

Direction des Études et Synthèses Économiques

G 2009 / 06

Faut-il pondérer ?... Ou l'éternelle
question de l'économètre confronté
à des données d'enquête

Laurent DAVEZIES - Xavier D'HAULTFOEUILLE

Document de travail



Institut National de la Statistique et des Études Économiques

INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

*Série des documents de travail
de la Direction des Études et Synthèses Économiques*

G 2009 / 06

Faut-il pondérer ?... Ou l'éternelle question de l'économètre confronté à des données d'enquête

Laurent DAVEZIES*
Xavier D'HAULTFOEUILLE**

JUILLET 2009

* Département des Études Économiques d'Ensemble - Division « Redistribution et Politiques Sociales »
Timbre G210 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX

** Département des Études Économiques d'Ensemble - Division « Marchés et Stratégies d'Entreprise »
Timbre G230 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX

Faut-il pondérer ?... Ou l'éternelle question de l'économètre confronté à des données d'enquête

Résumé

Un problème récurrent des travaux d'étude sur données d'enquête est de savoir quel usage doit être fait des pondérations associées aux observations individuelles. Ce texte précise dans quels cas ces pondérations sont à prendre en compte. Nous montrons que les estimateurs pondérés sont plus robustes que les estimateurs non pondérés, même s'ils sont moins précis lorsque ces derniers sont valides. Dans certains cas, la comparaison des deux estimateurs permet de tester la compatibilité entre les hypothèses concernant le mécanisme de sélection des observations (i.e., le plan de sondage et la non-réponse) et celles portant sur le modèle économétrique considéré. Enfin, quelques méthodes de calcul de précision des estimateurs pondérés sont présentées.

Mots-clés : poids de sondage, sélection, estimateurs pondérés, test d'Hausman

To Weight or not to Weight? The Eternal Question of Econometricians facing Survey Data

Abstract

A recurrent problem in economic studies based on survey data is to determine whether survey weights should be used or not. We show that weighted estimators are more robust than unweighted ones, but less accurate when unweighted estimators are consistent. In some circumstances, a compatibility test between the assumptions underlying the selection mechanism (i.e., the survey design and nonresponse) and the econometric model at hand can be implemented, by comparing these two estimators. Finally, we present some methods to compute the accuracy of weighted estimators.

Keywords: survey weights, sample selection, inverse probability weighting, Hausman test

Classification JEL : C13, C21, C83

1 Introduction

La situation du chargé d'études utilisant des données issues d'une enquête est quelque peu schizophrène. D'un côté, il est prêt à estimer les modèles économétriques qu'on lui a enseignés, où toutes les observations se valent puisque l'échantillonnage est supposé i.i.d¹. De l'autre, il ne peut ignorer que ces données sont issues d'un sondage, et qu'à ce titre, il devrait utiliser un estimateur basé sur des poids de sondage. Comment réconcilier ces deux approches, alors que leur formalisme semble incompatible ?

L'objet de cette note est de montrer que même si l'on reste dans le cadre pratique de données i.i.d., les variables de poids sont en général utiles voire indispensables. Ainsi, nous montrons qu'en général, les estimateurs pondérés sont plus robustes que les estimateurs non pondérés, dans le sens qu'ils convergent sous des hypothèses moins restrictives. La contrepartie, comme toujours en statistique, est une perte de précision. La note met également en exergue l'importance des variables utilisées dans le calage et éventuellement dans le plan de sondage lorsque le tirage de l'échantillon est à probabilités inégales. Ceci signifie deux choses : d'une part l'économètre se doit de connaître ces variables², et de l'autre, le choix des variables utilisées dans le redressement a une influence sur les outils économétriques utilisables par la suite.

Cette note s'articule comme suit. La deuxième partie rappelle la définition des pondérations en théorie des sondages, et définit le cadre d'analyse retenu ici. La troisième présente les résultats de convergence des estimateurs pondérés et non pondérés, suivant les hypothèses retenues sur le mécanisme de sélection et la nature du paramètre estimé. La quatrième partie indique comment calculer la précision d'estimateurs pondérés et les pièges à éviter dans ce cadre. Pour conclure, une feuille de route résume succinctement les différentes étapes de la démarche présentée. Les détails techniques des résultats sont renvoyés en annexe.

1. La question de l'utilisation de poids est peu abordée dans la littérature économétrique. Le papier se rapprochant le plus de la note ci-présente est celui de Wooldridge (2007). Voir également Lerman et Manski (1977) ou Imbens (1992) pour un traitement des "choice based samples", qui sont abordés dans l'exemple 6 ci-dessous.

2. En général, ces informations sont disponibles dans la documentation de l'enquête. Dans le cas contraire, on pourra retrouver ces informations à partir de notes du service producteur.

