



Les documents de recherche sont des documents de travail. A ce titre, ils n'engagent que leurs auteurs et ne reflètent pas la position de l'Acoss.

Par ailleurs, toutes erreurs ou omissions restent de la responsabilité des auteurs.

Fraude sociale et dispositif de répression

*Marianne Cornu Pauchet**
*Nadia Joubert**
*Thierry Senta***

N°2007- 01
Septembre 2007

**Mission recherche, études et publications - Direction des statistiques, des études et de la prévision - Acoss*

*** Mission recherche, études et publications - Acoss et Urssaf de Lyon*

Fraude sociale et dispositifs de répression

Marianne CORNU PAUCHET*, Nadia JOUBERT*, et Thierry SENTA**

Résumé

Les dispositifs de répression de la fraude consistent généralement à sélectionner un échantillon de ménages ou d'entreprises « à risque » sur la base d'informations administratives ou statistiques, puis à contrôler cet échantillon. Or, ces contrôles ciblés sont peu adaptés pour évaluer l'ampleur de la fraude. L'exploitation des informations obtenues est en effet susceptible de conduire à une estimation biaisée de l'observation des règles fiscales ou sociales dans le pays et de l'ampleur des activités souterraines au sens large. L'Acoss et les Urssaf ont donc développé des programmes d'audits aléatoires pour évaluer l'ampleur de la sous-déclaration des cotisations de Sécurité sociale, en s'affranchissant des biais liés aux programmes sélectifs. Cet article fournit, à partir des données recueillies lors de contrôles sur le secteur HCR, une première évaluation du montant des cotisations éludées du fait du recours au travail dissimulé. On montre aussi que le dispositif de contrôle aléatoire a conduit à une nette hausse des montants des salaires déclarés au Régime général de Sécurité sociale par les établissements contrôlés. Cet effet est mesuré à l'aide d'une méthode microéconométrique d'évaluation permettant de prendre en compte un éventuel biais de sélection dû à des variables inobservables.

Abstract

We use a direct micro approach to evaluate the magnitude of Social Security tax evasion in the Hotel and restaurant sector for two regions. Because traditional tax auditing (based on the selection of taxpayers as regards to elements indicating some likelihood of fraud) is likely to entail a selection bias, random audits have been implemented. Using recent econometrics methodology, we first estimate the effect of control on tax compliance. The control has led to a high increase in the amounts declared by the establishments to the Social Security. With this estimation, we can also conclude that our sample is actually a random one of the whole population of the Hotel and restaurant sector in the two regions considered. It is then possible to evaluate the amount of Social Security contributions hidden by this population.

Mots clefs : économie souterraine, fiscalité, transferts sociaux, impact du passage par un dispositif, méthode de sélection sur inobservable.

Code JEL : C34, D21, D81, H26 H32.

Cet article a été réalisé à partir des données de l'Agence centrale des organismes de Sécurité sociale (ACOSS). Nous sommes reconnaissants envers Alain Gubian, Directeur de la Direction des statistiques, des études et de la prévision, pour ses commentaires et suggestions fort utiles. Nous tenons également à remercier l'ensemble des membres de la Direction statistique de l'Acoss pour leurs remarques. Les éventuelles erreurs restent toutefois de notre responsabilité.

* Mission recherche de la Direction des statistiques, des études et de la prévision (DISEP), Agence centrale des organismes de Sécurité sociale (ACOSS)-65, Bd R. Lenoir, 75536 Paris Cedex 11. marianne.pauchet@acoss.fr

**Mission recherche de la DISEP-ACOSS et Département du contrôle interne et statistique (CIS), Urssaf de Lyon, 6, rue du 19 mars 1962, 69691 Vénissieux Cedex.

1. Introduction

Il est communément admis que de nombreuses infractions aux lois fiscales et sociales demeurent largement non détectées par les pouvoirs publics. Pourtant la plupart des pays tentent de limiter ces activités frauduleuses par le biais de contrôles intensifs.

L'objectif du présent papier est double. Il s'agit, d'une part, d'évaluer l'impact des audits aléatoires sur les déclarations volontaires et spontanées de masse salariale par les entreprises cotisant au Régime général de Sécurité sociale et d'autre part, d'utiliser l'information recueillie lors de ces audits pour apprécier le niveau d'observation des lois sociales en France (« *compliance* »).

Afin de répondre à cet objectif, une méthodologie originale a été retenue. Celle-ci s'inscrit dans le cadre des travaux de Feinstein (1999) et implique un recours massif aux audits aléatoires auprès de la population. Elle permet ainsi d'obtenir des informations non seulement extrêmement détaillées, mais également non biaisées, sur les comportements individuels. De ce fait, il est possible de mesurer l'ampleur de la sous-déclaration et l'efficacité des politiques répressives de lutte contre le travail au noir en France.

Pour évaluer l'ampleur de la fraude, les contrôles *ciblés* traditionnels restent peu adaptés. En premier lieu, l'échantillon contrôlé est limité aux contribuables ou aux entreprises connues par les services fiscaux ou sociaux, ce qui induit de fait un biais de sous-estimation de l'ampleur de la fraude. Par ailleurs, le choix des personnes à contrôler est, en général, fondé sur des signes permettant au regard des déclarations et de recoupements administratifs et statistiques, de soupçonner une fraude. Comme les résultats témoignent d'un ciblage délibéré sur les cas où les probabilités de fraude sont les plus fortes, le groupe contrôlé n'est pas représentatif des pratiques courantes de la population dans son ensemble et les résultats sont susceptibles de conduire à une sur-estimation de l'ampleur de la fraude. Tous ces facteurs sont donc source d'un biais potentiellement important de sélection, ce qui est susceptible de conduire à une estimation biaisée de l'observation des règles fiscales ou sociales dans le pays et de l'ampleur des activités souterraines au sens large. C'est pourquoi, le phénomène de la fraude ne saurait être estimé sur l'ensemble de la population par extrapolation des résultats constatés sur les contrôles fiscaux ou sociaux traditionnels (Courson *et al.*, 1996). En second lieu, les estimations fondées sur les contrôles fiscaux ou sociaux reflètent seulement la part du revenu dissimulé que les autorités parviennent à détecter. Or, il est fort probable que celle-ci ne constitue qu'une fraction des revenus non déclarés. En ce sens, elle conduit à une sous-estimation de l'économie souterraine. Enfin, l'identification de la fraude est réalisée à un instant *t*. Cette approche s'avère donc incapable de fournir une évaluation de la croissance de l'économie souterraine, ni de mesurer son évolution sur longue période. Elle ne permet pas non plus de suivre les individus dans le temps. En l'absence de données longitudinales et puisque seuls les fraudeurs sont

identifiés, il est alors particulièrement délicat, sinon impossible, d'analyser des comportements alternants la fraude et l'absence de fraude (Schneider et Enste, 2000).

En revanche, cette approche traditionnelle de contrôles ciblés détient un avantage indéniable sur bon nombre de méthodes disponibles, dans la mesure où elle procure une information détaillée sur le type d'activités sujettes à la fraude, ainsi que sur les caractéristiques particulières des individus s'adonnant à de telles activités. Elle a ainsi permis de mettre en évidence qu'aux Etats-Unis, les travailleurs indépendants ont tendance à verser volontairement une part plus faible de leur contribution fiscale que d'autres contribuables (Feinstein, 1999). De ce fait, une procédure d'échantillonnage par strates, avec pondération plus élevée pour les groupes à risque, pourrait se révéler plus efficace.

