle du chef de famille. Les auteurs considèrent le modèle d'offre de travail joint, en supposant que la contrainte budgétaire coudée est approximée par une fonction différentiable. Ils estiment le modèle par maximisation de la vraisemblance et utilisent l'estimateur de variables instrumentales pour estimer le modèle.

Les procédures de variables instrumentales non linéaires conventionnelles offrent, en effet, une méthode robuste pour l'estimation de fonctions d'offre de travail similaires à celle que nous avons spécifiée à l'équation (2.2). La forme fonctionnelle de cette équation permet la spécification d'équations structurelles non linéaires dans toutes les sources d'aléas. Le développement de telles spécifications constitue, néanmoins, un défi, dans la mesure où il est délicat de trouver une carte des préférences qui produise l'additivité dans les perturbations structurelles (les erreurs reflétant l'hétérogénéité inobservée parmi les individus) et qui, en même temps, permette aux erreurs de mesure d'entrer linéairement dans les heures et les salaires.

L'introduction d'un système fiscal non linéaire dans un modèle d'offre de travail soulève, en outre, quelques difficultés analytiques lorsque ce système génère une contrainte budgétaire strictement convexe avec des frontières doublement différentiables. La maximisation de l'utilité dans ce cas implique une simple caractérisation du choix des heures de travail. Admettons une fonction  $\tau$  approximant le système fiscal et spécifions le taux de salaire marginal et le revenu virtuel de la façon suivante :

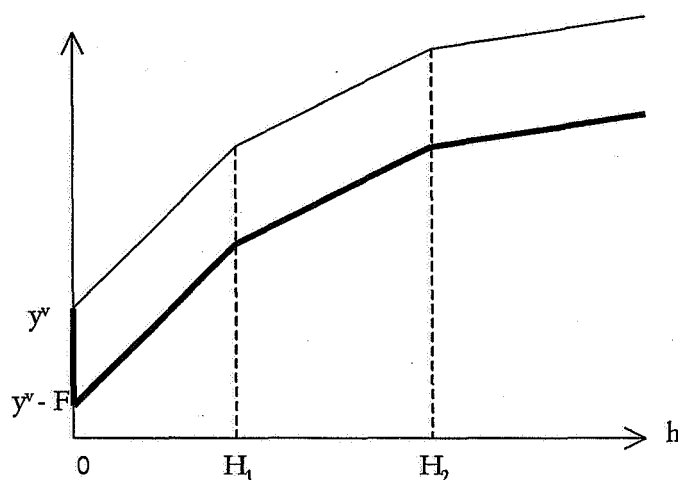
$$\begin{aligned} w &= w(h) = (1 - \tau)W \\ y^v &= y^v(h) = Y - F + Wh - T - wh = Y - F + \tau Wh - T = C - wh \end{aligned}$$

où  $F$  représente le coût de participation au marché du travail.  $T$  (montant de l'impôt net des transferts) et  $\tau$  (taux marginal d'imposition) sont évalués au niveau de revenu  $I = I(h) = Y + Wh - D$  qui dépend directement de  $h$ . Nous écrivons le salaire marginal  $w = w(h)$  et le revenu virtuel  $y^v = y^v(h)$  pour insister sur le fait qu'ils dépendent des heures de travail  $h$ . La maximisation de la fonction d'utilité implique une solution pour les heures de travail qui obéit à l'équation implicite suivante :

$$h = f(w(h), y^v(h), v) \quad (2.17)$$

La *figure 2.7* illustre la solution pour les heures optimales de travail lorsque la contrainte budgétaire est linéaire par segment, en raison de la présence d'un système progressif d'imposition et de coûts fixes.

Figure 2-7 :  
Coûts fixes et impôt progressif



Ces derniers entraînent un déplacement vers le bas de la contrainte budgétaire, excepté pour un nombre d'heures égal à zéro, où le coût n'est pas supporté. La contrainte est caractérisée par le point  $y^v$  lorsque l'individu ne travaille pas, puis dès que l'offre de travail devient supérieure à zéro, elle enregistre un saut vers le bas au point  $y^v - F$ . Les salaires nets sont les mêmes qu'en l'absence de coûts mais le revenu virtuel a diminué d'un montant équivalent à celui des coûts fixes.

La *figure 2.7* représente un individu confronté à une contrainte budgétaire linéaire en présence d'une fiscalité non linéaire. L'extrapolation de la contrainte sur l'axe des ordonnées équivaut à la linéariser. Cette contrainte linéarisée est construite de façon à la rendre tangente à l'ensemble des opportunités non linéaires effectives au niveau de la solution optimale des heures. La pente de cette contrainte linéarisée est  $w(h)$  et la valeur correspondante du revenu virtuel est  $y(h) - F$ .

L'équation (2.17) caractérise une relation structurelle qui détermine les heures de travail. En appliquant le théorème des fonctions implicites à cette spécification, nous pouvons résoudre pour  $h$  en termes de  $W$ ,  $Y$  et des autres paramètres associés aux fonctions  $\tau$  et  $f$ . Cette procédure génère une fonction d'offre de travail appropriée au

cas de l'imposition non linéaire. Elle fournit la base de la formulation d'une équation structurelle estimable par une procédure de variables instrumentales.

Considérons, par exemple, une spécification semi-logarithmique <sup>11</sup> :

$$h = f(w, y, v) = \alpha_1 + \alpha_2 Z + \alpha_3 \log w + \alpha_4 y + v \quad (2.18)$$

où  $\alpha_1, \alpha_2, \alpha_3$  et  $\alpha_4$  sont les paramètres à estimer,  $Z$  est un vecteur de déterminants observés de l'offre de travail (*e.g.* âge, taille de la famille, sexe...) et  $v$  est une perturbation structurelle captant les facteurs inobservés influençant les décisions d'offre de travail. Le taux de salaire marginal après-impôt  $w$  apparaît sous forme logarithmique dans l'équation. Ainsi,  $\alpha_3$  correspond à une forme hybride de l'effet substitution non compensé. Le coefficient  $\alpha_4$  caractérise l'effet revenu.

Les procédures de variables instrumentales procurent des estimateurs robustes des paramètres de la spécification semi-logarithmique de la fonction d'offre de travail donnée par l'équation (2.18). En l'absence d'erreur de mesure, l'examen de cette équation révèle que les termes d'erreur entrent linéairement dans la spécification. Par conséquent, les variables qui sont orthogonales à la perturbation  $v$  peuvent servir d'instruments pour estimer les paramètres déterminant les effets substitution et revenu. La mise en œuvre de telles procédures n'impose pas de restrictions paramétriques et permet de considérer une grande variété d'hypothèses d'exogénéité.

La majorité des études empiriques sur l'offre de travail, incorporant les taxes, traite le système fiscal comme une suite de tranches d'imposition. Cela implique une contrainte budgétaire linéaire par segment qui, la plupart du temps, est linéarisée pour faciliter les estimations. Néanmoins, une telle contrainte budgétaire génère l'endogénéité de certaines variables, notamment les heures et le salaire net d'imposition. Par conséquent, la méthode des variables instrumentales représente une solution adaptée à l'étude des coûts fixes sur l'offre de travail, tout en permettant de rendre compte de façon réaliste du système fiscal. Enfin, les variables endogènes sont, à la fois, discrètes et continues, ce qui complique la procédure et implique l'inclusion de termes de sélection de l'échantillon pour produire des perturbations de moyenne nulle.

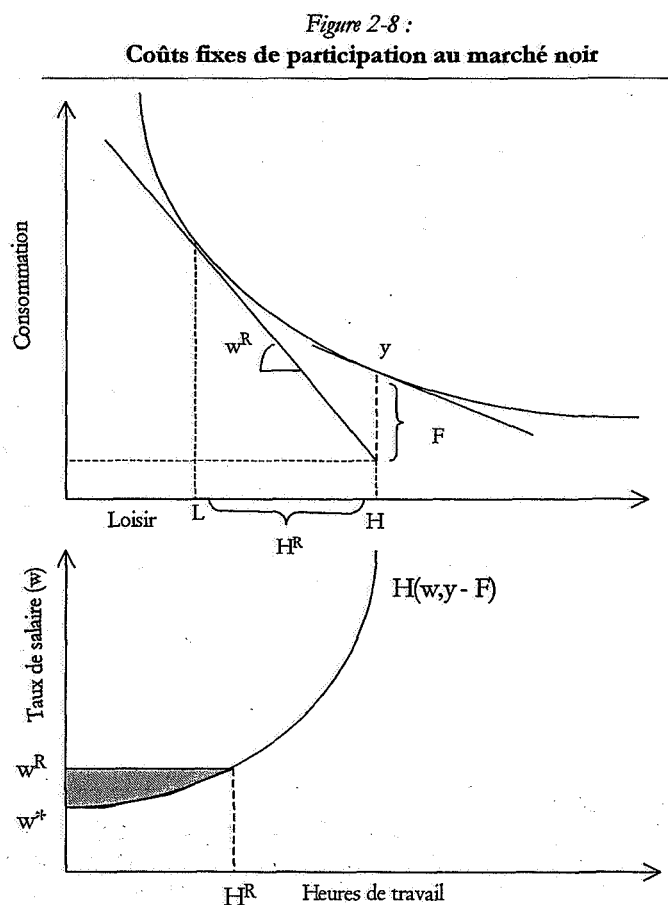
---

<sup>11</sup> Cette spécification de l'offre de travail est l'une des plus communément utilisées dans les travaux empiriques, en particulier dans le cadre de l'approche linéaire par segment (Blundell *et al.*, 1999). Elle permet un effet non linéaire du salaire. L'élasticité-salaire décline avec les heures, mais son signe est positif en tout point et il entre dans le revenu de façon linéaire. Cette formulation est particulièrement intéressante dans l'étude de phénomène de non-participation au marché du travail, de même que pour les cas d'endogénéité des salaires et du revenu. Enfin, la log-linéarité du salaire permet aux systèmes fiscaux proportionnels d'entrer linéairement dans la fonction d'offre de travail.

### 2.3 Un modèle d'offre de travail au noir en présence de coûts fixes

Notre modèle s'inspire de la littérature standard sur l'offre de travail. Il diffère néanmoins par l'introduction des paramètres de fraude fiscale avec revenus endogènes (Cowell 1985). Il admet, en outre, la possibilité de coûts fixes de participation à l'économie souterraine. En ce sens, il repose sur les travaux de Cogan (1981) et Moffit (1983).

Nous considérons la participation au marché noir comme un acte pouvant provoquer une perte d'utilité à l'individu. Cette désutilité peut alors le dissuader de participer, sans qu'elle n'affecte le niveau de participation parmi ceux qui choisissent de travailler au noir. Les implications de ces coûts d'entrée sur le marché noir peuvent être représentées à l'aide de la figure ci-dessous<sup>12</sup> :



<sup>12</sup> Afin de contourner un problème évident de lisibilité des représentations graphiques à trois dimensions, nous supposons l'existence d'un marché du travail unique.