2 Définition des pondérations et cadre retenu

2.1 Les pondérations en sondage

Dans le cadre standard de sondages, on considère une population notée $U = \{1, \dots, N\}$ dans laquelle on tire aléatoirement un échantillon $\mathcal{S} \subset U$ ³. En l'absence de non-réponse, la variable usuelle de pondération de l'individu k est définie par

$$w_k = \frac{1}{P(k \in \mathcal{S})}.$$

Ces poids sont connus puisqu'ils sont définis au moment du tirage de l'échantillon. Ils permettent de calculer l'estimateur de Horvitz-Thompson \hat{t} d'un total $t_y = \sum_{k \in U} y_k$ d'une variable y :

$$\hat{t} = \sum_{k \in \mathcal{S}} w_k y_k.$$

Rappelons que l'intérêt d'un tel estimateur est d'être sans biais, et ce quelles que soient les valeurs de $(y_k)_{k \in U}$. En pratique, une partie de l'échantillon est non-répondante. On note alors $\mathcal{R} \subset \mathcal{S}$ l'échantillon des répondants. Dans ce cas la variable de poids à utiliser pour calculer l'estimateur de Horvitz-Tompson serait $1/P(k \in \mathcal{R})$. Cependant, cette probabilité est inconnue puisqu'on ne connaît pas la probabilité de réponse des individus. La procédure habituelle consiste alors à estimer les poids par un modèle de non réponse et/ou par calage⁴. On a alors :

$$w_k = \frac{1}{\hat{P}(k \in \mathcal{R})}.$$

2.2 Le modèle adopté ici

Pour relier ces variables de poids aux modèles économétriques, on adopte dans la suite un formalisme un peu différent. Plus précisément, on considère un modèle de "superpopulation" où les valeurs individuelles sur la population U sont des réalisations de variables i.i.d. Par exemple, (y_1, \dots, y_N) sont les réalisations des variables aléatoires i.i.d. (Y_1, \dots, Y_N) , de même loi que Y . On suppose que l'échantillon des répondants \mathcal{R} est défini par $\mathcal{R} = \{i \in U / D_i =$

3. La présentation donnée ici est très succincte, pour davantage de détails sur le formalisme de la théorie des sondages, on se référera par exemple à Tillé (2001).

4. Soit \tilde{x}_{2k} des variables auxiliaires dont les totaux $t_{\tilde{x}_2}$ sur la population entière sont connus. Le calage a pour but de déterminer les poids w_k les plus proches possibles de $1/P(k \in \mathcal{S})$ tels que les estimateurs des totaux $\sum_{k \in \mathcal{S}} w_k \tilde{x}_{2k}$ soient exacts, i.e. égaux aux totaux $t_{\tilde{x}_2}$. Pour plus de détails, on se référera à Deville et Sarndal (1992) ou Sautory (1993) sur le calage, et à Deville (2002) sur le calage généralisé.

1}, où (D_1, \dots, D_N) sont également i.i.d., de même loi que D . On a $D = S \times R$ où S est l'indicatrice d'appartenance à l'échantillon initial et R est l'indicatrice de réponse à l'enquête. S est supposée dépendre de variables \tilde{X}_1 ⁵. On note \tilde{X}_2 les variables expliquant R , c'est-à-dire les variables ayant été incluses dans le modèle de non-réponse et/ou le calage. Soit alors $\tilde{X} = (\tilde{X}_1, \tilde{X}_2)$ l'ensemble des variables qui influent sur D . On considérera que la variable de poids W est alors définie par

$$W = \frac{1}{\widehat{P}(D = 1|\tilde{X})}.$$

Deux remarques s'imposent à ce stade. Premièrement, le modèle d'échantillonnage que l'on considère ici correspond, en théorie des sondages, à un plan poissonnien, c'est-à-dire à un plan où chaque individu est tiré indépendamment des autres. Cette hypothèse permet de s'assurer que l'on reste dans un cadre i.i.d. standard en économétrie. Elle est cependant peu réaliste pour un certain nombre de sondages, en particulier les sondages aréolaires. Notons que si cette hypothèse n'est pas vérifiée, les estimateurs de variance proposés en section 4 ne seront pas convergents ; en revanche la convergence des estimateurs présentés en section 3 ne sera pas affectée. Deuxièmement, on supposera par la suite que W tend vers $1/P(D = 1|\tilde{X})$. Cette hypothèse n'est pas anodine car la plupart du temps, l'estimation de cette probabilité repose sur une forme paramétrique (comme un logit ou un probit dans un modèle de non-réponse). Une mauvaise spécification conduit en général à des estimateurs non-convergents.

3 Pondérer ou non, telle est la question.

On s'intéresse maintenant à l'estimation d'un paramètre θ dépendant potentiellement d'une variable expliquée Y et de variables explicatives X , avec a priori $X \neq \tilde{X}$. Pour toutes variables aléatoires (U, V) , on note F_U (resp. $F_{U|V}$) la loi de U (resp. la loi conditionnelle de U sachant V) et $U \perp\!\!\!\perp V$ pour " U est indépendante de V ". Le paramètre θ est donc fonction de $F_{X,Y}$.