D'un point de vue administratif, l'exigence de représentativité pour évaluer l'ampleur de la fraude dans un pays signifie souvent qu'un programme avec une méthodologie de contrôle spécifique, et financé de façon indépendante, doit être mis en place pour collecter les données appropriées. Un exemple très connu d'estimation de l'inobservation des règles fiscales est le programme américain intitulé « *Taxpayer Compliance Measurement Program* » (« *TCMP* »). Les services fiscaux ont conduit différents programmes de mesure sur les ménages et les petites entreprises depuis 1973. Chaque programme implique des contrôles intensifs sur un échantillon tiré de manière aléatoire et stratifié à partir des déclarations. Il fournit des données détaillées sur les contrôles et les taux de poursuites, la sévérité des peines, les sources de revenus, la répartition géographique, démographique et sociale des contribuables.¹ Le programme le plus récent a eu lieu en 1988 sur 50 000 ménages.² L'information contenue dans les contrôles a ensuite été utilisée pour estimer la fréquence et l'amplitude de la fraude. Elle a également contribué à aider le personnel des services fiscaux dans l'élaboration de méthodes de contrôles. Les chiffres de 1988 indiquent qu'environ 40% des foyers américains procèdent intentionnellement à une sous-déclaration de leurs impôts, 53% se conforment normalement à la loi et 7% payent davantage d'impôts que nécessaire. Les paiements excessifs restent cependant modérés puisque les versements médians correspondent à 158 dollars US et sont certainement dus à des erreurs. Cela signifie qu'en 1988, près de deux-tiers des contribuables américains avaient l'intention de payer correctement leurs impôts. Toutefois, une proportion substantielle des sous-déclarations portait sur des montants élevés. Un quart des contribuables était ainsi détecté pour une fraude de plus de 1 500 dollars. Au-delà de la connaissance globale de l'ampleur des impôts non recouverts, de telles procédures de contrôle permettent de cerner les différents comportements de fraude.

¹ - Bien qu'il s'agisse d'informations relatives aux déclarations d'impôts individuelles, les données sont rendues accessibles aux chercheurs qui le souhaitent, sous la forme d'observations agrégées par région selon un code postal à trois chiffres.

² - Pour de plus amples détails, se reporter à IRS (1996).

Toutefois, les programmes de contrôles aléatoires destinés à l'évaluation du non-respect des lois et règlements sont en pratique très peu répandus, notamment en raison du coût extrêmement élevé nécessaire à la collecte de telles données³. Pourtant les enjeux de l'évaluation de l'ampleur de la fraude sont majeurs. En effet, cette dernière a un certain nombre de conséquences négatives. Elle se traduit tout d'abord par une érosion de l'assiette sur laquelle porte la charge fiscale et sont calculées les cotisations de Sécurité sociale, pouvant ainsi réduire les marges des recettes publiques disponibles pour financer les priorités et conduire à une hausse des taux de taxation et des taux de cotisation. L'économie souterraine a aussi des effets directs sur l'économie officielle parce qu'elle détourne les capacités productives mais aussi parce qu'elle fausse les évaluations des grands agrégats économiques (PIB, taux de chômage,...) et les politiques économiques qui découlent de ces évaluations. Enfin, l'existence d'une économie souterraine engendre également des distorsions concurrentielles, en particulier entre entreprises d'un même secteur. Il est donc essentiel de développer des programmes d'audits *aléatoires* pour évaluer l'ampleur de la dissimulation d'activité en s'affranchissant des biais liés aux programmes *sélectifs*. L'équivalent du programme « *Taxpayer Compliance Measurement Program* » (« *TCMP* ») a été développé par l'Acoss et les Urssaf pour évaluer la sous-déclaration de cotisations de Sécurité sociale.

Cet article présente les évaluations réalisées à partir de l'information recueillie lors de ce programme mené dans deux zones (méditerranéenne et atlantique) sur le secteur des Hôtels cafés restaurants (HCR). Les cotisations éludées en raison du travail dissimulé représentent entre 7 % et 14 % de l'ensemble des cotisations dues par les HCR dans les zones concernées. De plus, on évalue à partir de la méthode microéconométrique d'évaluation du passage par un dispositif, l'impact du contrôle sur les déclarations spontanées de masse salariale. On montre que les contrôles se sont traduits par un accroissement des déclarations de masse salariale de l'ordre de 3 à 4% pour les établissements contrôlés.

La structure de l'article est la suivante. Les données et une analyse descriptive de l'information sont réalisées dans la deuxième section. La méthode microéconométrique d'évaluation du passage par un dispositif est présentée dans la section 3. Enfin, les principaux résultats sont exposés dans la section 4.

2. Analyse descriptive

2.1 Récolte des informations

³ ■ Les programmes *TCMP* ont notamment montré que le rendement direct du contrôle aléatoire est beaucoup plus faible, puisqu'il est inférieur à 300\$ par contrôle alors qu'il s'élève à plus de 5500\$ pour le contrôle habituel.

De juillet à septembre 2004, des contrôles pour la recherche de travail dissimulé ont été menés par les Urssaf sur les établissements du secteur Hôtel, Café Restaurant (HCR) dans 2 zones géographiques : une zone méditerranéenne et une zone atlantique. Ces contrôles ont été réalisés de manière aléatoire. A l'inverse des contrôles traditionnels, aucune recherche préalable sur le fichier des établissements immatriculés en Urssaf ne devait être réalisée. Il était demandé aux inspecteurs de couvrir au mieux l'ensemble du territoire dans les départements concernés.

Les contrôles portaient sur l'établissement et les salariés trouvés sur place. Ils ont ainsi permis de recueillir des données très riches (*encadré 1*), avec des informations sur l'établissement contrôlé et sur les salariés auditionnés ; ces données étant ensuite complétées par l'Urssaf avec les suites données au contrôle. Ces informations ont été rapprochées d'autres informations disponibles à l'Acoss : plusieurs appariements ont été réalisés, avec les données des déclarations volontaires et spontanées de masse salariale (base Séquoia) ainsi qu'avec les données sur les DUE (Cèdre, *encadré*).

2.2 Construction des échantillons

Pour réaliser l'analyse d'impact du contrôle sur les déclarations spontanées de masse salariale, il est nécessaire de construire un échantillon témoin d'établissements non contrôlés ayant les mêmes caractéristiques que les établissements contrôlés. On obtient ainsi un échantillon d'environ 700 établissements contrôlés et un échantillon témoin de plus de 4000 établissements non contrôlés (*tableau 1*). Il s'agit d'établissements « pérennes », présents à la fois en septembre 2003 et septembre 2004⁴. Ces 5000 établissements représentent environ 37 % des établissements et 7% de la masse salariale du secteur HCR dans les deux zones concernées. Leur masse salariale s'est globalement accrue de 3,6 % entre septembre 2003 et septembre 2004.

Tableau 1 : Moyenne des taux de croissance individuels de la masse salariale des établissements

Zone géographique : Méditerranée			Zone géographique : Atlantique			Zones géographiques : Méditerranée et Atlantique		
Contrôlés	Non Contrôlés	Ensemble	Contrôlés	Non Contrôlés	Ensemble	Contrôlés	Non contrôlés	Ensemble
427	2 629	3 056	272	1 589	1 861	699	4218	4917

⁴ - Etablissements pour lesquels il est possible de construire le taux de croissance de la masse salariale entre septembre 2003 et septembre 2004 qui sera notre variable de résultat, voir en infra l'application économétrique.

Encadré : Les sources de données

Les données utilisées sont issues de trois sources :

La base issue des contrôles :

La base des contrôles aléatoires contient des données très riches sur les établissements contrôlés et les salariés auditionnés lors des audits aléatoires de l'été 2004 dans le secteur des Hôtels-Cafés-Restaurants (HCR). Pour chaque établissement, on dispose de la nature de son activité (saisonniers ou permanents), de sa date de création, du détail du secteur (en code Naf700) et du nombre de salariés auditionnés. Pour chaque salarié interrogé, l'information comporte le fait que le salarié a fait l'objet ou non d'une Déclaration unique d'embauche (DUE) ou d'une adhésion au titre emploi entreprise (TEE), l'âge du salarié, sa date d'entrée dans l'établissement, le type de contrat (CDD, CDI, intérim,...), le type d'activité et les horaires effectués. La base contient aussi des informations sur les suites données aux contrôles par les inspecteurs des Urssaf. Lorsqu'une suite est donnée au contrôle, elle prend la forme soit d'un procès verbal de travail dissimulé, soit d'un suivi de l'affaire avec deux types de préconisations : à revoir en contrôle comptable d'assiette ou en contrôle travail dissimulé.