La dotation en temps de l'individu et son revenu hors-travail sont notés respectivement  $H$  et  $y$ . La variable  $F$  caractérise la perte de revenu provoquée par l'entrée sur le marché noir. Si l'individu travaille, il devra dépenser  $F$  euros. La contrainte budgétaire est donc représentée par la droite définie à partir du point  $(H, y - F)$ . Si l'individu ne travaille pas, il n'aura pas à supporter le coût monétaire, de sorte que le point  $(H, y)$  est également atteignable. L'individu est alors indifférent entre travailler  $(H - L)$  heures et ne pas travailler du tout. L'effet d'une variation des coûts fixes sur le volume horaire des travailleurs est assez évident. Une élévation du coût  $F$  se traduit par une diminution des revenus de l'individu. Sous l'hypothèse que le loisir est un bien normal, les heures de travail offertes augmentent. En effet, lorsque le coût s'élève, la contrainte budgétaire enregistre un saut parallèle vers le bas. Le nombre d'heures de travail offertes se définit par un nouveau point de tangence avec une courbe d'indifférence inférieure à la précédente. Il s'ensuit un accroissement des heures de travail. Au même moment, l'alternative de ne pas travailler devient plus intéressante, provoquant une participation moindre au marché du travail. De ce point de vue, l'impact sur les heures de réserve est plus délicat à déterminer. Pour cela, il nous faut examiner la fonction de dépense de l'individu<sup>13</sup>.

Une première implication de la présence de coûts fixes sur le marché du travail est que l'individu n'est pas disposé à travailler en-dessous d'un certain nombre d'heures,  $h^R$ , qualifiées d'heures de réserve. Dès l'entrée sur le marché, l'individu subit d'importants coûts. Par conséquent, il choisit un volume horaire lui permettant de compenser les pertes subies. Le salaire de réserve, *i.e.* le salaire minimum nécessaire pour inciter l'individu à travailler, est représenté en  $w^R$ . Les heures de réserve sont, quant à elles, déterminées par  $h^R = H - L$  et sont décrites par la discontinuité de la fonction d'offre de travail au niveau du salaire de réserve. La non convexité de l'ensemble budgétaire crée une discontinuité dans l'offre de travail dont la taille dépend des préférences individuelles et des niveaux de  $y$  et de  $F$ .

Une deuxième implication de la présence de coûts fixes est la différence entre le salaire de réserve et la valeur que les individus assignent à leur temps (Cogan 1981). Cette dernière correspond à la pente de la fonction d'utilité au point  $y$ . De nombreux travaux ont tenté d'approcher cette valeur de différentes manières. Une première approche, fondée sur les participants au marché du travail, consiste à estimer le salaire de réserve  $w^R$ . Or, celui-ci est supérieur au salaire optimal  $w^*$ . La valeur attribuée au temps par les non travailleurs est donc surestimée. Une seconde approche consiste à estimer la fonction d'offre de travail pour un niveau nul d'heure. Dans ce cas, la pente

<sup>13</sup>Les propriétés de cette fonction seront détaillées plus loin.

de la courbe d'indifférence, nette des coûts fixes, est évaluée. Par conséquent, si le salaire offert  $w$  est inférieur au salaire optimal  $w^*$ , la valeur du temps correspondant à zéro heure de travail est sous-estimée. Mais, si le salaire offert  $w$  est compris entre le salaire de réserve  $w^R$  et le salaire optimal  $w^*$ , alors la valeur du temps est surestimée.

Admettons que les préférences individuelles soient représentées par la fonction d'utilité suivante :

$$U = U(h_1, h_2, C) \quad (2.19)$$

où  $h_1$  et  $h_2$  définissent respectivement, le nombre d'heures dans l'emploi officiel et le nombre d'heures de travail au noir.  $C$  désigne la consommation. La fonction d'utilité est supposée strictement concave, croissante en  $C$  et décroissante en  $h_1$  et  $h_2$ . La concavité provient du fait que l'individu est averse au risque.

Il est supposé que le taux de salaire  $W_1$  d'une activité déclarée est taxé à un taux proportionnel  $\tau$ . En revanche, un travail au noir est rémunéré au taux  $W_2$  et permet à l'individu d'échapper à l'imposition. Mais, il encourt le risque d'être contrôlé par les autorités fiscales -ce qui survient avec une probabilité  $p$ . A l'instar de Allingham et Sandmo (1972), cette probabilité est indépendante de  $h_2$ . Si la fraude est détectée, l'individu subit un redressement au taux de pénalité  $\theta$  (avec  $\theta > \tau$ ). Lorsque la pénalité s'applique sur les revenus dissimulés (Allingham et Sandmo 1972),  $\theta$  est fixe alors que lorsqu'elle s'applique sur l'impôt fraudé (Yitzhaki, 1974, Cowell 1990), on a :  $\theta = (1 + \lambda)\tau$  où  $\lambda$  est le taux de pénalité sur ce montant d'impôt fraudé. Par ailleurs, comme nous l'avons mentionné plus haut, la participation au marché noir engendre un coût fixe, noté  $F$ . L'utilité marginale du travail au noir devrait ainsi être inférieure à celle des revenus déclarés.

La nature dissimulée de l'activité souterraine implique également une limitation de son ampleur. Une fonction de gains non déclarés de type Cobb-Douglas et reliant ceux-ci aux heures non déclarées aurait pu être envisagée, tandis que les rémunérations du travail déclaré seraient modélisées par une fonction linéaire (Lemieux *et al.* 1994). Néanmoins, afin de rendre compte explicitement d'un coût variable relatif à l'activité souterraine, nous imposons la même forme fonctionnelle aux gains issus du marché noir et du marché officiel.

L'incertitude liée à la détection de l'activité souterraine rend la contrainte budgétaire stochastique, de sorte que celle-ci s'écrit :

$$C = m_1 + (N - F) = (y + w_1 h_1) + (w_2 h_2 - F) \quad (2.20)$$

$$\text{où } w_1 = W_1(1 - \tau) \text{ et } w_2 = \begin{cases} W_2 & \text{avec probabilité } (1 - p) \\ W_2(1 - \theta) & \text{avec probabilité } p \end{cases}$$

et où  $N$  et  $m_1$  désignent respectivement les revenus perçus sur le marché noir, nets des coûts fixes, et le revenu disponible de l'individu. Ce dernier est la somme du revenu hors-travail  $y$  et des revenus salariaux issus du marché officiel  $w_1 h_1$ . Enfin,  $W_1$  et  $W_2$  sont les salaires bruts de chaque activité.

Puisque nous souhaitons étudier l'impact des coûts fixes associés à l'économie souterraine sur l'offre de travail au noir, nous estimerons le modèle suivant l'approche conditionnelle initialement proposée par Pollack (1969) et plus récemment utilisée par Browning et Meghir (1991). Ainsi, nous dérivons l'offre de travail au noir conditionnelle à l'offre de travail sur le marché officiel. L'individu doit donc maximiser son espérance d'utilité sous la contrainte (2.20) en choisissant soit de participer, soit de ne pas participer, et en déterminant ses heures de travail. L'individu est alors confronté à deux décisions, l'entrée sur le marché noir et l'intensité de l'activité souterraine. Les déterminants de ces choix peuvent être sensiblement différents.

Une contrainte de non négativité est imposée sur les heures de travail dans chacun des marchés, de sorte que :  $h_1 \geq 0$  et  $h_2 \geq 0$ . Sous l'hypothèse que les axiomes de von Neuman-Morgenstern sous incertitude sont satisfaits, le programme à résoudre est alors le suivant :

$$\max_{\{h_2\}} EU(h_2, m_1 + (w_2 h_2 - F); h_1, z) \quad (2.21)$$

sous les contraintes de budget et de non négativité sur  $h_1$  et  $h_2$ .  $E$  est l'opérateur d'espérance mathématique et  $z$  un vecteur de caractéristiques individuelles.

Dans la mesure où la contrainte budgétaire (2.20) est respectée avec égalité stricte (en l'absence de saturation des préférences), il est possible de la substituer dans (2.21) et d'exprimer la fonction d'utilité espérée en termes de  $h_2$  et des variables exogènes. L'utilité marginale nette des heures de travail au noir est obtenue en dérivant partiellement la fonction d'utilité espérée obtenue par rapport à  $h_2$  :

$$\begin{aligned} m_2(h_1, h_2, z) &= EU_1 + E(U_2 w_2) \\ &= EU_1 + (1 - \tau)w_2 EU_2 + \tau w_2 E[U_2 R] \end{aligned} \quad (2.22)$$

où  $R$  désigne le taux de rendement d'un euro dissimulé aux autorités fiscales, avec :

$$R = \begin{cases} 1 & \text{avec probabilité } (1 - p) \\ -\lambda & \text{avec probabilité } p \end{cases}$$

$\lambda$  étant la pénalité sur l'impôt fraudé. Le terme  $EU_1$  représente l'effet d'une hausse de  $h_2$  sur l'espérance de l'utilité. Par conséquent,  $EU_1 < 0$ .

L'équation (2.22) nous permet de caractériser les conditions d'optimalité du travail sur le marché noir. Plus précisément, à l'optimum, les conditions de Kuhn-Tucker de premier ordre associées au programme (2.21) peuvent s'écrire de la façon suivante :

$$EU_1 + E(U_2 w_2) \leq 0 \quad (2.23)$$

$$h_2 [EU_1 + E(U_2 w_2)] = 0 \quad (2.24)$$

De (2.23) et (2.24) découle l'offre de travail au noir conditionnelle <sup>14</sup> :

$$h_2 = h_2(p, W_2, \lambda\tau, y, F, \bar{h}_1, z) \quad (2.25)$$

Les heures offertes sont fonction des taux de salaire sur chaque marché, des paramètres de répression de la fraude et du revenu non salarial. L'impact des différentes variables sur les heures de travail est donné par la statique comparative, notamment :

$$\frac{\delta h_2}{\delta w_1} < 0, \quad \frac{\delta h_2}{\delta \theta} < 0, \quad \frac{\delta h_2}{\delta p} < 0, \quad \frac{\delta h_2}{\delta w_2} > 0, \quad \frac{\delta h_2}{\delta \tau} \geq 0.$$

Une augmentation du taux salaire officiel ou du taux de redressement fiscal devrait réduire le nombre optimal d'heures de travail au noir. En revanche, une diminution de la probabilité de contrôle devrait avoir l'effet inverse, de même qu'une élévation du salaire non officiel. Quant au taux d'imposition, son impact sur les heures de travail au noir reste indéterminé.

Concernant la décision de participer au marché noir, définissons  $h_2^*$  la valeur de  $h_2$  maximisant l'utilité,  $y^*$  et  $N^*$  les montants correspondants des revenus déclarés et non déclarés calculés à partir de la contrainte budgétaire. L'existence de coûts fixes à l'entrée interdit de conclure à une renonciation au travail au noir si  $h_2^* < 0$ , et à une participation à hauteur de  $h_2^*$  heures de travail, si  $h_2^* > 0$ . En effet, même lorsque  $h_2^*$  est positif, l'individu peut s'abstenir de participer au marché noir, si la désutilité issue de l'entrée sur ce marché est suffisamment importante. Aussi, pour tenir compte de la présence des coûts fixes, l'individu compare l'utilité atteignable dans chacun des deux états et choisit celui qui lui procure la plus grande utilité. Considérons la fonction d'utilité indirecte  $V$  suivante :

$$V [W_2(1 - p\theta), y - F] = EU [h_2^*, y^* + N^* - F]$$

<sup>14</sup>Il convient de noter que la structure de la fonction d'offre de travail au noir est similaire à celle qui prévaut sur le marché officiel :  $h_1 [EU_1 + (1 - \tau)w_1 EU_2]$ . L'unique différence réside dans le terme correspondant au rendement de la fraude  $\tau w_2 E[U_2 R]$ .

où  $V$  correspond à l'utilité espérée maximale. Autrement dit,  $V(Ew_2, y - F) = EU[h_2(Ew_2, y - F), y - F + Ew_2h_2(Ew_2, y - F); \bar{h}_1, z]$ . L'individu choisira de travailler au noir si

$$EU(h_2^*, y^* - F + N^*; h_1, z) - EU(0, y; h_1, z) > 0$$

Afin d'accorder une attention plus précise à cette décision, nous suivons la méthode proposée par Cogan (1980) et reprise par Scott et Garen (1994). En isolant  $y$  de la fonction d'utilité indirecte, nous obtenons la fonction de dépense (ou de revenu hors-travail minimum) :

$$y = e(p, W_2, \lambda\tau, V; \bar{h}_1, z) + F$$

Celle-ci définit le revenu hors-travail permettant à l'individu avec un salaire  $w_2$  d'atteindre le niveau d'utilité espérée  $EU$ . Le lemme de Sheppard nous permet de définir la fonction d'offre de travail compensée ou hicksienne  $h_{2c}$  :

$$-\frac{\partial y}{\partial w_2} = -e_{w_2}(p, w_2, \lambda\tau, V; \bar{h}_1, z) = h_{2c}(p, W_2, \lambda\tau, V; \bar{h}_1, z)$$

Définissons un salaire de réserve à partir duquel l'individu décide de travailler au noir. Celui-ci nous est donné par l'expression suivante :

$$EU[h_2^*(w_2^R), y^* - F + N^*(w_2^R); h_1, z] - EU[0, y; h_1, z] = 0$$

En évaluant la fonction d'offre de travail compensée au salaire de réserve  $w_2^R$ , nous obtenons la fonction d'heure de travail de réserve  $h_2^R$  :

$$h_2^R = -e_{w_2}(p, W_2^R, \lambda\tau, V^1; \bar{h}_1, z) \quad (2.26)$$

où  $V^1$  est le niveau d'utilité atteint lorsque l'individu travaille.