3.1 Les variables \tilde{X} sont les facteurs pertinents de la sélection

On suppose tout d'abord que les concepteurs d'enquêtes ont réussi à capter, à travers \tilde{X} , les facteurs pertinents de la sélection de l'échantillon. Cette hypothèse se traduit formellement

5. Ainsi, \tilde{X}_1 inclut par exemple l'effectif en tranches dans les enquêtes annuelles d'entreprises de l'INSEE.

par :

H0. $(Y, X) \perp\!\!\!\perp D \mid \tilde{X}$.

Considérons par exemple une enquête ménages à probabilités de tirage égales et où la non-réponse a été corrigée par calage sur le type de ménage (variable TYPMEN) et l'âge de la personne de référence (variable AGEPR). Dans ce cas on a $\tilde{X}_1 = \emptyset$ et $\tilde{X}_2 = (\text{TYPMEN}, \text{AGEPR})$. Si l'on s'intéresse à un modèle où le salaire Y de la personne de référence est expliqué par le nombre d'années d'études X , l'hypothèse H0 stipule qu'à type de ménage et âge de la personne de référence fixés, les répondants et non-répondants de l'enquête ont la même distribution de salaires et du nombre d'années d'études. Ceci exclut par exemple que le salaire ait un effet "propre" (i.e. à âge et type de ménage fixé) sur la probabilité de réponse à l'enquête.

3.1.1 Un cas particulier : on peut pondérer, mais on ne doit pas.

Outre H0, on suppose vérifiée l'hypothèse suivante.

H1. θ dépend uniquement de $F_{Y|X}$ et $\tilde{X} \subset X$.

Résultat 1 : sous les hypothèses H0 et H1, les estimateurs seront convergents que l'on pondère ou non. L'estimateur non pondéré sera cependant plus précis.

La deuxième partie de l'hypothèse H1 suppose simplement que les variables \tilde{X} utilisées pour calculer les pondérations sont incluses dans l'estimation. Cela nécessite naturellement que l'économètre connaisse ces variables. Notons qu'il n'est pas toujours souhaitable d'inclure l'ensemble des \tilde{X} dans les explicatives du modèle. En effet, certaines composantes de \tilde{X} peuvent être endogènes pour le problème étudié. C'est le cas par exemple si \tilde{X} inclut la CSP et que l'on cherche à estimer une équation de salaire. Le cas où $\tilde{X} \not\subset X$ est traité dans la section suivante.

Illustrons maintenant, à travers des exemples, la première partie de l'hypothèse H1.

Exemple 1 : régression linéaire. On suppose ici

$$Y = X'\theta + \varepsilon, \quad E(\varepsilon|X) = 0 \tag{3.1}$$

Dans ce cas, θ dépend uniquement de $F_{Y|X}$ puisque

$$\theta = \frac{\partial E(Y|X = x)}{\partial x}.$$

Considérons maintenant une régression linéaire instrumentale, i.e. telle que $E(\varepsilon|Z) = 0$ ⁶. Alors sous la condition de rang que les différentes fonctions $(E(X_k|Z))_{k=1,\dots,K}$ sont linéairement indépendantes (où X_k désigne la k -ième composante de X), θ est défini par l'équation

$$E(Y|Z) = E(X|Z)'\theta.$$

Ainsi, θ ne dépend que de $F_{Y,X|Z}$. Lorsque $\tilde{X} \subset Z$, les estimateurs pondérés et non pondérés seront alors convergents, et l'estimateur non pondéré sera plus précis.

Exemple 2 : modèle logit / probit. Ici

$$Y = \mathbf{1}\{X'\theta + \varepsilon \geq 0\}$$

où $-\varepsilon \perp\!\!\!\perp X$, de fonction de répartition F . Dans ce cas,

$$\theta = \frac{\partial}{\partial x} [F^{-1}(E(Y|X = x))].$$

Plus généralement que le modèle logit / probit, considérons un modèle conditionnel paramétrique où $F_{Y|X}$ dépend d'un paramètre η , i.e. s'écrit $F_{Y|X,\eta}$. Si $\theta = f(\eta)$, alors on peut montrer⁷ que θ ne dépend que de $F_{Y|X}$.

Exemple 3 : régression non paramétrique. On s'intéresse ici à la fonction $\theta(x) = E(Y|X = x)$. $\theta(\cdot)$ dépend évidemment uniquement de $F_{Y|X}$.

3.1.2 Le cas général : on doit pondérer.

On considère maintenant les situations où H1 n'est plus satisfaite, H0 restant vérifiée.

Résultat 2 : Supposons que H0 soit satisfaite, mais pas H1. Dans ce cas, les estimateurs pondérés seront convergents, mais pas les estimateurs non pondérés en général.

L'intuition, dans ce cas, est que θ dépendra de la distribution des X . Or la distribution de X dans l'échantillon ne "correspond" pas à celle de la population (par exemple, les diplômés du supérieur sont sous représentés dans l'échantillon des répondants car leur probabilité de réponse est plus faible). L'estimateur non pondéré sera, de ce fait, non convergent en général.

6. Notons que l'hypothèse H0 s'écrit ici $(Y, X, Z) \perp\!\!\!\perp D|\tilde{X}$.

7. Pour cela, il faut également supposer que le modèle est identifiable, i.e., $F_{Y|X,\eta} = F_{Y|X,\eta'}$ implique $\eta = \eta'$.