La base Sequoia

La base Sequoia (Système pour l'Etude Quantitative et l'Observation des Assiettes) regroupe les données individuelles sur les effectifs et les salaires des établissements et des administrations cotisant au régime général de Sécurité sociale depuis 1997. Les employeurs affiliés au régime général déclarent et paient leurs cotisations sociales aux Urssaf à l'aide du Bordereau Récapitulatif des Cotisations (BRC) sur lequel ils portent les différentes assiettes salariales (plafonnée, déplafonnée, CSG) donnant lieu à cotisations ainsi que leurs effectifs. Cette déclaration est mensuelle si l'établissement a plus de 10 salariés et trimestrielle en deçà de ce seuil sauf demande contraire du cotisant (12 % des établissements sont dans ce cas). La base Sequoia comporte à ce jour 530 000 comptes mensuels et 1 380 000 comptes trimestriels. Elles sont désaisonnalisées pour corriger notamment l'impact des versements de primes et les fluctuations saisonnières de l'emploi. Le champ couvre l'ensemble des cotisants exerçant leur activité en France (Métropole et Dom) dans un secteur concurrentiel qui comprend tous les secteurs d'activité économique sauf les administrations publiques (codes 751 et 752 de la NAF), l'éducation non marchande (établissements d'enseignement relevant de l'Etat ou des collectivités locales), la santé non marchande et l'emploi par les ménages de salariés à domicile.

La base Cèdre

La base Cèdre regroupe les déclarations uniques d'embauche (DUE). Le champ couvert est l'ensemble des activités concurrentielles, non compris le secteur de l'intérim.

Les caractéristiques de l'échantillon des établissements contrôlés et non contrôlés figurent dans le *tableau 2*.

Tableau 2 : Structure des échantillons (établissements contrôlés et non contrôlés)

Caractéristiques	en %	
	Etablissements contrôlés	Etablissements non contrôlés
Nombre d'établissements		
Etablissements des Zones Géographiques : Méditerranée et Atlantique	14,22%	85,78%
Etablissements de la Zone Géographique : Méditerranée	13,97%	86,03%
Etablissements de la Zone Géographique : Atlantique	14,62%	85,38%
Nombre de salariés Trimestre 3 2004		
[0]	1,1%	1,2%
] 0 ; 2]	39,1%	41,9%
] 2 ; 4]	25,8%	21,6%
] 4 ; 6]	12,9%	10,8%
] 6 ; 8]	10,2%	7,4%
] 8 ; 10]	4,3%	5,0%
] 10 ; 20]	5,7%	8,2%
] 20 ; 43]	1,0%	3,9%
Codes APE (1)		
551A Hôtels touristiques avec restaurant	13,2%	16,2%
551C Hôtels touristiques sans restaurant	4,0%	4,1%
552A Auberges de jeunesse et refuges	0,1%	0,1%
552C Exploitation de terrains de camping	5,0%	6,7%
552E Autre hébergement touristique	0,1%	5,0%
553A Restauration de type traditionnel	44,8%	39,1%
553B Restauration de type rapide	19,2%	15,4%
554A Cafés tabacs	3,3%	3,0%
554B Débits de boissons	9,6%	9,3%
554C Discothèques	0,7%	1,1%
Type de déclaration		
Etablissements avec des Déclarations Mensuelles (plus de 10 salariés)	6,2%	12,4%
Etablissements avec des Déclarations Trimestrielles (moins de 10 salariés)	93,8%	87,6%
Zones d'Emploi		
9351 Toulon	44,7%	49,4%
9352 Fréjus Saint Raphael	32,3%	35,4%
9353 Draguignan	17,6%	7,1%
9354 Brignoles	5,4%	8,1%
Total	100,0%	100,0%
7206 Dax		20,3%
7272 Bordeaux Arcachonnais		7,4%
7281 Mont de Marsan Haute Lande		2,1%
7282 Mont de Marsan Est des Landes		9,3%
7291 Sud des Landes	2,9%	19,3%
7292 Bayonne Pyrénées	97,1%	41,6%
Total	100,0%	100,0%

(1) Les codes APE qui n'ont pas été contrôlés sont les suivants : Autres hôtels (551E), Hébergement collectif non touristique (552F), Cantines et restaurants d'entreprises (555A), Restauration Collective sous Contrat (555C) et Traiteurs, organisation de réception (555D)

La comparaison des taux de croissance moyen des établissements contrôlés et non contrôlés nous permet d'avoir une première indication de l'impact des contrôles sur la dynamique de la masse salariale.

Tableau 3 : Comparaison des moyennes des taux de croissance des établissements

Caractéristiques	Etablissements contrôlés	Etablissements non contrôlés
Nombre d'établissements		
Etablissements des Zones Géographiques : Méditerranée et Atlantique	32,9	21,1
Etablissements de la zones géographique Méditerranée	19,4	14,1
Etablissements de la zone géographique Atlantique	54,3	32,8
Nombre de salariés Trimestre 3 2004		
[0]	NS	13,0
] 0 ; 2]	24,5	30,0
] 2 ; 4]	64,5	19,8
] 4 ; 6]	31,5	14,8
] 6 ; 8]	3,9	11,5
] 8 ; 10]	8,4	11,0
] 10 ; 20]	9,8	10,8
] 20 ; 43]	-3,7	6,3
Codes APE		
551A Hôtels touristiques avec restaurant	11,7	8,4
551C Hôtels touristiques sans restaurant	16,7	11,7
552A Auberges de jeunesse et refuges	13,5	-3,2
552C Exploitation de terrains de camping	8,6	18,3
552E Autre hébergement touristique	NS	7,3
553A Restauration de type traditionnel	40,9	29,9
553B Restauration de type rapide	35,2	21,8
554A Cafés tabacs	29,1	15,7
554B Débits de boissons	42,4	23,2
554C Discothèques	4,2	3,7
Type de déclaration		
Etablissements avec des Déclarations Mensuelles (plus de 10 salariés)	2,2	3,3
Etablissements avec des Déclarations Trimestrielles (moins de 10 salariés)	35,0	23,7

NS= non significatif

On constate que les établissements contrôlés avec un nombre de salariés au troisième trimestre de l'année 2004 compris entre 2 et 6 salariés ont une moyenne des taux de croissance supérieure à celle des établissements non contrôlés. De manière plus générale, les établissements contrôlés de moins de 10 salariés (qui réalisent leurs déclarations de manière trimestrielle) ont une moyenne des taux de croissance supérieure à celle des établissements non contrôlés. Ce constat est également vrai pour les établissements des secteurs hôtels touristiques avec restaurant (551A), hôtels touristiques sans restaurant (551C), auberges de jeunesse et refuges (552A), restauration de type traditionnel (553A), restauration de type rapide (553B), cafés tabacs (554A), débits de boissons (554B) et discothèques (554C). La dynamique de la masse salariale déclarée aux Urssaf serait donc plus importante pour les établissements contrôlés que pour les établissements non contrôlés.

2.3 Comparaisons de distributions : le test de Li (1996)

	Zone Géographique :	Statistique de Li	Proba
Test de Li :	Méditerranée et Atlantique	665,88	0,00%
	Méditerranée	242,53	0,00%
	Atlantique	231,19	0,00%

Nous avons appliqué à notre échantillon un test de Li (1996)⁵ afin de comparer les distributions des taux de croissance des masses salariales des établissements contrôlés et non contrôlés. Les résultats du test de Li (*tableau 4*) démontrent que la distribution du taux de croissance de la masse salariale des établissements contrôlés est significativement différente de celle des établissements non contrôlés. Par ailleurs, ce résultat est robuste que l'on compare les distributions au niveau de l'ensemble de l'échantillon ou pour chaque zone géographique. Cela conforte l'idée selon laquelle le contrôle aurait un impact positif sur les déclarations volontaires et spontanées de masse salariale. Le développement d'un modèle micro-économétrique d'évaluation de l'impact du passage par un dispositif nous permettra de valider formellement cette hypothèse. Il permettra de plus d'en mesurer l'ampleur.