En supposant que la fonction d'offre de travail a la propriété standard de croissance avec les revenus, l'analyse suggère une caractérisation convenable des heures de travail et de la décision de participation au marché noir. En extrapolant la fonction d'offre de travail au noir effective sur l'axe du salaire, nous pouvons définir une fonction «allongée» qui s'écrit :

$$h_2^* = h_2(W_2(1 - p\theta), y - F; \bar{h}_1, z) \quad (2.27)$$

La fonction des heures de réserve est la suivante :

$$h_2^R = f(y, F; \bar{h}_1, z) \quad (2.28)$$

Une restriction importante est imposée à cette dernière équation. En effet, en l'absence de coûts fixes, les heures de réserve sont nulles. Afin de tester l'existence de coûts de participation, il suffirait de vérifier la validité de l'hypothèse nulle selon laquelle  $f(\cdot)$  est égale à zéro pour toutes les observations. Cependant, nous ne pouvons mettre en œuvre cette procédure que si les données nous fournissent une information sur les coûts, ce qui n'est pas notre cas.

La décision de participation au marché noir (notée  $D$ ) peut être exprimée en termes de comparaison entre  $h_2^*$  et  $h_2^R$  à partir des équations (2.27) et (2.28). Si le salaire offert sur le marché noir est inférieur au salaire de réserve, *i.e.*  $w_2 < w_2^R$ , le niveau d'utilité atteint en participant au marché noir est inférieur au niveau d'utilité atteint en n'y participant pas. Sous l'hypothèse que la fonction d'offre de travail au noir est de forme standard, les heures désirées sont supérieures aux heures de réserve, *i.e.*  $h_2^* > h_2^R$ . En revanche, si le salaire offert est supérieur au salaire de réserve ( $w_2 > w_2^R$ ), l'utilité obtenue en travaillant au noir est inférieure à celle qui est obtenue en dehors de ce marché. L'individu choisira alors de ne pas exercer d'activité souterraine et nous constaterons  $h_2^* < h_2^R$ . Par conséquent :

- si  $h_2^* > h_2^R$ , l'individu choisira de travailler au noir ( $D = 1$ )  
et les heures de travail observées seront égales à  $h_2^*$ .
- si  $h_2^* < h_2^R$ , l'individu préférera ne pas travailler au noir ( $D = 0$ )  
et aucune heure de travail ne sera observée.

Cette formulation généralise l'analyse des solutions de coin. En effet, le salaire de réserve correspond habituellement à la pente de la courbe d'indifférence au point où les heures de travail sont nulles et les heures de réserve sont, par définition, égales à zéro. Or, en présence de coûts fixes, les heures de réserve sont strictement supérieures à zéro. Cette formulation présente, en outre, l'avantage de conduire à une spécification du modèle économétrique qui facilite la comparaison avec de précédents travaux réalisés pour le marché officiel (*e.g.* Hausman 1980, Cogan 1981).

L'effet d'une modification de l'ampleur des coûts fixes sur les heures de réserve nous est donné par la dérivée de l'équation (2.26) par rapport à  $F$ , pour un niveau d'utilité  $V^1$  constant :

$$\frac{dh_2^R}{dF} = -e_{22} \frac{dW_2^R}{dF}$$

où  $e_{22}$  désigne la dérivée seconde de  $e$  par rapport à  $W_2^R$ . Le terme  $dW_2^R/dF$  est la variation du salaire de réserve consécutive à un changement du coût fixe. Intuitivement, il paraît assez évident qu'une augmentation du coût à l'entrée se traduise par une

hausse du salaire de réserve. Autrement dit,  $dW_2^R/dF > 0$ . Le terme restant  $-e_{22}$  est l'effet de substitution d'une variation de salaire sur la quantité de travail offerte. Il est nécessairement supérieur à zéro, de sorte que les heures de réserve augmentent avec les coûts fixes à l'entrée sur le marché noir.

En différenciant l'équation (2.26) par rapport à  $y$ , tout en permettant à l'utilité de varier, nous pouvons déterminer l'effet du revenu hors-travail sur les heures de réserve. L'effet est le suivant :

$$\frac{dh_2^R}{dy} = -\frac{\partial e/\partial W_2^R}{\partial W_2^R} \frac{dW_2^R}{dy} - \frac{\partial e/\partial W_2^R}{\partial V^1} \frac{dV^1}{dy} = -\left[ e_{22} \frac{dW_2^R}{dy} + e_{12} \frac{dV^1}{dy} \right]$$

où  $e_{12} \equiv \partial^2 e/\partial W_2 \partial V$ . Sous l'hypothèse que le loisir est un bien normal, une augmentation du revenu hors-travail s'accompagne d'une hausse du salaire de réserve. D'où,  $\frac{dW_2^R}{dy}$  est positif. Dans la mesure où une augmentation du revenu hors-travail accroît l'utilité atteinte lorsque l'individu ne travaille pas,  $\frac{dV^0}{dy}$  est également positif. Mais, si le loisir est normal,  $e_{22}$  et  $e_{12}$  sont de signe opposé. Par conséquent, l'effet du revenu hors-travail sur les heures de réserve est ambigu.

L'analyse théorique nous a permis d'identifier les différents paramètres de décision pour l'exercice d'une activité souterraine. Néanmoins, elle ne nous permet pas de déterminer les effets de chacune des variables sur les heures de travail au noir. L'objet de la section suivante est de procéder aux estimations économétriques de façon à lever certaines des ambiguïtés du modèle théorique.

## 2.4 Estimation du modèle

Selon la nature des données disponibles, deux procédures d'estimation économétriques peuvent être suivies afin d'estimer les fonctions d'offre de travail en présence de coûts fixes à l'entrée. Si l'information sur les coûts est disponible, alors ils peuvent être utilisés pour estimer les paramètres structurels de la fonction d'offre de travail et de la fonction de coût du travail<sup>15</sup>. En revanche, si cette information n'est pas disponible, nous ne pouvons estimer qu'une fonction d'offre de travail de forme réduite. L'estimation reflète alors non seulement les paramètres de la fonction d'utilité, mais également les effets de variations dans le coût du travail.

Dans notre cas, les coûts ne sont pas directement observables dans les données. L'estimation des paramètres du modèle nécessite donc de spécifier les équations appropriées et la structure stochastique sous-jacente.

<sup>15</sup>L'estimation des paramètres structurels est discutée par Cogan (1977).

### 2.4.1 Spécification économétrique

Le modèle théorique, développé à la section précédente, stipule deux équations pour caractériser l'offre de travail sur le marché noir : une fonction «allongée» pour les heures de travail désirées et une fonction pour les heures de travail de réserve. A l'instar de Fortin *et al.* (2000), nous spécifions un modèle de forme quasi-réduite.

Nous postulons la forme fonctionnelle suivante pour l'équation des heures de travail désirées :

$$\begin{aligned} h_2^* &= \gamma_0 + \gamma_1 \log W_2 + \gamma_2 m_1 + \gamma_3 y + \gamma_4 h_1 + \gamma_5 \lambda \tau + \gamma_6 p + \dots + \varepsilon_1 \\ &\equiv \gamma_0 + \gamma_1 \log W_2 + \gamma_2 m_1 + \gamma_3 y + \gamma_4 h_1 + \gamma' X_1 + \varepsilon_1 \end{aligned} \quad (2.29)$$

où  $\log W_2$  est le logarithme du salaire obtenu sur le marché noir.  $m_1$  représente les revenus de l'activité déclarée et  $y$  caractérise le revenu hors-travail.<sup>16</sup> Les heures réalisées dans l'emploi déclaré et la probabilité d'être repéré par les autorités fiscales sont désignées respectivement par  $h_1$  et  $p$ .  $\tau$  est le taux d'imposition marginal effectif sur le salaire déclaré et  $\lambda \tau$  le taux de pénalité sur l'impôt impayé. Enfin,  $X_1$  est un vecteur de variables explicatives correspondant aux paramètres du modèle théorique et aux variables traditionnellement utilisées dans la littérature, telles que l'âge, le sexe et le niveau d'éducation de l'individu. La pertinence théorique et empirique de chacune de ces variables, en l'absence de coûts du travail, est bien connue pour le marché officiel. L'équation des heures de réserve est la suivante :

$$h_2^R = \beta_0 + \beta_1 m_1 + \beta_2 y + \beta_3 h_1 + \beta_4 \lambda \tau + \beta_5 p + \dots + \varepsilon_2 \equiv \beta' X_1 + \varepsilon_2 \quad (2.30)$$

En dépit de l'intuition que l'on peut avoir sur l'effet de chacune des variables, le modèle théorique ne nous apporte que peu d'éclaircissements sur les effets attendus.

Les équations (2.29) et (2.30) impliquent une équation de salaire de réserve de la forme :

$$W_2^R = \delta_0 + \delta_1 m_1 + \delta_2 y + \delta_3 h_1 + \delta_4 \lambda \tau + \delta_5 p + \dots + \mu_3$$

$$\text{où } \delta_0 = \frac{\beta_0 - \gamma_0}{\gamma_1} \quad \text{et} \quad \delta_j = \frac{1}{\gamma_1} (\beta_j - \gamma_{j+1}) \quad \text{avec } j = 1, \dots, 5.$$

La spécification de l'équation d'offre de salaire est la suivante :

$$\log W_2 = \alpha' X_2 + \varepsilon_3 \quad (2.31)$$

<sup>16</sup>La prise en compte du système fiscal nécessite, en effet, de distinguer les revenus de travail et les revenus hors-travail sur le marché officiel. Nous proposons une justification de cette formulation au paragraphe suivant.



où  $X_2$  est un vecteur de variables exogènes. Les équations de salaire sont spécifiées sous forme semi-logarithmique, comme il est d'usage dans la littérature tant théorique qu'empirique.

Afin de rendre compte de la participation ou non au marché officiel, nous spécifions deux équations de forme réduite. La première caractérise l'offre de travail déclaré et la seconde, le revenu d'activité issu de ce marché :

$$h_1 = b'X_3 + \varepsilon_4 \quad (2.32)$$

$$m_1 = c'X_4 + \varepsilon_5 \quad (2.33)$$

où  $X_3$  et  $X_4$  sont des vecteurs de variables exogènes.