Exemple 4 : statistiques simples. On s'intéresse par exemple à $\theta = E(Y)$. Ce paramètre ne dépend pas uniquement de $F_{Y|X}$, mais aussi de la distribution des X (puisque $E(Y) = E(E(Y|X))$). Ici, l'estimateur pondéré n'est autre que l'estimateur de Horvitz-Thompson. Ceci vaut en fait pour toute fonction de la distribution de Y telle que la variance, les quantiles...

Exemple 1 (suite). Revenons sur le modèle linéaire, mais avec une hypothèse plus faible que (3.1) :

$$Y = X'\theta + \varepsilon, \quad E(X\varepsilon) = 0 \quad (3.2)$$

L'hypothèse $E(\varepsilon|X) = 0$ revient à supposer que la meilleure approximation en X de Y est linéaire et vaut $X'\theta$, alors que $E(X'\varepsilon) = 0$ revient juste à supposer que $X'\theta$ est la meilleure approximation linéaire en X de Y . Cette dernière hypothèse est donc beaucoup plus faible⁸. Dans ce cas θ dépend en général de la loi jointe de (X, Y) et non uniquement de la loi de $Y|X$. Ainsi, l'estimateur non pondéré sera non convergent en général. Ceci peut être illustré par l'exemple suivant. Considérons le modèle simple :

$$\ln(\text{salaire}) = a + b \times \text{années d'éducation} + \varepsilon$$

Supposons qu'en fait, les rendements de l'éducation soient décroissants⁹. Le paramètre b correspond alors à une sorte de rendement moyen sur l'ensemble de la population (cf. graphique 1)¹⁰. Supposons maintenant que la probabilité de répondre à l'enquête diminue avec le nombre d'années d'étude. Dans ce cas l'échantillon comprendra davantage de personnes peu éduquées, pour qui le rendement est élevé. Un estimateur non pondéré surestimerait alors en général b . L'estimateur pondéré, en revanche, accentuerait l'importance des quelques individus très éduqués présents dans l'enquête, et conduirait à un estimateur convergent.

Notons qu'il est également nécessaire de pondérer lorsque $E(\varepsilon|X) = 0$ mais $\tilde{X} \not\subset X$. Ceci vaut également pour la régression instrumentale décrite précédemment. Dès lors que $\tilde{X} \not\subset Z$, les deux étapes de cette régression doivent être pondérées.

Exemple 2 (suite). On s'intéresse, dans le cadre des modèles logit ou probit, à l'effet marginal moyen $\tilde{\theta}_k$ de la variable X_k (on note les autres variables X_{-k}). $\tilde{\theta}_k$ est alors défini

8. En particulier cette hypothèse n'est pas testable alors que $E(\varepsilon|X) = 0$ l'est.

9. Ce n'est pas tant la décroissance qui importe ici, mais le fait que les rendements sont non constants avec le nombre d'année d'éducation.

10. Ceci n'est pas tout à fait exact car en fait, $b \neq E(\partial E(Y|X)/\partial x)$ en général.

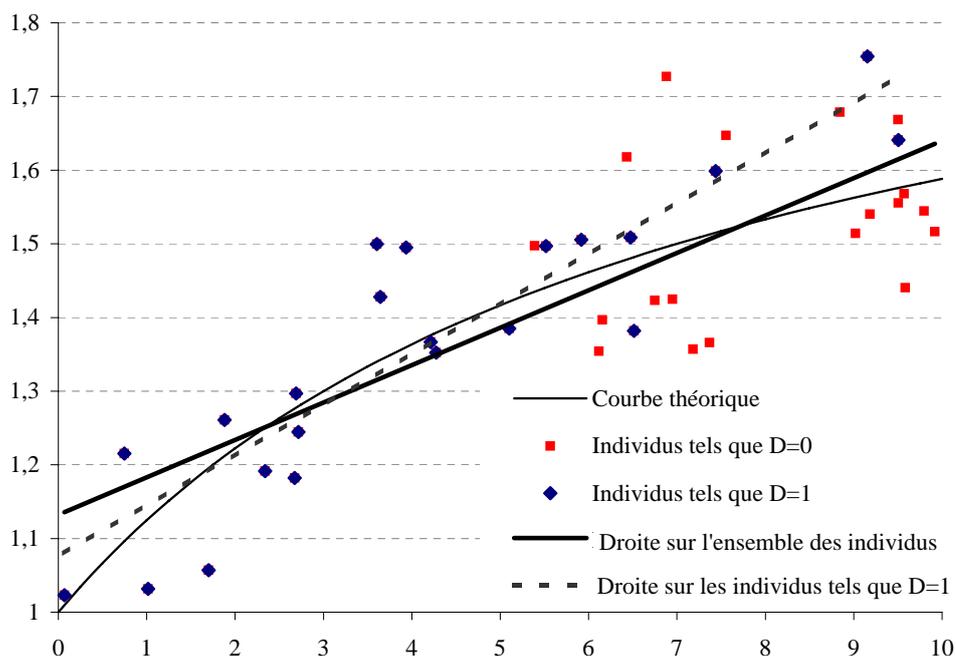


FIGURE 1: Exemple 1, estimation des rendements de l'éducation.

par

$$\tilde{\theta}_k = E \left[\frac{\partial E(Y|X)}{\partial x_k} \right].$$

que l'on peut réécrire

$$\tilde{\theta}_k = E [F'(X'\theta)] \theta_k$$

où θ_k est la k -ième composante de θ . Ainsi, $\tilde{\theta}_k$ dépend également de la loi de X , à travers le premier terme. Dans ce cas, on peut adopter l'approche hybride d'estimer θ sans poids, puis utiliser ceux-ci dans l'estimation de $E [F'(X'\theta)]$.