Tableau 4 : extrapolation des montants de cotisations éludés

Total des salariés HCR des deux zones	Taux de salariés dissimulés détectés	Extrapolation du nombre de salariés dissimulés	Rémunération mensuelle par salarié dissimulé (en euros)	Extrapolation de la dissimulation pour l'été 2004		
				Rémunérations non déclarées En millions d'euros	Cotisations éludées En millions d'euros	En % des cotisations dues
49 510	19 %	9 650	916	26,5	11,9	11,5 %

3. Méthode microéconométrique d'évaluation

La méthode économétrique employée est le modèle de micro-économétrie d'évaluation du passage par un dispositif. Dans notre étude, le « dispositif » est le contrôle aléatoire. Il existe deux états potentiels de la nature : un établissement est soit contrôlé, soit non contrôlé :

⁵ Le test de Li (1996) est un test non paramétrique de comparaison de deux distributions dont on ne connaît pas la fonction de distribution. Sous H_0 , les 2 distributions ne sont pas significativement différentes :

$$H_0 : \hat{f}(x^f) = \hat{g}(x^g)$$

$$H_1 : \hat{f}(x^f) \neq \hat{g}(x^g)$$

La statistique de Li est la suivante :

$$\xi = \frac{N \cdot h^{0,5} \cdot \hat{I}}{\hat{\sigma}} \rightarrow N(0,1)$$

$$D = \begin{cases} 1 & \text{si l'établissement est contrôlé (Etat 1)} \\ 0 & \text{si l'établissement n'est pas contrôlé (Etat 0)} \end{cases}$$

où D est la variable de sélection (variable indicatrice du Contrôle).

Chaque établissement est caractérisé par le couple de variables aléatoires (Y_1, Y_0) où Y_1 est la variable de résultat des établissements en cas de contrôle et Y_0 est celle des établissements en cas de non contrôle. Dans notre cas, cette variable de résultat correspond au facteur de croissance de la masse salariale. La variable effectivement observée Y_{obs} correspond à l'état effectif de l'entreprise et s'écrit :

$$Y_{obs} = D \cdot Y_1 + (1-D) \cdot Y_0 = \begin{cases} Y_1 & \text{si } D = 1 \\ Y_0 & \text{si } D = 0 \end{cases}$$

Pour évaluer l'impact du contrôle, il faudrait pouvoir comparer la situation des établissements contrôlés avec leur situation s'ils n'avaient pas été contrôlés. Or, ces données ne sont pas disponibles. On est donc confronté à un problème d'identification puisque seule une des 2 variables Y_0 ou Y_1 est observée. Pour évaluer l'effet moyen du contrôle sur les contrôlés qui correspond ici à la variation moyenne de masse salariale qui résulte du contrôle, on utilise l'information disponible sur un groupe « témoin » d'entreprises c-a-d sur un groupe d'entreprises non contrôlées ayant des caractéristiques similaires au groupe des entreprises contrôlées.

Le modèle utilisé ici est le modèle causal de Rubin (1974), dit aussi de « sélection sur les inobservables ». Ce modèle repose sur deux équations. L'équation dite « de sélection » définit et explique la probabilité d'être contrôlé pour chaque établissement. L'équation de résultat modélise la variable de résultat (Y).

Le modèle de sélection s'écrit :

$$D_i = \begin{cases} 1 & \text{si } Z_i \cdot \gamma + v_i \geq 0 \\ 0 & \text{si } Z_i \cdot \gamma + v_i < 0 \end{cases}$$

où D est une indicatrice ($D = 1$ si l'établissement est contrôlé et $D = 0$ sinon) et Z est l'ensemble de variables exogènes observées et « v » un résidu centré réduit⁶.

Le modèle de résultat s'écrit :

$$\begin{cases} Y_{1,i} = c_1 + X_i \cdot \beta_1 + U_{1,i} \\ Y_{0,i} = c_0 + X_i \cdot \beta_0 + U_{0,i} \end{cases}$$

⁶ $\sigma_v^2 = 1$ est nécessaire pour l'estimation de γ (condition d'identification du modèle probit). Par ailleurs, La première hypothèse posée est celle de la séparabilité entre variables observées et inobservées. On suppose également que X et Z sont indépendants de U_0 , U_1 et v. Enfin, conditionnellement à X, la décision de sélection d'un établissement est une fonction monotone et non constante de Z.

où i désigne l'établissement, Y est la variable de résultat, X est l'ensemble des variables explicatives observées et (U_0, U_1) les termes d'erreurs non observés de moyenne nulle. Les résidus V , U_0 et U_1 sont supposés suivre une loi normale de moyenne nulle et de matrice de variances et covariances Ω :

$$\begin{pmatrix} V_i \\ U_{0,i} \\ U_{1,i} \end{pmatrix} \rightarrow N(0, \Omega)$$

avec

$$\Omega = \begin{pmatrix} 1 & \rho_0 \cdot \sigma_0 & \rho_1 \cdot \sigma_1 \\ \rho_0 \cdot \sigma_0 & \sigma_0^2 & \sigma_{01} \\ \rho_1 \cdot \sigma_1 & \sigma_{01} & \sigma_1^2 \end{pmatrix} = \begin{pmatrix} V(v_i) & \text{cov}(v_i, U_{0,i}) & \text{cov}(v_i, U_{1,i}) \\ \text{cov}(v_i, U_{0,i}) & V(U_{0,i}) & \text{cov}(U_{0,i}, U_{1,i}) \\ \text{cov}(v_i, U_{1,i}) & \text{cov}(U_{0,i}, U_{1,i}) & V(U_{1,i}) \end{pmatrix}$$

où σ_j^2 est la variance de U_j ($j = 0, 1$), σ_{01} est la covariance entre U_0 et U_1 , et ρ_j est le coefficient de corrélation linéaire entre V et U_j

Différentes mesures sont à notre disposition pour évaluer l'impact des contrôles. Le Gain individuel associé au Contrôle (Effet Causal) représente la différence pour un établissement entre ce que serait sa situation s'il était contrôlé et ce qu'elle serait si cet établissement n'était pas contrôlé ($\Delta = Y_1 - Y_0$). Ce gain individuel est inobservable pour chaque établissement puisque soit l'établissement a été contrôlé, soit il ne l'a pas été.

L'effet naïf du contrôle est l'écart entre la situation moyenne des établissements contrôlés et celle des établissements non contrôlés ($\Delta_{\text{Naïf}} = E(Y|D=1) - E(Y|D=0) = \bar{Y}_1 - \bar{Y}_0$). Cet effet a l'avantage d'être directement observable sur les données.

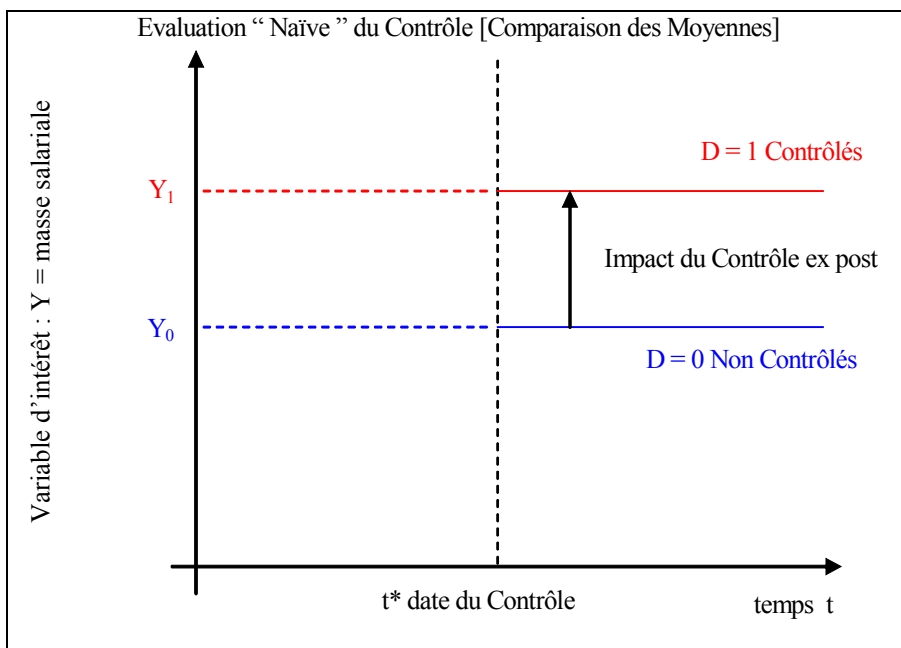
L'effet moyen du contrôle est la mesure de l'impact du contrôle si celui-ci est étendu à l'ensemble de la population ($\Delta_M = E(Y_1 - Y_0)$). L'effet moyen du contrôle sur les contrôlés est la mesure de l'impact du contrôle sur les établissements contrôlés. Il permet de juger au mieux de la performance du contrôle ($\Delta_{\text{TT}} = E(Y_1 - Y_0 | D=1)$). L'effet moyen du contrôle sur les non contrôlés est la mesure de l'impact sur les établissements non contrôlés si ces derniers étaient passés par le contrôle ($\Delta_{\text{TNT}} = E(Y_1 - Y_0 | D=0)$). L'effet moyen du contrôle sur les non contrôlés ne mesure pas l'effet réel qu'a eu le contrôle sur les non contrôlés mais un effet potentiel. Il ne doit pas être confondu avec l'effet de diffusion que peut avoir le contrôle sur les établissements non contrôlés (« peur du gendarme » et accroissement des salaires déclarés).