Enfin, il convient d'introduire une structure fiscale au modèle afin d'explicitier complètement le comportement d'offre de travail au noir. L'interaction des systèmes d'imposition et de transfert donne lieu à une contrainte budgétaire linéaire par segment, ce qui implique que le taux marginal d'imposition auquel fait face un individu est endogène. La création de cette variable est, par conséquent, une tâche très complexe. Comme nous l'avons mentionné au chapitre précédent, le Québec, tout comme le Canada dans son ensemble ou la France, est doté d'un système d'imposition progressif selon lequel le taux marginal d'imposition augmente avec le revenu imposable. À l'inverse, les transferts sociaux sont mis en place de façon à verser un revenu complémentaire à certains individus. Le taux implicite d'imposition résulte de ces deux mécanismes distincts, chacun étant défini en fonction du revenu et des caractéristiques socio-démographiques des individus. Une augmentation des heures travaillées peut avoir pour effet, d'une part, de réduire le montant des transferts reçus et d'autre part, d'augmenter l'impôt qu'il doit verser. De tels cas mènent parfois à des taux d'imposition marginaux avoisinant et même dépassant 100%. En outre, la juxtaposition de la fiscalité provinciale et fédérale complique davantage encore la détermination de cette variable.

Un modèle comptable a été utilisé par le Ministère des Finances du Québec pour calculer un taux marginal d'imposition propre à chaque individu. Cette variable est définie à partir des revenus de travail, des revenus hors-travail et des caractéristiques démographiques de chaque individu à l'intérieur du ménage.

Le principal problème économétrique, lié à l'introduction de la fiscalité, est la multiplicité des taux de salaire nets auxquels l'individu fait face lorsqu'il détermine son offre de travail. À l'instar de Hausman (1978), nous linéarisons la contrainte budgétaire

de sorte que le revenu hors-travail doit être remplacé par le revenu hors-travail virtuel  $y^v$ . Ce dernier est défini de la manière suivante :

$$y^v = y + \tau W_1 h_1 - T$$

où  $y$  est le revenu hors-travail de l'individu,  $T$  définit le montant de l'impôt net des transferts et  $\tau$  le taux marginal effectif de taxation sur les revenus du travail. Ces deux derniers éléments sont fonction du salaire brut global  $W_1 h_1$ . Le revenu hors-travail, quant à lui, a été calculé à partir des salaires et traitements, des revenus de placement et gains nets en capital, et des diverses prestations. En ce qui concerne les couples, la règle de partage du revenu hors-travail, qui a été appliquée, repose sur le résultat obtenu par Chiappori, Fortin et Lacroix (2001). Elle attribue 70% des ressources à la femme et 30% à l'homme.

Afin d'endogénéiser le revenu hors-travail virtuel, nous spécifions l'équation suivante :

$$y^v = d'X_5 + \varepsilon_6 \quad (2.34)$$

Les équations (2.32) à (2.34) sont estimées par moindres carrés ordinaires. Les estimateurs obtenus servent à générer les valeurs prédites de  $h_1$ ,  $m_1$  et  $y^v$  qui seront remplacées dans l'équation (2.29). De la même manière, la variable  $\log W_2$  est instrumentée à l'aide de l'équation (2.31) et remplacée dans l'équation (2.29) par sa valeur prédite. L'inverse du ratio de Mills, servant à corriger le biais de sélection lors de l'estimation du salaire au noir, sera alors incluse à la spécification.

Le vecteur des termes aléatoires  $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6)$  est supposé normalement distribué, avec un vecteur de zéros pour moyenne et une matrice de covariance constante  $\Omega$ .

Si les variables endogènes  $h^*$ ,  $h^R$  et  $W_2$  étaient observées pour tous les individus, les paramètres des équations (2.29) à (2.31) pourraient être estimés par moindres carrés ordinaires à deux ou trois étapes. Malheureusement, ces variables ne sont pas observées pour tous les individus. D'une part, les heures de réserve  $h^R$  ne sont jamais observées dans les données. D'autre part, les heures désirées  $h^*$  et les salaires  $W_2$  ne sont observés que pour les travailleurs. Par conséquent, nous sommes confrontés à un problème de censure et une approche alternative, telle la méthode du maximum de vraisemblance, doit être mise en œuvre pour estimer les paramètres de notre modèle.

Sous l'hypothèse d'une fonction d'offre de travail au noir croissante avec le taux de salaire, un individu choisira de travailler sur ce marché si  $h_2^* > h_2^R$ . Les heures

observées sont définies de la façon suivante :

$$h_2 = \begin{cases} h_2^* & \text{si } h_2^* - h_2^R > 0 \\ 0 & \text{sinon} \end{cases} \quad (2.35)$$

et l'équation de participation est donnée par :

$$D = \begin{cases} 1 & \text{si } h_2^* - h_2^R > 0 \\ 0 & \text{sinon} \end{cases} \quad (2.36)$$

avec  $h_2^* - h_2^R$ , caractérisant la propension à travailler de l'individu, exprimée ainsi :

$$\begin{aligned} h_2^* - h_2^R &= \gamma_1 \log W_2 + (\gamma' - \beta')X_1 + (\varepsilon_1 - \varepsilon_2) \\ &= \kappa_1 \log W_2 + \kappa'X_1 + v \end{aligned} \quad (2.37)$$

Les équations (2.35) et (2.36) permettent la présence de coûts fixes à l'entrée sur le marché noir et par conséquent, la discontinuité de l'offre de travail au niveau des heures de réserve. Il convient de noter que lorsque  $(h_2^* - h_2^R > 0)$  est remplacé par  $(h_2^* > 0)$  dans l'équation (2.35), ce qui signifie que les heures de réserve sont nulles, le modèle se ramène à un tobit sans coût fixe.

En substituant l'équation d'offre de salaire (2.31) dans la fonction d'heure de travail, la décision de participation s'exprime uniquement en termes de variables exogènes. Elle s'écrit :

$$\gamma_1 \alpha' X_2 + X_1 (\gamma' - \beta') > \varepsilon_2 - (\gamma_1 \varepsilon_3 + \varepsilon_1)$$

ou, de façon plus concise :  $\pi Z > u$ , avec  $\pi = \begin{pmatrix} \gamma_1 \alpha' \\ \gamma' - \beta' \end{pmatrix}$ ,  $Z = (X_2, X_1)$  et  $u \sim N(0, \Sigma)$ .<sup>17</sup> La probabilité de participer au marché noir est donnée par  $P = \Phi\left(\frac{\pi Z}{\sigma_u}\right)$ .

## 2.4.2 Procédure d'estimation

L'analyse développée précédemment souligne l'importance de distinguer la décision de travailler au noir et l'intensité du travail sur ce marché. Cependant, le choix d'entrer n'est pas indépendant de celui du volume de l'activité non déclarée. Les deux décisions doivent donc être déterminées simultanément. Au point de vue statistique cela revient à spécifier un modèle tobit type 2 (Amemyia, 1985). Cette spécification nous permet de prendre en considération la présence de coûts fixes à l'entrée sur le marché noir.

Par ailleurs, notre modèle de participation à l'économie souterraine est conditionné à la participation sur le marché officiel. Cette forme conditionnelle simplifie grandement l'estimation économétrique. Nous évitons ainsi d'estimer la forme structurelle

<sup>17</sup> où  $\Sigma = \sigma_2^2 + \sigma_1^2 + \sigma_3^2 \gamma_1^2 - 2\sigma_{21} - 2\sigma_{23} \gamma_1 + 2\sigma_{13} \gamma_1$ .

complète du modèle, ce qui serait inutile pour identifier l'effet qui nous intéresse. Toutefois, l'estimation d'une équation d'offre de travail conditionnelle soulève un problème d'endogénéité. En effet, comme le soulignent Browning et Meghir (1991), l'hypothèse d'exogénéité est peu probable pour certaines variables explicatives. Dans notre cas, l'analyse empirique tient compte du fait que quatre variables indépendantes sont endogènes. Ces variables sont les heures de travail sur le marché officiel ( $h_1$ ), le revenu de l'activité déclarée ( $m_1$ ), le revenu hors-travail virtuel ( $y^v$ ) et le logarithme du taux de salaire horaire sur le marché noir ( $W_2$ ).

Nous avons retenu une procédure en trois étapes afin d'estimer les paramètres du modèle. En premier lieu, nous estimons l'équation de salaire (2.31) à partir de l'information des individus ayant une prestation positive sur le marché noir. La méthode proposée par Heckman (1979) permet de tenir compte du biais de sélection et d'obtenir des paramètres convergents. En second lieu, les individus ne travaillant pas sur ce marché se voient imputer un salaire basé sur cette équation de salaire. Nous générons également pour chaque individu la valeur prédite des trois autres variables endogènes ( $h_1$ ,  $m_1$  et  $y^v$ ). Puis, nous remplaçons ces variables par leurs valeurs prédites dans l'équation d'offre de travail au noir. Enfin, une troisième étape consiste à maximiser la fonction de vraisemblance, conditionnelle aux valeurs estimées.

Afin de mettre en œuvre la procédure d'estimation du maximum de vraisemblance conditionnelle, il est nécessaire d'obtenir des estimateurs convergents des paramètres du salaire au noir. Or, l'exclusion d'une partie de l'échantillon, à la première étape, est susceptible de soulever un problème de sélection. En effet, les paramètres estimés sont susceptibles d'être biaisés dans la mesure où la sélection de l'échantillon repose sur un critère qui est lui-même le résultat d'un choix individuel (travailler au noir ou non). Posons le problème en termes plus formels afin de mieux saisir la nature du problème.

Les salaires ne sont observés que pour les individus ayant effectivement travaillé ( $h_2 > 0$ ). Si nous estimons une équation de salaire à partir de ce sous-échantillon, la régression obtenue est la suivante :

$$E(W_2|X_2; h_2 > 0) = X_2\beta + E(\varepsilon_{3i}|h_2 > 0) \quad (2.38)$$

Heckman (1979) a montré que l'espérance du terme de droite est généralement non nulle et corrélée avec les variables de  $X_2$ . Donc l'application des moindres carrés ordinaires à l'équation (2.38) donnera des estimateurs biaisés. Il a alors proposé d'estimer cette équation en deux étapes. A la première, nous estimons l'espérance conditionnelle  $E(\varepsilon_{3i}|h_2 > 0)$  à l'aide d'une régression de type probit. Nous appliquons l'analyse à la

décision de participation au marché noir afin de générer l'inverse du ratio de Mills<sup>18</sup> pour chaque observation. Cette étape nous fournit également une estimation de l'équation de forme réduite de participation. A la seconde étape, nous régressons (2.38) par les moindres carrés ordinaires en traitant les estimateurs obtenus à la première étape comme une variable manquante. Les estimateurs des paramètres ainsi obtenus sont convergents.

La procédure en trois étapes permet de corriger le biais d'endogénéité et de sélection de l'estimateur d'une équation avec variables explicatives endogènes. Cependant, les écart-types obtenus sont asymptotiquement incorrects. La matrice de variance-covariance des estimateurs est biaisée vers zéro. Récemment Fernandez *et al.* (2001), ont démontré la convergence en distribution d'un estimateur en trois étapes pour un modèle structurel d'offre de travail. Ils définissent un algorithme efficace pour calculer la matrice de variance-covariance. Mais, le recours à une telle technique est particulièrement complexe, d'autant que notre modèle comporte trois variables endogènes supplémentaires dans l'équation d'offre de travail ( $h_1, m_1$  et  $y^v$ ). Ces trois variables compliquent davantage encore la dérivation de la distribution asymptotique<sup>19</sup>.