Exemple 5 : matching. Le même argument s'applique à l'estimation des "effets de traitement" par matching. Supposons que l'on veuille mesurer l'efficacité d'une politique publique. Soit $T = 1$ si l'individu est concerné par la politique (i.e., "traité"), $T = 0$ sinon. Supposons que l'on veuille mesurer (par exemple) l'effet moyen sur les traités

$$\Delta^{TT} = E(Y_1 - Y_0|T = 1)$$

où Y_1 (resp. Y_0) est la variable d'intérêt qu'aurait l'individu s'il était traité (resp. non traité). On observe seulement $Y = TY_1 + (1-T)Y_0$ mais on fait l'hypothèse d'indépendance conditionnelle :

$$Y_0 \perp\!\!\!\perp T|X.$$

Cette hypothèse est nécessaire pour que les estimateurs de matching soient convergents. On a, sous cette condition,

$$\Delta^{TT} = E(Y - E(Y|T = 0, X)|T = 1).$$

Le terme $E(Y|T = 0, X)$ correspond à l'exemple 3 de la régression non-paramétrique¹¹. Il peut donc être estimé sans les poids, par matching ou par un estimateur à noyau par exemple. En revanche, Δ^{TT} est une moyenne simple (de la variable $Y - E(Y|T = 0, X)$) qui correspond à l'exemple 4. Pour l'estimer de manière convergente, il faut donc pondérer.

Exemple 6 : “choice-based sample”. Lorsque la variable d'intérêt binaire Y intervient dans l'échantillonnage, on parle de “choice-based sample” (cf. Lerman et Manski, 1977). C'est le cas par exemple si l'on s'intéresse à la probabilité d'être cadre à partir d'un échantillon où les cadres sont surreprésentés¹². On a alors $Y \subset \tilde{X}$, et on ne peut dans ce cas avoir $\tilde{X} \subset X$ ¹³. L'estimateur non pondéré sera donc en général non convergent, contrairement à l'estimateur pondéré.

3.2 Certains facteurs de la sélection ont été omis de \tilde{X}

On peut identifier deux situations principales où l'hypothèse H0 n'est pas vérifiée :

1. Le concepteur d'enquête a été “négligent”, par exemple en redressant la non-réponse sur un nombre insuffisant de variables. Dès lors, \tilde{X} n'inclut pas tous les facteurs pertinents de la sélection. En revanche, les X considérés par l'économètre incluent bien l'ensemble de ces facteurs. Dans ce cas, on aura

H0'. $Y \perp\!\!\!\perp D|X$.

Lorsque $\tilde{X} \subset X$, cette hypothèse est moins forte que H0 puisque cette dernière implique H0' (cf. annexe). Lorsque $\tilde{X} \not\subset X$, en revanche, aucune des deux hypothèses n'est plus restrictive a priori, et seul le contexte peut permettre de trancher entre les deux.

2. Même conditionnellement à (X, \tilde{X}) , le processus de sélection est lié à Y . Il est possible, dans une enquête sur le patrimoine des ménages, que les ménages fortunés soient plus

11. Il s'agit en effet de la régression non-paramétrique sur la sous-population des individus tels que $T = 0$.

12. cf. l'enquête ECMOSS (Enquête sur le Coût de la Main d'Oeuvre et la Structure des Salaires) de l'INSEE.

13. En revanche, l'hypothèse H0 peut tout à fait rester valide.

récalcitrants à répondre toutes choses égales par ailleurs. Dans ce cas, ni H_0 ni H_0' ne seront satisfaites.

Considérons ces deux situations successivement.

3.2.1 Cas 1

Dans ce cas, une première solution est de reconstruire une nouvelle variable de poids $W' = 1/P(D = 1|X)$. L'hypothèse H_0 sera satisfaite si l'on remplace \tilde{X} par X et les résultats de la section précédente s'appliquent.

La construction de cette variable peut cependant s'avérer fastidieuse. Elle est même impossible en l'absence d'information auxiliaire (i.e., sur les non-répondants ou via une autre source comme des fichiers administratifs) sur X . Dans ce cas, on a le résultat suivant.

Résultat 3 : Sous l'hypothèse H_0' et si θ dépend uniquement de $F_{Y|X}$, un estimateur non pondéré de θ sera convergent et "efficace". En revanche, les paramètres dépendant de la loi jointe de (X, Y) ne pourront pas être estimés de manière convergente en général.