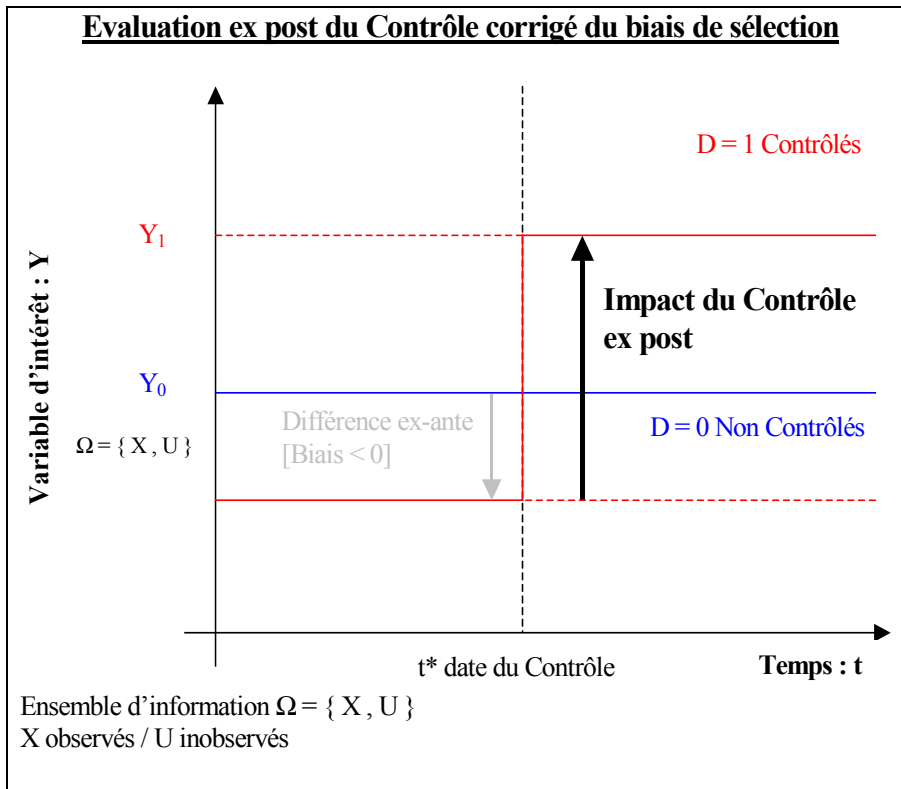
Lorsque l'effet moyen du contrôle sur les contrôlés « Δ_{TT} » et l'effet moyen du contrôle « Δ_M » sont égaux, la mesure de la performance du contrôle sur une population limitée

donnée par Δ_{TT} fournit directement la mesure d'impact du contrôle s'il était étendu à l'ensemble de la population. La différence entre ces deux paramètres mesure le biais de dit de « sélectivité ». Dans notre contexte, il est possible que « l'expérience » des inspecteurs de l'URSSAF les aient conduit à sélectionner certains établissements. Cette sélection qui repose sur des éléments inobservables peut être source de biais. Le modèle économétrique de sélectivité sur éléments inobservables nous permettra de contrôler cet éventuel biais.

Les deux graphiques suivants permettent de caractériser l'impact du contrôle.

Impact du Contrôle : cas où le contrôle a un impact positif sur les déclarations spontanées de masse salariale





4. Résultats

4.1 Evaluation de la dissimulation

Il convient de noter que, dans le contexte des opérations de contrôle aléatoire, il s'avère impossible de détecter la dissimulation partielle (e.g. heures supplémentaires non déclarées). L'évaluation présente ne concerne donc que la dissimulation totale de salariés, au sein d'établissements totalement dissimulés, mais également au sein d'établissements immatriculés auprès des Urssaf. En conséquence, l'on doit s'attendre à ce que cette évaluation minore l'ampleur globale de la dissimulation, en raison d'une part du fait qu'elle ne permet d'identifier qu'une fraction de la fraude et d'autre part, que la seule fraude qui soit identifiée est celle qui a été détectée, donc vraisemblablement celle qui est la moins efficace.

Tableau 5 : échantillon des salariés auditionnés lors des contrôles

Nbre de salariés HCR	Nbre salariés auditionnés	Taux de sondage	Nbre indiv. employés au noir	Part de travailleurs dissimulés détectés
(1)	(2)	(3)=(1)/(2)	(4)	(5)=(4)/(2)
49 510	3276	6.6 %	638	19 %

Au total, 638 individus ont été trouvés en situation d'emploi dissimulé. Parmi ceux-ci, 178 ont rapporté un montant de rémunération pour un total de 163 000 euros, soit une rémunération mensuelle moyenne d'environ 920 euros.

En appliquant le taux de rémunération moyen par secteur fin d'activité (NAF 700) aux individus dissimulés n'ayant déclaré aucune rémunération, l'on estime l'ampleur des rémunérations mensuelles dissimulées pour l'ensemble des salariés en situation de fraude à 580 000 euros, soit avec un taux de cotisations de 45%, 261 000 euros de cotisations éludées.

- **Montant annuel des cotisations éludées sur l'échantillon des établissements audités**

Nous détenons l'information relative à la durée du contrat de travail des individus en situation de fraude. De ce fait, il est possible d'extrapoler à l'ensemble de l'année, le montant des rémunérations dissimulées des salariés détectés dans le secteur HCR. La durée moyenne du contrat de travail pour ces individus est de 19 à 24 semaines selon le secteur d'activité. En supposant d'une part que ces individus ne sont pas employés dans le secteur durant le reste de l'année et d'autre part qu'ils exercent une activité dans le secteur HCR toute l'année, on obtient :

- Sous l'**hypothèse 1** (*borne inférieure*) : les individus n'exercent pas d'activité similaire durant la saison hivernale, ce montant s'élève à **3 100 000 euros**.

3 100 000 €	x	45%	=	1 400 000 €
rémunération globale au noir		taux cotisation		cotisations éludées pour l'année 2004.

- Sous l'**hypothèse 2** (*borne supérieure*) : les individus qui exercent une activité saisonnière estivale dans le secteur HCR, occupent également un emploi saisonnier dans le même secteur HCR sur la période hivernale. On suppose néanmoins que ces individus s'accordent un mois de repos. Dès lors, la rémunération dissimulée totale sur l'année s'élève à **4 300 000 €**.