L'objectif de ce chapitre n'est pas de déterminer l'impact de chacune des variables du modèle sur l'offre de travail au noir. Cet aspect constituera l'enjeu du prochain chapitre et sera abordé dans un cadre structurel plus complet. Pour l'heure, il s'agit de tester la présence de coûts fixes à l'entrée sur le marché noir. La procédure d'estimation mise en œuvre nous fournit une information suffisante pour appréhender l'impact de ces coûts sur l'offre de travail non déclaré. La comparaison d'estimations, qui successivement tiennent compte de ces coûts ou les négligent, nous apportera un élément de réponse. L'existence de ces coûts n'est pas facile à tester et la mesure de leur ampleur n'est pas, non plus, une tâche aisée. Nous suivons alors la méthode proposée par Scott et Garen (1994). Celle-ci consiste à comparer la somme des résidus des carrées des régressions contrainte et non contrainte par le test de Chow (après correction de l'hétéroscédasticité).

### 2.4.3 Identification des paramètres du modèle

Avant de présenter les résultats de nos estimations, il importe de discuter de l'identification des paramètres du modèle. Comme nous l'avons mentionné, nous avons recours à une procédure à variables instrumentales à trois étapes pour résoudre les problèmes

<sup>18</sup>Le ratio de Mills est défini par la fonction de distribution cumulative à  $\tau$  divisée par la fonction de densité à  $\tau$  :  $-\frac{\Phi(\tau)}{\phi(\tau)}$ .

<sup>19</sup>Se reporter à l'annexe de l'article de Fernandez *et al.* (2001) pour plus de détails sur la distribution asymptotique, dans le cas proposé par ces derniers.

d'endogénéité et de sélection. Dans la mesure où aucune variable endogène n'apparaît dans l'équation des heures de travail officiel  $h_1^*$ , dans celle du revenu de l'activité déclarée  $m_1$  et du revenu hors-travail  $y^v$ , elles sont automatiquement identifiées. Cependant, certaines restrictions sont nécessaires afin d'identifier les paramètres de nos équations de participation (2.36) et d'offre de travail au noir (2.29). De façon générale, pour ce type de procédure, le nombre de variables servant d'instruments à la première étape doit être égal au nombre de variables endogènes pour que le modèle soit parfaitement identifiable. Dans notre cas, la matrice de variables exogènes  $X_2$  doit contenir au moins trois variables n'apparaissant pas dans la matrice  $X_1$ <sup>20</sup>. Comme le souligne l'article de Fortin *et al.* (2000), le choix des variables doit être basé sur des critères économiques et statistiques. Suivant les travaux de ces auteurs, nous excluons de notre équation d'offre de travail, la variable dichotomique prenant la valeur 1 lorsque l'individu affirme qu'il aurait souhaité travailler un plus grand nombre d'heures sur le marché officiel.

En outre, nous ajoutons des variables supplémentaires à notre matrice d'instruments dans le but d'accroître le pouvoir explicatif de nos équations de première étape. Ainsi, la liste de nos instruments inclut également une variable d'interaction entre le sexe ( $femme = 1$ ) et la présence d'enfants de moins de 6 ans, le statut marital au sein du ménage<sup>21</sup> ainsi que des variables régionales muettes (*Montréal et Bas-du-Fleuve*<sup>22</sup>). Afin d'examiner le rendement de l'éducation en termes de salaire, nous incluons les variables de scolarité et scolarité au carré. Enfin, les variables d'âge au carré et au cube nous permettront de saisir la nature de la relation, concave ou non, entre l'âge et les rémunérations salariales. Le choix de ces variables s'inspire de la théorie du capital humain.

Parmi les variables explicatives du volume horaire de l'activité souterraine,  $h_2^*$ , nous introduisons au-delà des caractéristiques individuelles objectives traditionnelles, les croyances subjectives relatives à l'économie souterraine. Nous prenons ainsi en considération le sentiment des individus sur l'exercice d'une activité dissimulée et le jugement que, selon eux, la société porte sur ce phénomène. En effet, l'enquête rapporte l'évaluation par les individus du pourcentage de la population active participant à l'économie souterraine parmi les membres de l'entourage. Des questions d'ordre moral ont également été posées. D'une part, l'exercice d'un emploi non officiel est jugé, selon

<sup>20</sup> Notre modèle est un cas particulier du modèle tobit à équations simultanées développé par Amemyia (1974). Nos conditions d'identification correspondent à celles qui sont discutées dans son papier, notamment l'hypothèse 3.3.

<sup>21</sup> Les catégories sont les suivantes : personne vivant seule, famille monoparentale, personne vivant avec conjoint et enfant à la maison. La référence est la catégorie "autres" qui regroupe les individus vivant chez leurs parents.

<sup>22</sup> La référence étant la ville de Québec.

une échelle de valeur, très moral à très immoral. D'autre part, l'approbation ou la désapprobation des proches, à l'annonce de l'exercice d'une activité souterraine, est appréhendée par une variable dichotomique.

Au cours de l'enquête, il a été demandé aux personnes interrogées d'évaluer la probabilité de détection d'une activité non déclarée et le taux de pénalité encouru. Cette information est utilisée pour estimer la perception du risque du travail au noir et, par conséquent, l'impact du système fiscal et de répression de la fraude sur l'exercice d'une activité souterraine. Nous aurions pu exploiter les publications officielles des services fiscaux, et les divers textes de loi relatifs au travail dissimulé, pour appréhender la probabilité de détection par secteur d'activité et le montant réel des pénalités. Mais, les décisions individuelles sont certainement davantage fondées sur une évaluation subjective que sur une connaissance précise de ces éléments. Il est donc certainement plus judicieux de procéder de cette manière.

## 2.5 Résultats économétriques

Nous présentons les résultats de nos estimations étape par étape. Nous analysons successivement la décision de participer à l'économie souterraine, les déterminants du salaire offert sur le marché noir et le choix des heures de travail. Les estimations de la probabilité d'entrer sur le marché noir et du temps à y consacrer confirment l'existence de critères de choix significativement différents. D'une part, l'influence des variables explicatives n'est pas la même selon qu'il s'agit de participer au marché noir ou de déterminer l'intensité de l'activité souterraine. D'autre part, certaines de ces variables n'affectent que l'une des deux décisions. Enfin, l'estimation de l'équation des heures de réserve révèle la présence de coûts fixes importants à l'entrée sur le marché noir. L'ampleur de ces coûts reste inférieure à celle que l'on retrouve dans les travaux empiriques sur le marché officiel. Cependant, ils conditionnent très fortement l'offre de travail au noir.

### 2.5.1 La probabilité de travailler au noir

La décision de travailler au noir est supposée dépendre à la fois des attributs personnels et de considérations purement financières. L'impact des croyances personnelles en matière de répression de la fraude et en matière de jugement par la société est également testé. Enfin, le choix d'une activité souterraine est examiné au regard de la perception du fardeau fiscal.

Le tableau de contingence nous permet de comparer les valeurs prédites et les valeurs réelles. Elle nous donne une indication de la qualité du modèle.

Tableau 2-1 :  
Tableau de contingence

	Prédiction		Réalisation
	$h_2 = 0$	$h_2 > 0$	
$h_2 = 0$	4755	1	4756
$h_2 > 0$	69	163	232
Total	4824	164	4988

Le modèle prédit correctement 98 % des observations.

Le tableau 2.2 présente les effets marginaux de chaque variable explicative sur la décision de travailler au noir.<sup>23</sup>

Tableau 2-2 :  
Décision d'entrer sur le marché noir

Variables	Coefficient	Ecart-type	
Constante	-0.4811	0.2229	*
Age	-0.0136	0.0032	***
Femme	0.0645	0.0730	
Femme x (enfants <6 ans)	-0.1917	0.1011	*
Primaire	-0.2961	0.1600	*
Secondaire	-0.2224	0.0934	**
Collège	0.0624	0.0888	
Université	référence	-	
Conjoint	-0.3086	0.0814	***
Monoparent	-0.0971	0.1600	
Célibataire	référence	-	
Probabilité de détection	-0.7135	0.1481	***
Fardeau fiscal	-0.4125	0.0774	***
Prestations sociales	0.1826	0.0862	**
Rationnement sur le marché officiel	0.1288	0.0758	*
Pénalités	0.0080	0.0630	
Proportion dans l'entourage	0.1037	0.0167	***
Réaction de l'entourage	0.3010	0.0867	***
Immoralité	-0.1970	0.0405	***
Pseudo-R <sup>2</sup>	0.1901		
Nombre d'observations	4988		

\*\*\* statistiquement significatif à 1%,

\*\* statistiquement significatif à 5%,

\* statistiquement significatif à 10%.

Au regard de ces résultats, la décision de travailler au noir est très fortement influencée par des considérations d'ordre moral. Le jugement que porte la société sur

<sup>23</sup> Dans les modèles probit, les coefficients ne reflètent les effets marginaux des variables explicatives. Ces derniers ont donc été calculés et rapportés dans le tableau ci-dessous.



l'économie souterraine, et plus encore la réaction de son entourage face au travail au noir sont déterminants. Ainsi, lorsqu'ils sont persuadés de l'approbation de leur entourage (*réaction entourage*), les individus exercent plus volontiers une activité non officielle. De même, plus ils estiment importante la participation de leurs proches à l'économie souterraine (*proportion entourage*), plus ils participent. A l'inverse, le sentiment de réaliser une tâche immorale (*immoralité*) les en dissuade. Les résultats confirment également que les travailleurs au noir sont majoritairement jeunes. Le coefficient associé à la variable *âge* est significativement négatif attestant que le travail au noir est essentiellement concentré sur les classes d'âge les plus faibles. De même, les individus n'ayant obtenu qu'un diplôme du primaire ou du secondaire tendent à moins participer que les individus ayant atteint le niveau du collège.

Par ailleurs, le risque associé à une activité souterraine (*probabilité de détection*) affecte la décision de travailler au noir très significativement. En revanche, le montant de l'amende encourue (*pénalités*) ne semble pas avoir d'effet sur la décision de participer à l'économie souterraine. Le fait d'avoir des enfants en bas âge agit négativement sur la participation des femmes au marché noir, comme en témoigne la variable d'interaction «*femme × enfant < 6 ans*». Du point de vue de la situation matrimoniale, il s'avère qu'être en couple diminue également la probabilité d'accepter un emploi non déclaré. Ce résultat peut s'expliquer de deux façons alternatives. D'une part, la vie de famille peut dissuader les individus d'encourir le risque d'une activité frauduleuse. D'autre part, ce statut peut permettre à l'individu de bénéficier des ressources du conjoint et, par conséquent, rendre la recherche de revenus supplémentaires moins indispensable.

De manière surprenante, le sentiment d'être trop imposé réduit la probabilité de travailler au noir (*imposup*). L'explication d'un tel résultat peut être trouvée dans le fait que les personnes se plaignant de payer trop d'impôt sont celles qui payent le plus d'impôt. Et, puisqu'il existe de nombreuses possibilités d'évasion fiscale, ces personnes n'ont probablement pas besoin de recourir à une activité non déclarée pour réduire leur charge fiscale. Elles ne participent donc probablement pas au marché noir. Au contraire, la perception d'allocation-chômage (*prestations*) incite les individus à travailler au noir afin d'obtenir un revenu additionnel tout en conservant leurs indemnités.

Enfin, le choix d'une activité non déclarée semble également dépendre de contraintes institutionnelles sur les heures officielles de travail. En effet, les individus qui se déclarent contraints de travailler moins d'heures qu'ils ne le désireraient sur le marché officiel (*rationnement*) ont davantage d'incitations à travailler au noir.