Par ailleurs, l'estimateur pondéré (en utilisant les poids W) sera convergent mais inefficace lorsque $\tilde{X} \subset X$. Il sera non convergent en général lorsque $\tilde{X} \not\subset X$.

3.2.2 Cas 2

Dans cette dernière situation, les estimateurs pondérés et non pondérés seront en général non convergents. Pour résoudre ce problème, il existe deux grandes classes de solutions :

1. Les modèles de sélection ;
2. Les méthodes de calage généralisé.

Ces deux solutions nécessitent l'existence d'instruments Z vérifiant des relations d'exclusion et disponibles sur toute la population (c'est-à-dire qu'on les observe sur les répondants et soit sur les non-répondants, soit dans une autre source comme un fichier administratif). Elles diffèrent quant à la nature de la relation d'exclusion. Dans le premier cas, on suppose que Z joue sur l'appartenance à l'échantillon des répondants mais pas directement sur Y :

$$Y \perp\!\!\!\perp Z|X \tag{3.3}$$

Par exemple, Z pourrait correspondre à des caractéristiques d'enquêteur. Ces caractéristiques sont susceptibles d'expliquer la non-réponse, mais ne jouent pas directement sur Y .

On peut alors obtenir des estimateurs convergents en corrigeant de la sélection par des modèles à la Heckman. Pour plus de détails sur les modèles de sélection, on se référera par exemple à Wooldridge (2002).

La méthode de calage généralisé (Deville, 2002), quant à elle, repose sur l'hypothèse que Z joue sur Y , mais est indépendant de la sélection conditionnellement à (\tilde{X}_2, Y) :

$$Z \perp\!\!\!\perp D | \tilde{X}_2, Y \quad (3.4)$$

Par exemple, dans une enquête sur la santé, la sélection peut être corrélée à un indicateur de santé Y car les personnes hospitalisées ne sont pas interrogées. On pourra utiliser comme instrument Z , si cette variable est connue, le nombre de médecins dans la zone. Cette variable est en effet corrélée a priori avec la santé des individus, mais pas directement à la réponse à l'enquête. De façon générale, si les concepteurs d'enquête pensent qu'un phénomène de ce type est en jeu, ils ont intérêt à utiliser le calage généralisé en incluant dans \tilde{X}_2 l'ensemble des facteurs déterminant la sélection. On est alors ramené au cadre de la partie 3.1, et tous les résultats correspondants sont valides.

On peut mettre en œuvre le calage généralisé sous SAS grâce à la macro `calmar2`¹⁴, avec la syntaxe suivante¹⁵ :

```
%calmar2(datamen=table_entree, marmen=table_marge, poids=inv_prob,
         nonrep=oui, datapoi=table_sortie, poidsfin=W)
```

Dans cet exemple, la table `table_entree` contient Y, \tilde{X}_2, Z et les inverses des probabilités de tirage `inv_prob` (i.e., $1/P(k \in S)$). La table `table_marge` contient les totaux des variables \tilde{X}_2 et Z . La macro `calmar2` crée la table `table_sortie` qui contient les mêmes variables que `table_entree`, ainsi que les poids W issus du calage généralisé.

3.3 Comparaison des estimateurs pondérés et non pondérés

L'objet de cette partie est de montrer qu'il est souvent possible de tester la compatibilité des hypothèses retenues sur la sélection et celles du modèle que l'on cherche à estimer. On suppose pour ce faire que $\tilde{X} \subset X$ et que H_0 ou H_0' est vérifiée.

14. Au moment où cette note a été écrite, la version la plus récente de la macro `calmar2` ne permettait pas d'estimer des poids en cas de sur-identification, c'est à dire lorsque la dimension de Z est supérieure à celle de Y .

15. Cette syntaxe n'inclut pas toutes les possibilités de la macro `calmar2`. Pour davantage de détails, cf. Le Guennec et Sautory (2005). On portera une attention particulière à leurs notations, qui diffèrent de celles adoptées ici.

3.3.1 Modèles paramétriques

Considérons ici, comme dans l'exemple 2, un modèle paramétrique $F_{Y|X,\eta}$ avec $\theta = f(\eta)$. Les estimateurs pondérés et non pondérés convergent vers le même paramètre. De plus, si θ est estimé par maximum de vraisemblance sans les pondérations, il est asymptotiquement efficace, c'est-à-dire de variance asymptotique minimale.