6 400 000 €	x	45%	=	2 900 000 €
de rémunération globale au noir dans le secteur HCR année 2004		taux de cotisation		cotisations éludées pour toute l'année 2004

- **Extrapolation à l'ensemble du secteur HCR des zones géographiques concernées**

En l'absence de biais de sélection, ces résultats peuvent être extrapolés à l'ensemble de la population étudiée en tenant compte du taux de fraude identifiée lors des contrôles

aléatoires. Dans ce cas, la fraude concerne 9 500 salariés. En appliquant aux salariés non déclarés, le montant de la rémunération mensuelle moyenne dissimulée, on obtient 3,9 millions d'euros.

Nbre de salariés HCR	Taux de fraude	Extrapolation du nombre d'indiv. en fraude	Rémunération dissimulée / indiv. détecté	Extrapolation de la dissimulation de rémunération mensuelle (en millions)	Extrapolation des cotisations mensuelles éludées (45%) (en millions)
49 510	19%	9 407	920.0	8,6	3,9

Pour l'ensemble de l'année, les résultats sont alors les suivants :

Nbre de salariés	Cotisations éludées (en millions d'euros)	
	Annuelles - Borne inférieure -	Annuelles - Borne supérieure -
49 510	21	43

La dissimulation globale s'élève donc à un montant global compris entre 21 et 43 millions d'euros de cotisations éludée, soit entre 6.7% et 13.8% des cotisations annuelles dues au Régime général de Sécurité sociale.

4.2 Evaluation d'impact du contrôle sur les déclarations de masse salariale

- **Déterminants de la probabilité d'être contrôlé**

Dans l'équation de sélection (*tableaux 6*), nous avons choisi comme facteur déterminant du contrôle les codes APE, la taille de l'établissement (nombre de salarié en classes d'effectifs) et les zones d'emplois. La méthode de sélection sur inobservables requiert des variables instrumentales. Ces variables doivent déterminer la probabilité d'être contrôlé d'un établissement sans affecter la variable d'intérêt, ici égale à l'évolution des déclarations de masse salariale des cotisants. Nous avons choisis comme variables instrumentales les zones d'emploi.

En effet, les zones d'emploi situées dans le périmètre géographique de nos opérations de contrôle ont des dynamiques de croissance de leur masse salariale identiques (Acosse Stat n°36). Le fait d'appartenir à l'une de ces zones plutôt qu'à une autre n'est donc pas discriminant pour expliquer le taux de croissance de la masse salariale d'un établissement. Cela affecte en revanche la probabilité d'être contrôlé dans la mesure où chaque zone d'emploi a pu faire l'objet de différences dans les modalités de contrôle (du fait par exemple de l'intervention d'équipes d'inspecteurs différentes,...).

Tableau 6 : estimation des paramètres d'intérêt (MCO et méthode de sélection sur inobservables)

Libellés des Variables :	Coefficients estimés ⁽¹⁾	Ecart-types estimés
Constante	-1,3285 ***	0,0801
551A Hôtels touristiques avec restaurant	-0,1752 *	0,0964
551C Hôtels touristiques sans restaurant	-0,0854 ns	0,1338
552A Auberges de jeunesse et refuges	-0,1419 ns	0,6077
552C Exploitation de terrains de camping	-0,2256 *	0,1175
552E Autre hébergement touristique	-1,6139 ***	0,3614
553A Restauration de type traditionnel	0,0297 ns	0,0820
553B Restauration de type rapide	0,1324 ns	0,0920
554A Cafés tabacs	0,0394 ns	0,1485
554B Débits de Boisson	Référence	
554C Discothèques	-0,3410 ns	0,2473
Dummy : Effectif trimestre 3 2004 égal à 0	0,0039 ns	0,2158
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle]0 ; 2]	Référence	
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle]4 ; 10]	0,1055 *	0,0542
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle]10 ;43]	-0,2638 ***	0,0862
7291 Sud des Landes et 7292 Bayonne Pyrénées	0,6223 ***	0,0561
9351 Toulon	Référence	
9352 Fréjus Saint Raphael	0,2345 ***	0,0655
9353 Draguignan	0,7852 ***	0,0926
9354 Brignoles	0,0666 ns	0,1211

(1) Pour rappel, les coefficients du modèle probit ne sont pas directement interprétables. Les niveaux de significativité statistique sont indiqués par *** (1%), ** (5%), * (10%) et ns (non significatif).

Toutes choses égales par ailleurs, ce sont les établissements du secteur des “autres hébergements toutistiques” qui avaient une probabilité plus faible d’être contrôlés. De même, le fait d’appartenir au secteur des “hôtels touristiques avec restaurant” ou encore à celui des “exploitations de terrains de camping” diminue sensiblement la probabilité d’être contrôlé. Par ailleurs, la taille de l’établissement est une variable discriminante, le choix des inspecteurs de contrôle ayant porté de fait sur des établissements de petite taille. Plus l’établissement est de taille importante, moins sa probabilité d’être audité est élevée. Le fait d’être un établissement HCR de plus de 10 salariés a même un impact négatif sur la probabilité d’être contrôlé.

Enfin, nos variables instrumentales (les zones d’emploi) sont très significatives. Sur le périmètre méditerranéen, c’est dans la zone d’emploi de Draguignan que les contrôles ont été les plus massifs. Le fait d’appartenir à cette zone plutôt qu’à la zone d’emploi de Toulon accroît la probabilité d’être contrôlé.

- **Impact du contrôle sur les déclarations volontaires et spontanées de masse salariale**

La méthode de sélection sur inobservables nous permet de tester l'existence d'un biais lié à des variables inobservables. Dans notre cadre d'étude de l'impact des contrôles aléatoires saisonniers de lutte contre le travail dissimulé, il s'agit de tester l'existence d'un éventuel biais de sélection. En effet, même si les audits ont été réalisés de manière aléatoire, il est probable qu'un certain nombre de facteurs non observables aient influencé le choix des établissements à contrôler. Le mode d'organisation des contrôles (mode de quadrillage géographique, organisation des équipes) peut avoir influencé le choix des établissements. Enfin, les inspecteurs ont été sensibilisés à l'importance du fait de procéder à la sélection des établissements "au hasard" mais aucune règle statistique précise de sélection aléatoire n'a été imposée aux inspecteurs pour garantir le caractère aléatoire de l'opération. La subjectivité des contrôleurs peut ainsi avoir influencé le choix des établissements à contrôler. La mise en oeuvre de la méthode de sélection sur inobservables nous permet de vérifier si ce biais existe ou non.

A partir de l'estimation de l'effet moyen du contrôle sur les établissements contrôlés (*tableau 7*), qui prend en compte un éventuel biais de sélection, on observe que l'impact du contrôle pour ces établissements a été d'accélérer la croissance de leur masse salariale de l'ordre de 3 à 4%. De plus, si le contrôle était généralisé à l'ensemble des établissements HCR des deux zones géographiques, l'impact serait de l'ordre de 4 à 7 % (effet moyen du contrôle). Enfin, l'effet moyen du contrôle sur les non contrôlés est également important. Si ces établissements étaient passés par le programme de contrôle, ils auraient connu une croissance de leur masse salariale plus dynamique.