## 2.5.2 L'équation de salaire au noir

Les régressions des équations de salaire au noir semi-logarithmiques sont présentés dans le *tableau 2.3*. Pour les deux premières colonnes, les estimations du probit ont été utilisées pour évaluer l'inverse du ratio de Mills et tenir compte d'un biais de sélection potentiel dans l'équation de salaire. La matrice de variance-covariance a été calculée selon la procédure de White (1980) pour tenir compte de l'hétéroscédasticité introduite par l'inverse du ratio de Mills.

Tableau 2-3 :  
Salaire au noir

Variables	Estimateurs par		Estimateurs par	
	Moindres Carrés Ordinaires <sup>a</sup>		Moindres Carrés Ordinaires <sup>b</sup>	
	Coefficient	Ecart-type	Coefficient	Ecart-type
Constante	-3.7304	1.2020	-0.7058	1.0582
Age	0.2174	0.0836 ***	0.1801	0.0924 *
Age au carré	-5.0797	2.0879 **	-4.1095	2.2856 *
Age au cube	3.5936	1.5556 **	2.9411	1.6733 *
Femme	0.0076	0.1386	0.0117	0.1286
Montréal	-0.0143	0.1193	-0.0048	0.1400
Bas-du-Fleuve	-0.2819	0.1631 *	-0.2880	0.1715 *
Années de scolarité	3.5542	1.5903 **	3.3629	1.4665 **
Années de scolarité au carré	-3.5790	1.9572 **	-3.7084	1.8767 *
Ratio de Mills	4.2681	1.0276 ***	-	-
R <sup>2</sup>	0.1074		0.0981	
Salaire moyen prédit <sup>c</sup>				
Travailleurs		2.21		2.24
Chômeurs		2.08		2.09
Travailleurs au noir		2.20		2.20
Non-participants		2.17		2.19
Total		2.17		2.20

<sup>a</sup> Les écarts-type sont corrigés pour toute forme d'hétéroscédasticité.

<sup>b</sup> Sans correction de sélection.

<sup>c</sup> Etant donné la forme de l'équation, les salaires sont log normalement distribués et les salaires moyens sont calculés par :

$$\bar{w} = \exp^{\alpha} X_2 + (1/2)\sigma_2^2$$

\*\*\* Statistiquement significatif à 1% \*\* Statistiquement significatif à 5%, \* Statistiquement significatif à 10%.

A l'instar de Cogan (1981), nous rapportons les estimations par moindres carrés ordinaires sans correction de sélection pour appréhender l'impact du biais de sélection sur les paramètres estimés. La comparaison des résultats issus des estimations, avec et sans correction de ce biais, apporte un appui à l'existence d'un biais de sélection important. Le coefficient de l'inverse du ratio de Mills est significativement très élevé. Son signe est, de surcroît, positif, indiquant que les travailleurs au noir reçoivent

en moyenne des offres de salaires plus importantes que ceux qui ne participent pas. Nous constatons donc la présence d'une auto-sélection sur le marché noir.<sup>24</sup> Le biais de sélection s'avère particulièrement conséquent dans l'estimation du terme constant qui reflète la différence de prédiction entre les salaires moyens conditionnels et non conditionnels à la moyenne de l'échantillon. La différence dans l'estimation de l'effet de l'âge, bien que faible en valeur absolue, est relativement grande du point de vue du coefficient (12%). L'estimation sans correction du biais de sélection sous-estime, en outre, la décroissance du rendement de l'âge sur les rémunérations. En revanche, aucune conclusion ne peut être déduite pour les variables *femme* et *Montréal*, dans la mesure où nous sommes très loin du seuil de significativité. Enfin, le biais de sélection semble n'avoir qu'un impact modéré sur le niveau d'éducation, ce qui est encourageant.

La plupart des variables ont le signe attendu et sont statistiquement significatives. Les résultats sont, en outre, conformes à ce qui est observé dans la littérature standard. Les rémunérations augmentent avec l'âge, mais à un taux décroissant, reflétant la concavité de la fonction de salaire sur le marché noir. De même, il existe une relation positive entre le niveau d'éducation et le salaire. Mais, le rendement du diplôme est décroissant. Le coefficient associé au niveau de scolarité doit être interprété comme l'effet d'une année supplémentaire dans le système éducatif relativement à une année additionnelle sur le marché noir.

Un résultat intéressant est que la variable *femme* n'est pas significative. Alors qu'il est communément admis que les femmes souffrent de forte discrimination salariale sur le marché officiel, cette dernière ne semble pas présente sur le marché noir. Enfin, les variables régionales font apparaître des différences de salaire entre les zones rurales et urbaines. Ainsi, la région du Bas-du-Fleuve offre des rémunérations inférieures aux travailleurs au noir comparativement à Québec. La ville de Montréal semble également offrir des salaires plus faibles qu'à Québec, mais la variable n'est pas statistiquement significative.

Enfin, il convient de noter que nous n'avons pas introduit de variable d'expérience dans nos estimations, comme il est coutume de le faire. L'hypothèse que l'expérience est exogène aux heures de travail a été initialement rejetée par Heckman (1979), et par d'autres ensuite (*e.g.* Mroz 1987), notamment pour les femmes. Il est, en effet, concevable que l'expérience soit fortement corrélée avec certaines caractéristiques non observables qui affectent la situation présente sur le marché. Par exemple, les individus

<sup>24</sup>Une procédure identique a été déployée pour les revenus issus du marché officiel. Les résultats obtenus rejettent l'hypothèse d'auto-sélection sur ce marché. Nous n'avons donc pas à corriger d'auto-sélection pour nos équations  $m_1$  et  $h_1$ .

qui ont un fort attachement au marché du travail accumulent beaucoup d'expérience et sont susceptibles d'avoir un salaire horaire élevé. Ce salaire élevé peut, en retour, être responsable de leur présence actuelle sur le marché.

Afin d'éviter de biaiser inutilement les estimations des paramètres de l'équation de salaire, nous avons préféré ne pas inclure cette variable dans nos estimations. Deux raisons ont, en outre, motivé ce choix. En premier lieu, la variance de l'expérience sur ce marché est tout aussi grande pour les hommes que pour les femmes. Dans la mesure où cette dispersion est due à des différences non observables dans le goût pour ce type de travail, les problèmes d'endogénéité évoqués ci-dessus sont susceptibles de se produire. En second lieu, la variable d'expérience sur le marché noir est probablement sujette à des erreurs de mesure. Ainsi, comme les heures travaillées sur ce marché sont en moyenne assez faibles, et souvent concentrées sur quelques semaines, la signification exacte du terme «années» d'expérience peut porter à confusion. Certains individus peuvent considérer une présence très brève sur le marché au cours d'une année comme équivalent à une année d'expérience, alors que d'autres pourront plutôt être portés à cumuler les heures jusqu'à ce qu'elles correspondent aux heures annuelles «normales» sur le marché officiel avant de les considérer comme équivalentes à une année d'expérience.<sup>25</sup>

### 2.5.3 Le volume horaire de l'activité souterraine

La variable dépendante, retenue pour estimer l'offre de travail au noir, est le nombre d'heures annuelles. Ce choix nécessite quelques commentaires. En effet, les heures journalières semblent *a priori* plus appropriées pour capter les effets des coûts fixes du travail, qui sont habituellement quotidiens ou convertibles sur une base quotidienne. Ils concernent les déplacements domicile-travail, les frais de garde des enfants et les dépenses d'équipements liés à l'activité elle-même. Malheureusement, les bases de données existantes ne fournissent pas d'information sur cette dimension de l'offre de travail. Le choix des heures de travail annuelles se justifie alors par le fait qu'elles constituent la meilleure alternative possible. Et, parce qu'elles donnent une mesure compréhensible de l'offre de travail et facilitent la comparaison avec les études précédentes, elles sont également les plus communément utilisées.

Dans le cas du travail au noir, un certain nombre de raisons supplémentaires viennent étayer ce choix. Les coûts engendrés par l'exercice d'une activité souterraine sont vraisemblablement basés sur un horizon plus long que la seule journée. D'une part,

<sup>25</sup>L'examen de l'endogénéité de la variable d'expérience est développé en *annexe 2A*.

le travail au noir est, par définition, dissimulé. Le coût psychologique, que doivent subir les travailleurs du fait du jugement négatif de la société, n'est alors pas supporté quotidiennement. D'autre part, les dépenses occasionnées par l'achat de matériels, ne pouvant être assimilées à des frais professionnels, sont généralement amorties sur une durée à moyen ou long terme. De même, les sommes nécessaires à la dissimulation de l'activité ne peuvent être engagées au quotidien. Enfin, l'emploi lui-même n'est pas effectué sur une base journalière. Il est étalonné sur l'ensemble de l'année, à raison de quelques heures par semaine ou de quelques semaines par an. Par conséquent, une variable annuelle semble plus adéquate pour appréhender les coûts monétaires et psychologiques, associés à l'emploi non déclaré. C'est pourquoi, l'offre de travail au noir est estimée à partir des heures travaillées durant l'année.

L'estimation des paramètres des heures de travail et des heures de réserve pour le marché noir sont présentées dans le *tableau 2-4*.

*Tableau 2-4 :*  
**Heures de travail et heures de réserve au noir**

<i>Variables</i>	<i>Heures de travail au noir</i>		<i>Heures de réserve au noir</i>	
	<i>Coefficient</i>	<i>Ecart-type</i>	<i>Coefficient</i>	<i>Ecart-type</i>
Constante	2.7441	0.7401	2.9642	0.7225
Age	-2.1680	0.6833	-2.2063	0.6828
Age au carré	4.1334	1.3522	4.1582	1.3522
Age au cube	-2.3135	0.8136	-2.3332	0.8136
Femme	2.2240	0.9371	2.0818	0.9314
Nbre d'enfants <6 ans	-4.7431	4.5216	-6.1362	4.4060
Années de scolarité	-6.7741	2.8964	-5.7449	2.7974
Revenu disponible	-0.0010	0.0004	-0.0009	0.0004
Heures officielles	-0.0559	0.0311	-0.0557	0.0324
Log(salaire) au noir	1.0627	0.7762	-	-
Pénalités	-0.6970	0.4110	-0.6359	0.3800
Probabilité de détection	-2.2443	1.1135	-3.1111	0.9160
Proportion dans l'entourage	4.4210	1.4284	4.5330	1.4262
Réaction de l'entourage	1.1083	0.5207	1.1092	5.2076
Immoralité	-1.3555	0.2839	-1.3642	2.8394
<i>Heures moyennes</i>				
Travailleurs officiels		544.48		610.85
Chômeurs		650.17		567.98
Travailleurs au noir		581.44		567.61
Non participants		544.33		689.84
Total échantillon		579.71		608.61

\*\*\* Statistiquement significatif à 1%, \*\* Statistiquement significatif à 5%, \* Statistiquement significatif à 10%.