Dans ce genre de situations, on peut, à l'aide d'un test d'Hausman, tester la compatibilité du processus de sélection (i.e. l'hypothèse H0 ou H0') et du modèle paramétrique retenu $F_{Y|X,\eta}$. Ainsi, dans l'exemple 2, il se peut ou que l'hypothèse sur la sélection soit fautive, ou que la forme fonctionnelle retenue $P(Y = 1|X) = F(X'\theta)$ ne soit pas correcte. Le test d'Hausman est basé sur la différence des estimateurs pondéré et non pondéré. Sous l'hypothèse nulle que H0 ou H0' et le modèle paramétrique sont compatibles, ils doivent en effet être proches l'un de l'autre alors que sinon, ils ne convergeront pas a priori vers la même valeur. Si on note $\hat{\theta}_W$, l'estimateur calculé avec utilisation des poids et $\hat{\theta}$ celui sans utilisation des poids, la statistique du test d'Hausman est la suivante :

$$T_H = (\hat{\theta}_W - \hat{\theta})' [\hat{V}(\hat{\theta}_W) - \hat{V}(\hat{\theta})]^{-1} (\hat{\theta}_W - \hat{\theta}).$$

Sous l'hypothèse nulle, on a, lorsque la taille de l'échantillon tend vers l'infini, on a $T_H \xrightarrow{L} \chi^2(r)$ avec

$$r = \text{rg} \left(V(\hat{\theta}_W) - V(\hat{\theta}) \right).$$

De plus, T_H tend en général vers $+\infty$ sous l'hypothèse alternative. Ainsi, on rejettera au niveau α la compatibilité des hypothèses si $T_H > \chi_r^2(1 - \alpha)$, où $\chi_r^2(1 - \alpha)$ est le quantile d'ordre $1 - \alpha$ d'un χ^2 à r degrés de liberté.

Dans le cas où l'hypothèse est rejetée, trois interprétations sont possibles. Soit on maintient l'hypothèse que le processus de sélection vérifie H0 ou H0', ce qui revient à rejeter la validité du modèle. Soit on maintient l'hypothèse que le modèle est valide, ce qui conduit à rejeter H0 ou H0'. On peut évidemment également rejeter les deux hypothèses. Mais dans tous les cas, on ne peut maintenir simultanément les deux hypothèses.

3.3.2 Modèles linéaires

La discussion précédente portait sur les modèles paramétriques, ce qui exclut l'exemple important des modèles linéaires puisque le modèle est semi-paramétrique. Cependant le test est également valable dans ce cas car l'estimateur des MCO non pondéré est également optimal dans la classe des estimateurs linéaires sans biais sous l'hypothèse H0 ou H0',

d'après le théorème de Gauss-Markov. La statistique T_H définie ci-dessus converge donc là-aussi, sous l'hypothèse nulle, vers un χ^2 à r degrés de liberté.

Si l'implémentation du test nécessite de façon générale le recours à la programmation matricielle (via la PROC IML sous SAS), il n'en est rien pour le modèle linéaire. Au lieu de faire les deux regressions séparées et de calculer "à la main" la différence des variances des deux estimateurs, on peut remarquer qu'une solution équivalente est d'estimer le système d'équation suivant :

$$\begin{cases} Y = X'\theta_1 + \varepsilon_1 \\ \sqrt{W}Y = \sqrt{W}X'\theta_2 + \varepsilon_2 \end{cases}$$

Dans ce cadre, le test d'Hausman revient au test de $\theta_1 = \theta_2$. Sous SAS, on peut alors utiliser la PROC SYSLIN pour estimer les deux modèles et implémenter le test¹⁶, en utilisant la syntaxe suivante :

```
data a;
  set a;
  sqrtw=sqrt(w);          /* Racine carré des poids */
  xp=sqrt(w)*x;
  yp=sqrt(w)*y;
run;

proc syslin data=a;
  model y=x;              /* Première équation du système ci-dessus */
  model yp=sqrtw xp/noint; /* Deuxième équation, sans la constante */
                          /* (elle est capturée par sqrtw) */
  stest y.x=yp.xp;       /* Test d'Hausman */
run;
```

16. Dans la proc syslin, on estime $V(\hat{\theta}_W - \hat{\theta}|X)$ sans utiliser le fait que, quand $\hat{\theta}$ est efficace sous l'hypothèse nulle, on a :

$$V(\hat{\theta}_W - \hat{\theta}|X) = V(\hat{\theta}_W|X) - V(\hat{\theta}|X).$$

Cependant, l'estimateur de la variance reste convergent et le test est donc valide.

Hyp. sur la sélection	$\tilde{X} \subset X$?	θ dépend de...	Est. pond. convergent ? (si oui, efficace?)	Est. non pond. convergent ? (si oui, efficace?)	Test possible entre les deux ?
$(Y, X) \perp\!\!\!\perp D \tilde{X}$	oui	$F_{Y X}$	oui (non)	oui (oui)	oui
$(Y, X) \perp\!\!\!\perp D \tilde{X}$	non	$F_{Y X}$ ou $F_{X,Y}$	oui (non)	non	non
$(Y, X) \perp\!\!\!\perp D \tilde{X}$	n'importe	$F_{X,Y}$	oui (non)	non	non
$Y \perp\!\!\!\perp D X$	oui	$F_{Y X}$	oui (non)	oui (oui)	oui
$Y \perp\!\!\!\perp D X$	non	$F_{Y X}$	non	oui (oui)	non
$Y \perp\!\!\!\perp D X$	n'importe	$F_{X,Y}$	non	non	non
Non ignorable	n'importe	$F_{Y X}$ ou $F_{X,Y}$	non*	non**	non

* : sauf en calculant les poids par calage généralisé, et sous l'hypothèse (3.4).