Tableau 7 : estimation des paramètres d'intérêt (MCO et méthode de sélection sur inobservables

	Estimation des effets	Zones		
		Méditerranéenne et Atlantique	Zone Méditerranéenne	Zone Atlantique
Sans correction du biais de sélection [Estimation par MCO]	Effet Naïf du Contrôle	3,05%	2,67%	3,50%
	Effet Moyen du Contrôle	5,43%	4,53%	6,92%
	Effet Moyen du Contrôle sur les Contrôlés	3,33%	3,17%	3,59%
	Effet Moyen du Contrôle sur les Non Contrôlés	5,78%	4,76%	7,50%
	Effet Moyen du Contrôle	5,43%	4,54%	6,92%
Avec correction du biais de sélection	Effet Moyen du Contrôle sur les Contrôlés	3,31%	3,15%	3,57%
	Effet Moyen du Contrôle sur les Non Contrôlés	5,79%	4,76%	7,51%
	Biais	-0,26%	-0,47%	-0,08%

Modèle d'intérêt pour les établissements Contrôlés :

Libellés des Variables :	Coefficients estimés	Ecart-types estimés :
Constante	0,2819 ns	0,2607
551A Hôtels touristiques avec restaurant	-0,0090 ns	0,0924
551C Hôtels touristiques sans restaurant	0,0790 ns	0,0948
552A Auberges de jeunesse et refuges	0,3269 ***	0,1102
552C Exploitation de terrains de camping	-0,0765 ns	0,1159
552E Autre hébergement touristique	0,6273 ***	0,2249
553A Restauration de type traditionnel	0,0365 ns	0,0814
553B Restauration de type rapide	0,0771 ns	0,0870
554A Cafés tabacs	0,0209 ns	0,1323
554B Débits de Boisson	Référence	
554C Discothèques	0,0471 ns	0,1240
Etablissements de la zone géographique : Méditerranéenne	Référence	
Etablissements de la zone géographique : Atlantique	0,0595 ns	0,0696

Dummy : Effectif trimestre 3 2004 égal à 0	0,0466 ns	0,3402
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle] 0 ; 2]	Référence	
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle] 4 ; 10]	0,1898 ***	0,0467
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle] 10 ; 43]	0,3252 ***	0,0905
Log [1 + Effectifs trimestre 3 2003]	-0,2351 ***	0,0464

Modèle d'intérêt pour les établissements Non Contrôlés :

<u>Libellés des Variables :</u>	<u>Coefficients estimés</u>	<u>Ecart-types estimés :</u>
Constante	0,2073 ns	0,0341
551A Hôtels touristiques avec restaurant	-0,0463 ns	0,0318
551C Hôtels touristiques sans restaurant	-0,0248 ns	0,0393
552A Auberges de jeunesse et refuges	-0,1392 ns	0,1321
552C Exploitation de terrains de camping	0,0316 ns	0,0361
552E Autre hébergement touristique	-0,0046 ns	0,0341
553A Restauration de type traditionnel	-0,0174 ns	0,0292
553B Restauration de type rapide	0,0119 ns	0,0330
554A Cafés tabacs	-0,0047 ns	0,0478
554B Débits de Boisson	Référence	
554C Discothèques	-0,0960 ns	0,0668
Etablissements de la zone géographique : Méditerranéenne	Référence	
Etablissements de la zone géographique : Atlantique	0,0469 ***	0,0149
Dummy : Effectif trimestre 3 2004 égal à 0	-0,4425 ***	0,1379
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle] 0 ; 2]	Référence	
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle] 4 ; 10]	0,1796 ***	0,0190
Dummy : Effectif trimestre 3 2004 compris dans l'intervalle] 10 ; 43]	0,3112 ***	0,0305
Log [1 + Effectifs trimestre 3 2003]	-0,1682 ***	0,0165

Les variables explicatives du modèle d'intérêts sont les codes APE INSEE, la taille de l'établissement en terme de nombre de salariés en 2004 (en classes d'effectifs), la taille de l'établissement en terme de nombre de salariés en 2003 et les zones géographiques.

Les niveaux de significativité statistique sont indiqués par *** (1%), ** (5%), * (10%) et ns (non significatif). Les résultats montrent qu'il n'y a pas de biais lié à la sélection. En effet, les

paramètres ρ_0 et ρ_1 intervenant dans l'expression du biais sont nuls (cf annexe). Le paramètre ρ_0 rend compte de la corrélation entre le choix de l'établissement par le contrôleur et la croissance de la masse salariale dans un établissement non contrôlé. ρ_1 rend compte de la corrélation entre le choix de l'établissement par le contrôleur et la croissance de la masse salariale dans un établissement contrôlé

Ecart-types et coefficients de corrélations :

<u>Libellés des écart-types et coefficients de corrélation :</u>	<u>Coefficients estimés :</u>
Coefficient de corrélation entre U_0 et v : ρ_0	0,0002
Coefficient de corrélation entre U_1 et v : ρ_1	-0,0001
Ecart-Type de U_0 : σ_0	0,4591
Ecart-Type de U_1 : σ_1	0,4934

5. Conclusion

Nous avons réalisé deux évaluations complémentaires à partir des contrôles aléatoires menés sur le secteur hôtel, café, restaurant (HCR) des zones géographiques Méditerranéenne et Atlantique. La première évaluation consistait à chiffrer l'assiette dissimulée de cotisations. La seconde consistait à mesurer l'impact des contrôles aléatoires sur les déclarations volontaires et spontanées de masse salariale.

L'assiette dissimulée sur l'ensemble de l'année représente 7 à 14 % de l'ensemble des cotisations du secteur HCR dans les zones concernées. L'impact des contrôles aléatoires sur les déclarations volontaires et spontanées de masse salariale des établissements contrôlés est de l'ordre de 3 à 4 %. Si le contrôle était généralisé à l'ensemble des établissements HCR des zones géographiques, l'impact serait de l'ordre de 4 à 7 %.

Les évaluations réalisées confirment l'absence de biais de sélection. Les inspecteurs d'URSSAF ont respecté la méthodologie de contrôle au hasard. Ces résultats sont les premiers éléments d'évaluation de l'impact des contrôles aléatoires. Ils constituent un minorant de l'impact des contrôles. En effet, une perspective de recherche est de prendre en compte l'éventuel effet de diffusion des contrôles aléatoires c'est-à-dire de mesurer l'accroissement de la masse salariale issu de la « peur du gendarme » des établissements non contrôlés. Par ailleurs, le constat de l'absence de biais de sélection valide l'extrapolation de l'assiette dissimulée à l'ensemble du secteur HCR des zones géographiques.

Il reste que nos deux évaluations comportent un biais de sous-estimation lié à la difficulté à détecter la fraude. Réalisant ce même constat, les services fiscaux américains (IRS, 1996) ont d'ailleurs développé plusieurs multiplicateurs permettant d'accroître le volume de la fraude détectée dans le calcul des impôts impayés. La valeur moyenne de ces multiplicateurs est approximativement de deux, mais elle varie considérablement selon les composantes fiscales envisagées. Dans le cas de données collectées par contrôles et inspections, un moyen de résoudre ce problème est d'analyser les données en procédant à une estimation de la détection contrôlée.⁷

Enfin, l'Acoss et les Urssaf ont décidé de reconduire ces contrôles aléatoires en 2005 mais cette fois sur l'ensemble de la France. L'exploitation des informations issues de ces opérations de contrôles aléatoires nous permettra d'affiner nos évaluations de l'assiette dissimulée et de l'impact des contrôles.

⁷ - Cette procédure est décrite en détail dans Feinstein (1999).

Références bibliographiques

Brodaty T., Crépon B. et Fougère D. (2004), « Les méthodes micro-économétriques d'évaluation et leurs applications aux politiques actives de l'emploi », *Economie et Prévision*, à paraître.

Cavaco S. et Cornu Pauchet M. (2005), « L'effet des allègements de la loi Aubry 2 sur l'emploi », Miméo pour le colloque de l'AFSE 2005.

Cornu Pauchet M. (2003). "L'évaluation de l'économie souterraine, un recensement des études", *Acoss Stat* n°8

Feinstein (1999), « Approaches for Estimating Noncompliance : Examples from Federal Taxation in the United States », *The Economic Journal*, 109, 360-369.

Internal Revenue Service (1996), « Federal Tax Compliance Research : Individual Income Tax Gap Estimates for 1985, 1988, and 1992 », publication 1415, Washington D.C.

Qi Li (1996), "Nonparametric testing of closeness between two unknown distribution functions", *Econometric Reviews*, 15(3), 261-274.

Rubin, D. (1974): "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies", *Journal of Educational Psychology*, 66, pp. 688-701.

Schneider et Enste (2000) "Shadows Economies : Size, causes and Consequences", *Journal of economic Literature*, Vol XXXVIII.