Les résultats sont conformes, au moins qualitativement, à ce qui est traditionnellement obtenu dans les travaux empiriques sur l'offre de travail au noir. L'introduction

de coûts fixes n'altère donc pas la nature qualitative des paramètres estimés de l'offre de travail. En revanche, les heures de réserve estimées sont plus élevées pour les individus ne participant pas à l'économie souterraine, indiquant que les coûts fixes à l'entrée sur le marché noir sont déterminants dans les décisions d'offre de travail. Les individus les plus actifs sur le marché noir sont en majorité de jeunes gens dont le niveau d'éducation est peu élevé (Fortin *et al.*, 1996). La relation entre l'âge et l'intensité au travail sur le marché noir est négative, contrairement à ce qui est observé sur le marché officiel. L'offre de travail au noir diminue à un taux croissant avec l'âge. L'obtention d'un diplôme, ayant *a priori* pour effet d'accroître les chances d'obtenir un emploi officiel, réduit logiquement le temps consacré à une activité souterraine. Dans le même ordre d'idée, une augmentation du revenu disponible diminue la participation au marché noir, tandis qu'un accroissement des salaires offerts au noir incite l'individu à développer son activité souterraine. Nous avons constaté en première étape que les travailleurs sur le marché officiel avaient une probabilité plus grande de travailler au noir. De nombreuses professions facilitent, en effet, l'exercice d'activités souterraines en rapport avec l'emploi déclaré, ou permet de réaliser des heures supplémentaires non déclarées. Ce résultat confirme l'existence d'incitations à cumuler les emplois déclarés et non déclarés. En revanche, l'exercice d'un emploi officiel a pour effet de diminuer le temps disponible et, par conséquent, l'activité souterraine, de sorte que le coefficient associé aux heures officielles est statistiquement négatif.

Par ailleurs, les résultats indiquent clairement que l'intensité du travail est fortement conditionné par l'entourage et par l'image plus ou moins négative associée à la fraude. Ainsi, la perception d'un emploi non déclaré comme étant très immoral (variable *immoralité*) restreint l'activité souterraine de l'individu. En revanche, l'approbation d'une activité non déclarée de la part de l'entourage (variable *réaction*) accroît le volume horaire de l'emploi dissimulé. Alors qu'un certain opprobre social peut être associé aux emplois souterrains, notamment pour les non participants, une attitude moins négative à l'égard de ces activités peut motiver leur choix. De même, la conviction que le travail au noir est un phénomène très répandu, en particulier parmi ses proches, incite à travailler davantage sur ce marché. La variable «*proportion de l'entourage*» caractérise l'impact du voisinage sur l'intensité de l'activité souterraine. En l'occurrence, il semble que la détermination des heures de travail au noir soit influencée par le comportement des proches de l'individu. Cette variable est, en effet, très significative et met en évidence l'interaction entre les choix individuels et le comportement des proches. Elle révèle le poids des normes sociales dans la décision de fraude et apporte un éclaircissement sur l'*effet d'épidémie* propre au développement de l'économie souterraine.

Il est, en effet, souvent avancé que le travail au noir se développe au sein d'un réseau de relations très denses.<sup>26</sup> Par ailleurs, les membres d'un même groupe social tendent à se comporter de façon similaire (Manski, 2000). Le groupe auquel l'individu appartient représente alors non seulement un bassin d'emplois potentiels non négligeable, mais génère également davantage de tolérance à l'égard des activités souterraines. Autrement dit, un milieu constitué de nombreux travailleurs au noir a tendance à inciter ses membres à exercer le même type d'activités et, par un effet de réseau, conduit à la fois à l'accroissement des opportunités d'emploi sur ce marché et à la levée d'éventuelles réticences à l'égard des activités frauduleuses. Ce résultat confirme l'hypothèse d'interdépendance des comportements individuels pour laquelle les préférences individuelles sont affectées par le comportement des autres individus (Aronsson *et al.*, 1999). L'individu interagit avec son environnement lors de ses prises de décision et les interactions sociales parmi les membres d'un groupe sont de nature à conditionner le choix de l'activité. Les préférences individuelles en termes d'heures de travail au noir sont donc influencées par le comportement et l'approbation des membres du groupe auquel l'individu appartient. Ces résultats justifient que l'on poursuive l'analyse du rôle des normes sociales au *chapitre 4*.

Par ailleurs, les résultats confirment l'impact du système fiscal sur l'offre de travail au noir. La variable «*probabilité de détection*» limite l'ampleur de l'activité souterraine. Le coefficient associé à la variable «*pénalités*» indique également une relation négative entre l'offre de travail au noir et le montant de l'amende à payer en cas de détection. La forme conditionnelle du modèle implique que le taux marginal d'imposition  $\tau$  affecte l'offre de travail au noir uniquement par l'intermédiaire du taux de pénalité. Lacroix *et al.* (1992) ont démontré qu'une augmentation du taux marginal de taxation menait à une diminution de l'offre de travail sur le marché noir. Cependant, puisque nous conditionnons sur l'offre de travail officiel, la substitution entre les deux marchés est nulle et une augmentation du taux marginal d'imposition a pour effet d'augmenter le taux de pénalité ( $\lambda\tau$ ) dans le cas d'un contrôle fiscal. La variable définie comme étant le multiple de l'impôt non déclaré que doit rembourser l'agent ( $\lambda$ ) dans le cas d'un contrôle fiscal varie entre 0 et 25. Puisque l'amende n'est évidemment pas plus élevée que deux fois l'impôt impayé, nous avons appliqué la transformation suivante à cette variable : nous utilisons la fonction cumulative normale de la variable centrée-réduite des valeurs comprises entre 1 et 25. Ce faisant, nous obtenons comme taux de pénalité des valeurs comprises entre 0 et 2 fois l'impôt impayé.

---

<sup>26</sup> Spiro (1993) montre que les individus, une fois sur le marché noir, développent leur propre réseau social et nouent d'importantes relations personnelles. Cela explique qu'ils éprouvent certaines difficultés à retourner sur le marché officiel, même à plus long terme.

Les paramètres estimés peuvent être utilisés pour mesurer les coûts de participation. Pour cela, il est nécessaire de déterminer la forme de la fonction d'offre de travail compensée à partir de la forme supposée de la fonction d'offre de travail non compensée. Cette dernière (équation (2.29)) peut être écrite succinctement de la façon suivante :

$$h_2 = \gamma_1 \log W_2 + \gamma' X_1 + \gamma_2 y + \varepsilon_1 \quad (2.39)$$

où  $X_1$  est le vecteur des autres variables incluses dans le terme de droite. En utilisant la décomposition de Slutsky, nous pouvons écrire la pente d'une fonction d'offre de travail compensée associée à (2.39) :

$$\left( \frac{\partial h_2}{\partial w_2} \right)^c = \frac{\gamma_1}{w_2} - h_2 \gamma_2 \quad (2.40)$$

La forme générale de la fonction compensée peut facilement être obtenue de l'équation (2.40). Celle-ci est une équation différentielle linéaire de premier ordre dont la solution générale sur un intervalle  $I$  est donnée par :

$$h_2^c(w_2) = e^{-\gamma_2 w_2} \int_{w_2^*}^{w_2} e^{\gamma_2 t} \frac{\gamma_1}{t} dt + c e^{-\gamma_2 (w_2 - w_2^*)}$$

où  $c$  est une constante arbitraire et  $w_2^*$  est l'intersection sur l'axe des ordonnées de la fonction d'offre de travail. Cette dernière est extrapolée vers le bas et correspond à zéro heure de travail. Cette solution particulière au point du salaire de réserve et des heures de réserve peut être déterminée numériquement. Les coûts fixes annuels sont calculés séparément pour les participants et les non participants au marché noir. Ces estimations sont présentées dans le *tableau 2.5*. A des fins de comparaisons, nous avons rapporté les estimations des coûts fixes sur le marché officiel réalisées par Cogan (1981). Nous avons également converti les montants en euros pour plus de lisibilité.

Tableau 2-5 :

**Coûts fixes annuels**

Participants	226.09 \$can.	142.38 euros	947.29 euros <sup>a</sup>
Non-participants	1 130.07 \$can.	711.95 euros	1 022.62 euros <sup>a</sup>
Total	1 088.03 \$can.	685.46 euros	949.70 euros <sup>a</sup>

<sup>a</sup> Estimations de Cogan (1981), pour le marché officiel.

Les coûts fixes sont, à première vue, très faibles. Leur ordre de grandeur est bien moindre que celui des coûts observés pour le marché officiel par Cogan (1981), *i.e.* 30% de moins. Néanmoins, lorsque nous les rapportons aux gains annuels estimés sur le marché noir, ils deviennent considérables. En effet, ils représentent près du tiers des revenus potentiels des non participants (1 130,07/3443 \$can.), soit environ 1.7 euro



par heure travaillée (711,95 euros/424 heures annuelles moyennes travaillées au noir). Ce montant n'est, en revanche, que de 0.33 euro pour les travailleurs au noir, soit environ 6% des revenus annuels obtenus sur ce marché. Comme nous pouvions nous y attendre, les non participants ne travaillent pas au noir parce qu'ils font face à des coûts excessivement plus élevés que les participants. La présence de coûts fixes à l'entrée sur le marché noir contribue ainsi à expliquer le taux de participation relativement limité à l'économie souterraine. Malheureusement, nous n'avons pas été en mesure d'évaluer l'effet d'un changement dans la situation personnelle ou familiale, tel que l'obtention d'un diplôme ou l'arrivée d'un nouvel enfant.

Le tableau 2-6 reporte les paramètres estimés correspondant à différentes spécifications du modèle. Il montre également la sensibilité des paramètres à la spécification de l'amende à payer en cas de détection. Quelle que soit la spécification retenue, lorsque la variable ( $\lambda\tau$  = pénalités sur l'impôt impayé) est remplacée uniquement par la variable  $\tau$ , nous obtenons une relation plus significative, ce qui suggère que la variable  $\lambda$  n'est pas correctement mesurée et par conséquent ajoute du bruit autour de la variable du taux marginal d'imposition.

Tableau 2-6 :  
Estimations Tobit

Variables	Tobit (sans coûts fixes)		Décision d'entrer (iii)	Tobit type 2 (avec coûts fixes)		
	Volume horaire (i)	Volume horaire (ii)		Volume horaire (iv)	Décision d'entrer (v)	Volume horaire (vi)
Constante	2.1369 (1.561)	2.1811 (1.557)	2.0500 * (1.230)	-1.8954 (1.521)	2.1070 (1.220)	-1.8634 (1.396)
Age	-1.0543 ** (0.537)	-1.0431 * (0.535)	-0.7900 ** (0.336)	1.0132 (0.639)	-0.0790 ** (0.330)	1.0135 * (0.577)
Age au carré	1.4054 * (0.730)	1.3833 * (0.718)	1.0700 ** (0.513)	-1.3912 * (0.821)	1.0700 ** (0.510)	-1.4010 * (0.744)
Femme	3.8945 * (2.092)	3.7736 * (2.086)	0.2570 * (0.150)	-2.7469 (1.910)	0.2500 * (0.150)	-2.9406 (1.932)
Nb enfant <6ans	-3.0510 (0.523)	-3.5950 (0.499)	-0.2300 (0.750)	-1.8467 ** (0.851)	-0.0270 (0.870)	-1.9722 ** (0.927)
Années de scolarité	-3.1634 * (1.880)	-3.0347 (1.8753)	-2.4100 * (1.430)	2.3351 ** (6.527)	-2.3508 (1.440)	2.6356 (2.948)
Années de scolarité au carré	6.1885 (4.989)	6.0709 (4.985)	6.6100 (4.410)	-1.3026 ** (0.652)	6.6100 (4.141)	-1.3422 ** (0.644)
Revenu déclaré	-0.0040 ** (0.001)	-0.0040 ** (0.001)	-3.3300 ** (1.480)	0.0020 (0.001)	-3.3261 ** (1.497)	0.0020 (0.001)
Heures officielles	-2.6100 * (1.400)	-2.5800 * (1.440)	-0.002 ** (0.001)	1.7500 ** (1.310)	-0.0020 ** (0.001)	1.8117 (1.310)
Revenu virtuel	0.0300 ** (0.012)	0.0320 ** (0.012)	0.0010 (0.600)	0.0170 *** (0.008)	0.0001 (0.007)	0.0191 *** (0.009)
Log(salaire) au noir	1.8592 (1.480)	1.9341 (1.485)	0.2000 * (0.110)	-2.5836 ** (1.085)	0.2111 * (0.112)	-2.6264 ** (1.073)
Tx marg. d'impôt ( $\tau$ )	-	-3.1700 ** (1.680)	-	-	-0.0023 * (0.001)	-2.4432 ** (1.209)
Pénalités ( $\lambda\tau$ )	-0.3620 (0.30)	-	-0.0008 (0.010)	-0.1700 (0.710)	-	-
Proba de détection	-7.8759 *** (2.669)	-7.8252 *** (2.658)	-0.4400 ** (0.193)	-4.6824 ** (2.485)	-0.447 ** (0.190)	-4.7445 ** (2.483)
Propor. entourage	1.0380 *** (0.222)	1.0109 *** (0.221)	0.1050 *** (0.015)	-1.4100 (1.728)	0.1034 *** (0.015)	-1.6126 (1.741)
Réaction entourage	5.0652 *** (1.210)	5.1355 *** (1.2152)	0.2700 *** (0.080)	1.9191 * (1.093)	0.2817 *** (0.080)	1.9600 * (1.093)
Immoralité	-2.4946 *** (0.531)	-2.4498 *** (0.5294)	-0.1800 *** (0.040)	-6.2330 (4.573)	-0.1876 *** (0.043)	-6.6041 (4.544)
Montréal	2.6182 * (1.568)	2.5513 (1.5636)	0.2500 ** (0.110)	2.8390 (11.976)	0.2439 ** (0.114)	3.5085 (11.785)
Bas-du-Fleuve	-3.9558 (14.229)	-3.9990 (14.185)	-0.0500 (0.100)	-4.2271 (14.210)	-0.0582 (0.104)	-3.1566 (10.244)
Log-Vraisemblance	-2193.9	-2192.2	-2576.3	-2576.3	-2572.7	-2572.7

\* Statistiquement significatif à 10%

\*\* Statistiquement significatif à 5%

\*\*\* Statistiquement significatif à 1%.

Les deux premières colonnes retracent l'estimation d'un tobit standard sous l'hypothèse qu'il n'y a pas de coûts fixes associés à l'entrée sur le marché noir. Les variables ont le signe généralement attendu et confirment les résultats précédemment relatés. Toutefois, ce type de modèle ne permet pas de distinguer la décision d'entrer sur le marché noir et l'intensité du travail. Une telle distinction n'est possible que dans le cadre de tobit généralisé. Les paramètres estimés selon cette spécification particulière sont reportés aux colonnes (iii) à (vi).

Le modèle prévoit une relation négative entre le taux de salaire horaire sur le marché noir et l'offre de travail non déclaré. Ce résultat *a priori* contre-intuitif ne peut être attribuable à un problème de «biais de division»<sup>27</sup> puisque les heures, le revenu et le taux de salaire sont obtenus de façon indépendante dans l'enquête. Bien au contraire, il est en conformité avec les résultats précédemment obtenus par Lemieux *et al.* (1994). Comme l'ont expliqué ces auteurs, la corrélation négative observée reflète une relation concave entre le revenu et les heures travaillées, en raison des coûts croissants nécessaires à la dissimulation de l'activité souterraine.

Les résultats changent quelque peu lorsque la prise en compte des coûts fixes est permise. La statistique du rapport des vraisemblances pour tester les restrictions de sur-identification est de 3.41, alors que la valeur critique au seuil de 5% est  $\chi_{0.05}^2(1) = 3.84$ . Sur la base de ce test, nous ne pouvons rejeter la validité des restrictions apportées au modèle. Par ailleurs, la littérature récente sur l'utilisation d'instruments faiblement corrélés avec la variable dépendante endogène<sup>28</sup> montre que les estimations peuvent être fortement biaisées. Par conséquent, il importe de s'assurer qu'un tel problème ne se rencontre pas dans nos estimations. Nous devons vérifier que les restrictions supplémentaires (sur-identification du modèle), n'affectent pas la convergence de nos estimateurs. Nous avons donc testé l'hypothèse que l'ajout de ces instruments n'avait aucun effet sur les équations de première étape. Pour ce faire, nous avons utilisé la statistique  $F$  basée sur chacune des équations (2.31), (2.32) et (2.33) de notre modèle empirique. A partir de l'estimation par moindres carrés ordinaires de l'équation de salaire (2.31), nous obtenons une statistique de 3.42, ce qui est au-dessus de la valeur critique du test à 5% ( $F_{0.05} = 1.83$ ). Pour les équations (2.32) et (2.33), nous obtenons des statistiques calculées de 25.28 et 42.09. La valeur critique  $F_{0.05}$  étant de 1.75, nous rejetons l'hypothèse que nos instruments ne sont pas corrélés avec les variables endogènes du modèle.

Enfin, nous devons tester la présence de coûts fixes à l'entrée sur le marché noir. Or, ceci n'est pas facile à tester dans la mesure où le modèle tobit (avec ou sans régresseurs

<sup>27</sup> Voir Borjas (1969) à ce sujet.

<sup>28</sup> Voir, par exemple, Bound *et al.* (1995) et Staiger et Stock (1997).

endogènes) n'est pas emboîté dans un modèle tobit de type 2 (Blundell et Smith 1994). En supposant l'exogénéité des variables conditionnantes, le test de Chow<sup>29</sup> proposé par Scott et Garen (1994) peut être appliqué à notre cas. Comme nous l'avons mentionné à la *section 2.4.1* du présent chapitre, lorsqu'il n'y a pas de coûts fixes, les paramètres de l'équation latente associée à la probabilité de travailler au noir (paramètres  $\kappa$ ) sont égaux à ceux de l'équation latente associée aux heures (paramètres  $\gamma$ ).

$$\begin{aligned} h_2^* - h_2^R &= \kappa_1 \log W_2 + \kappa X_1 + v \\ h_2 &= h_2^* = \gamma_1 \log W_2 + \gamma X_1 + \varepsilon_1 \end{aligned}$$

Une estimation convergente de  $\bar{\kappa} = \kappa/\sigma_v$  peut être obtenue par une procédure de type probit. Puisque  $\gamma = \sigma_v \bar{\kappa}$ , les paramètres obtenus par l'estimation en trois étapes des heures conditionnelles devraient être proportionnels à ceux obtenus à partir du probit. Ces restrictions s'appliquent à l'estimation de l'équation de troisième étape. A l'instar de Scott et Garen (1994), nous comparerons la somme des résidus carrés des régressions contrainte et non contrainte par le test de Chow (avec correction de l'hétéroscédasticité). La statistique de ce test est de 5.30, tandis que la valeur critique est  $F_{0.05} = 1.75$ . Par conséquent, nous ne rejetons pas la spécification associée aux coûts fixes.

## 2.6 Conclusion

Bien que de nombreux travaux aient posé l'existence de coûts fixes à l'entrée sur le marché noir, aucun n'a, à notre connaissance, apporté un appui empirique à cette hypothèse. Dans ce chapitre, nous avons tenté d'étendre un modèle d'allocation du temps de travail entre activité déclarée et activité non déclarée, en incorporant un stigmate associé à la participation à l'économie souterraine. En ce sens, notre modèle essaie de généraliser l'analyse standard des solutions en coin. Le coût fixe se définit comme une perte d'utilité provoquée par l'entrée sur le marché noir. Il s'apparente à un coût fixe et intervient à différents niveaux. Non seulement, il affecte la décision de participer au marché noir, mais il conditionne également le volume horaire de l'activité souterraine. Il entraîne une discontinuité dans la fonction d'offre de travail telle que l'individu choisit de ne pas travailler en-dessous d'un certain nombre d'heures. Ce coût modifie les conditions selon lesquelles il est plus ou moins intéressant de travailler au noir que d'occuper un emploi officiel. Il implique de distinguer les heures de travail désirées et les heures de réserve.

<sup>29</sup>Pour plus de détails sur ce test de stabilité, se reporter à Wooldridge (2001).

Une approche conditionnelle, basée sur les variables instrumentales, nous a permis de développer un modèle économétrique qui tient compte de la présence de ces coûts à l'entrée. Nous procédons aux estimations selon une méthode en trois étapes, ce qui nous permet de contourner les problèmes de sélection de l'activité souterraine et d'endogénéité de variables salariales et relatives à l'activité officielle.

Les résultats de nos estimations économétriques donnent une indication de la nature de l'offre de travail au noir et révèlent l'importance de la prise en compte de ces coûts. Ces derniers, bien qu'inférieurs à ceux qui sont observés sur le marché officiel, sont néanmoins relativement conséquents. Ils contribuent dès lors à expliquer les taux de participation modérés à l'économie souterraine. Il caractérise l'opprobre jeté sur la dissimulation de la fraude et explique la très grande sensibilité des individus au jugement de la société. Nous observons, en effet, que les individus sont très sensibles au comportement de l'entourage et de la société dans son ensemble.

Par ailleurs, nous constatons que le système fiscal et de répression de la fraude affecte considérablement les comportements individuels sur le marché noir. Le système fiscal et pénal semble intervenir uniquement comme un frein à l'élargissement de l'économie souterraine, mais en aucun cas comme facteur dissuasif de la participation. Néanmoins, de nombreux effets sont contradictoires et, à la lumière de ce chapitre, nous ne sommes pas en mesure d'apporter des conclusions non ambiguës sur l'impact de la fiscalité. En outre, les questions touchant à la probabilité de détection et l'amende sont de nature à soulever des problèmes d'endogénéité. Cela est dû au fait que les individus ayant travaillé sur le marché noir sont susceptibles de sous-estimer les valeurs de  $\theta$  et  $p$ , de façon à réduire le stress associé à cette activité. Une telle attitude est à la source d'un biais de dissonance cognitive. En effet, selon la théorie de la dissonance cognitive empruntée à Akerlof et Dickens (1982), les individus ont des préférences quant à l'état de la nature et quant à leurs croyances en l'état de la nature. Elle nous enseigne, en outre, qu'ils peuvent exercer un certain contrôle sur leurs croyances. Ils peuvent, en effet, manipuler leurs croyances en sélectionnant les sources d'information qui confirment leurs désirs. Tout comme les travailleurs d'emplois dangereux sous-estiment le danger inhérent à leur activité, les participants à l'économie souterraine peuvent se convaincre du caractère peu risqué de l'activité non déclarée. Ils rejettent généralement tout sentiment d'insécurité et déclarent de faibles probabilités de sanctions. Par conséquent, le risque de détection et l'amende associée sont endogènes à la participation au marché noir et doivent être traitées comme telles. C'est pourquoi, nous proposons de prolonger l'analyse de l'impact du système fiscal sur l'offre individuelle de travail au noir. Celle-ci doit permettre de tenir compte de l'endogénéité de ces paramètres et d'écartier ainsi le biais potentiel de dissonance cognitive.

Enfin, nous avons procédé à l'étude du comportement d'offre de travail au noir conditionnelle au statut de l'individu sur le marché officiel. Cette hypothèse restrictive s'est révélée très utile pour la détermination des coûts de participation au marché noir. Néanmoins, il paraît évident que l'individu détermine simultanément son offre de travail sur le marché noir et sur le marché officiel. Par conséquent, l'hypothèse de décisions séparables doit être levée. Au chapitre suivant, nous proposons une modélisation des comportements individuels simultanés sur chacun des deux marchés afin de rendre compte plus explicitement du poids de la fiscalité dans l'offre de travail au noir.