Annexes

1. Les paramètres d'intérêt

L'effet moyen du « contrôle sur les contrôlés » est représenté ici par la variation moyenne de la masse salariale induite par le contrôle pour les établissements visités. Compte tenu des hypothèses précédentes sur les équations de passage par le contrôle et d'évolution de la masse salariale, cet effet se réécrit :

$$\Delta_{TT}(X) = E(Y_1 - Y_0 | X, D = 1) = E[(c_1 - c_0) + (X\beta_1 - X\beta_0) | X, D = 1] + E(U_1 - U_0 | X, D = 1)$$

$$\Leftrightarrow \Delta_{TT}(X) = E \left[(c_1 - c_0) + (X\beta_1 - X\beta_0) + (\rho_1 \cdot \sigma_1 - \rho_0 \cdot \sigma_0) \cdot \frac{\phi(Z\gamma)}{\Phi(Z\gamma)} \right]$$

où ϕ est la fonction de densité de la loi normale standard et Φ est sa fonction de répartition. Le dernier terme de cette équation représente le biais lié à la corrélation entre le résidu de l'équation de passage par le contrôle et le résidu de l'équation explicative de l'évolution de la masse salariale (termes ρ_0 et ρ_1). De même, l'effet du contrôle sur les non contrôlés, qui mesure la variation de masse salariale qui aurait été induite par le contrôle pour les établissements non contrôlés s'ils avaient été contrôlés, s'écrit :

$$\Delta_{TNT}(X) = E(\Delta | X, D = 0) = E[(c_1 - c_0) + (X\beta_1 - X\beta_0) | X, D = 0] + E(U_1 - U_0 | X, D = 0)$$

$$\Leftrightarrow \Delta_{TNT}(X) = E \left[(c_1 - c_0) + (X\beta_1 - X\beta_0) - (\rho_1 \cdot \sigma_1 - \rho_0 \cdot \sigma_0) \cdot \frac{\phi(Z\gamma)}{1 - \Phi(Z\gamma)} \right]$$

2. Le biais de sélectivité

L'effet moyen du contrôle sur les contrôlés « Δ_{TT} » et l'effet moyen du contrôle « Δ_M » ne sont toutefois égaux que sous certaines hypothèses techniques restrictives. Si les deux variables de résultat sont indépendantes de l'affectation au Contrôle $\Leftrightarrow Y_0, Y_1 \perp D$, on peut écrire :

$$\Delta_M = E(Y_1 - Y_0) = E(Y_1) - E(Y_0) = E(Y_1 | D = 1) - E(Y_0 | D = 0) = E(Y | D = 1) - E(Y | D = 0)$$

$$\Delta_{TT} = E(Y_1 - Y_0 | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 1) = E(Y_1 | D = 1) - E(Y_0 | D = 0) = E(Y | D = 1) - E(Y | D = 0)$$

Dans ce cas, les deux paramètres sont égaux et sont identifiés par le calcul de la différence des Moyennes. Ces deux paramètres d'intérêts sont égaux à l'Effet Naïf. Pour l'Effet Moyen du Contrôle sur les Contrôlés « Δ_{TT} », la condition $Y_0 \perp D$ est suffisante.

Si on supprime la condition d'indépendance suivante : $Y_0 \perp D$, on obtient :

$$\begin{aligned} E(Y|D=1) - E(Y|D=0) &= E(Y_1|D=1) - E(Y_0|D=0) \\ E(Y|D=1) - E(Y|D=0) &= E(Y_1|D=1) - E(Y_0|D=1) + E(Y_0|D=1) - E(Y_0|D=0) \\ E(Y|D=1) - E(Y|D=0) &= \underbrace{E(Y_1 - Y_0 | D=1)}_{\Delta_{TT}} + \underbrace{E(Y_0 | D=1) - E(Y_0 | D=0)}_{\text{Biais}} \end{aligned}$$

$$\text{Biais} = E(Y_0|D=1) - E(Y_0|D=0)$$

Ce biais trouve son origine dans le fait que la situation moyenne des établissements Contrôlés n'aurait pas été la même en cas d'absence de Contrôle que celle des établissements Non Contrôlés.

$$\begin{aligned} B(X) &= E(Y_0|X, D=1) - E(Y_0|X, D=0) \\ B(X) &= [E(c_0 + X\beta_0|X, D=1) + E(U_0|X, D=1)] - [E(c_0 + X\beta_0|X, D=0) + E(U_0|X, D=0)] \end{aligned}$$

$$\begin{aligned} B(X) &= E(Y_0|X, D=1) - E(Y_0|X, D=0) \\ B(X) &= E \left[c_0 + X\beta_0 + \rho_0 \cdot \sigma_0 \cdot \frac{\phi(Z\gamma)}{\Phi(Z\gamma)} \right] - E \left[c_0 + X\beta_0 - \rho_0 \cdot \sigma_0 \cdot \frac{\phi(Z\gamma)}{1 - \Phi(Z\gamma)} \right] \end{aligned}$$

3. Vraisemblance du modèle et technique d'estimation :

La vraisemblance du modèle s'écrit :

$$L(Y, D|X, Z) = \prod_{j=0,1} \prod_{\{D_i=j\}} \Pr(D_i = j | Y_{j,i}, X_i, Z_i) \times f(Y_{j,i} | X_i)$$

$$(Y_{i,j} | X_i) \sim N[X_i \beta_j + c_j, \sigma_j^2]$$

$$\text{On a donc que } f(Y_{i,j}) = \frac{1}{\sigma_j} \cdot \phi \left[\frac{Y_i - (X_i \beta_j + c_j)}{\sigma_j} \right]$$

Calcul de la probabilité d'être Contrôlé : $\Pr(D_i = j | Y_{j,i}, X_i, Z_i)$

Rappel sur l'hypothèse de normalité du modèle :

$$\begin{pmatrix} U_{j,i} \\ v_i \end{pmatrix} \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_j^2 & \rho_j \cdot \sigma_j \\ \rho_j \cdot \sigma_j & 1 \end{pmatrix} \right] \quad (v_i | U_{i,j} = u) \sim N \left[\frac{\rho_j}{\sigma_j} \cdot u, (1 - \rho_j^2) \right]$$

Probabilité d'être Contrôlé

$$\Pr(D_i = 1 | Y_{1,i}, X_i, Z_i) = \Pr(v_i \geq -Z_i \cdot \gamma | U_{1,i} = Y_{1,i} - c_1 - X_i \cdot \beta_1, X_i, Z_i)$$

$$\Pr(D_i = 1 | Y_{1,i}, X_i, Z_i) = 1 - \Pr(v_i \leq -Z_i \cdot \gamma | U_{1,i} = Y_{1,i} - c_1 - X_i \cdot \beta_1, X_i, Z_i)$$

$$\Pr(D_i = 1 | Y_{1,i}, X_i, Z_i) = 1 - \Phi \left[\frac{-Z_i \cdot \gamma - \frac{\rho_1}{\sigma_1} (Y_{1,i} - c_1 - X_i \cdot \beta_1)}{\sqrt{1 - \rho_1^2}} \right]$$

$$\Leftrightarrow \Pr(D_i = 1 | Y_{1,i}, X_i, Z_i) = \Phi \left[\frac{Z_i \cdot \gamma + \frac{\rho_1}{\sigma_1} (Y_{1,i} - c_1 - X_i \cdot \beta_1)}{\sqrt{1 - \rho_1^2}} \right]$$

Probabilité de ne pas être Contrôlé :

$$\Pr(D_i = 0 | Y_{0,i}, X_i, Z_i) = \Pr(v_i \leq -Z_i \cdot \gamma | U_{0,i} = Y_{0,i} - c_0 - X_i \cdot \beta_0, X_i, Z_i)$$

$$\Leftrightarrow \Pr(D_i = 0 | Y_{0,i}, X_i, Z_i) = 1 - \Phi \left[\frac{Z_i \cdot \gamma + \frac{\rho_0}{\sigma_0} (Y_{0,i} - c_0 - X_i \cdot \beta_0)}{\sqrt{1 - \rho_0^2}} \right]$$

4. Technique d'estimation :

Le modèle d'évaluation est estimé par maximum de vraisemblance. Le logiciel utilisé pour réaliser les estimations est le logiciel SAS. La procédure pour maximiser la vraisemblance est la procédure NLP (Non Linear Programming). La technique d'optimisation choisie est la méthode du Quasi Newton - BFGS