** : sauf en utilisant une procédure à la Heckman, et sous l'hypothèse (3.3).

TABLE 1: Propriétés des estimateurs pondérés et non pondérés

3.4 Récapitulatif

Le tableau de la page précédente présente, en fonction du mécanisme de sélection et de la nature du paramètre θ , si les estimateurs pondérés et non pondérés sont convergents ou non. Dans le cas où ils le sont, il précise lequel est asymptotiquement efficace. A l'exception d'un seul cas, les estimateurs pondérés seront convergents à chaque fois que les estimateurs non pondérés le sont. Il existe en revanche des situations importantes où les estimateurs pondérés seront convergents alors que les estimateurs non pondérés ne le seront pas.

4 Calcul de précision avec des poids

Il s'agit ici de savoir si la précision calculée par les logiciels d'estimateurs pondérés est "correcte", i.e. si les estimateurs des écarts-types correspondants sont convergents ou non. Comme précisé précédemment, on néglige ici le plan de sondage précis de l'enquête en supposant que les indicatrices d'appartenance à l'échantillon des répondants (D_1, \dots, D_N) sont i.i.d¹⁷. La règle simple (mais inconfortable!) à retenir est la suivante :

Même si les logiciels de statistique prévoient en général l'inclusion de poids (via par exemple l'instruction `weight` sous SAS), les estimateurs de variance calculés sont non convergents en général.

Pour obtenir des estimateurs convergents des écarts-types sous les hypothèses précédentes, il est possible d'utiliser l'algorithme de bootstrap de la page suivante, dont la validité est établie en annexe. Cette procédure est très proche du bootstrap classique, mais s'en distingue par l'aléa sur la taille de l'échantillon. Intuitivement, cet aléa provient du fait que la taille de l'échantillon des répondants est aléatoire.

17. Cette approximation n'est pas correcte lorsque le sondage consiste à tirer des grappes d'individus, et non des individus. Dans ce cas, il faut tenir compte des corrélations au sein des grappes, à l'aide par exemple de l'option `Cluster` sous Stata.

Algorithme de bootstrap tenant compte des pondérations.

Pour $b = 1$ à B :

1. Tirer $n_b \sim \text{Binomiale}(N, n/N)$, où n la taille de l'échantillon des répondants ;
2. Tirer à probabilités égales et avec remise un échantillon de taille n_b issu de l'échantillon initial. On peut utiliser pour cela la commande suivante sous SAS (ici on échantillonne dans a et $n_b = 2500$) :

```
proc surveysselect data=a method=urs sampsize=2500 out=boot ;
run ;
```

 On note U_i^b le nombre de fois où l'individu i a été tiré dans l'échantillon bootstrap.
3. Estimer les poids $W_i^b = 1/P(D_i = 1|\tilde{X}_i)$, par un modèle de non-réponse et/ou un calage identique à celui effectué sur l'échantillon initial mais en utilisant les pondérations U_i^b (ou en construisant une table ayant U_i^b observations pour l'individu i de la table initiale) ;
4. Estimer le paramètre θ avec les poids $W_i^b U_i^b$. On note $\hat{\theta}_b$ l'estimateur obtenu.

Fin.

On peut ensuite estimer (par exemple) la variance de $\hat{\theta}$ par

$$\hat{V} = \frac{1}{B} \sum_{b=1}^B (\hat{\theta} - \hat{\theta}_b)^2 .$$

Même si la procédure ci-dessus est simple, elle peut s'avérer coûteuse en temps lorsque l'échantillon est de grande taille¹⁸. Pour certains exemples simples, comme ceux considérés ci-dessous, il est possible d'obtenir plus simplement des estimateurs des écarts-types, si l'on est prêt à négliger la variance due à l'estimation des pondérations (i.e., on assimile $\hat{P}(D = 1|\tilde{X})$ à $P(D = 1|\tilde{X})$).

Exemple 1 (suite).

Le modèle linéaire constitue un exemple important. On repart du modèle linéaire (3.1) sous l'hypothèse $E(X'\varepsilon) = 0$. On suppose que le processus de sélection satisfait H0. Alors on peut montrer (cf. annexe) que la matrice de variance de l'estimateur correspond à la matrice de White pour le modèle pondéré. Pour obtenir cette matrice sous SAS, il suffit

¹⁸. Pour accélérer la procédure, on pourra supprimer la phase 3 de l'algorithme précédent, et utiliser pour tous les échantillons bootstrap les poids W_i au lieu des poids W_i^b . Nous postulons que cette omission n'est pas trop problématique dans le sens qu'elle conduit en général à surestimer les variances des estimateurs.

d'utiliser l'option ACOV dans la proc REG :

```
proc reg data=a;
model y=x/acov;
weight w;
run;quit;
```

Exemple 2 (suite).

Supposons que l'on souhaite estimer un modèle logit ou probit en pondérant, par exemple parce que $\tilde{X} \not\subset X$. Dans ce cas, l'estimateur $\widehat{\theta}^W$ peut être obtenu sous SAS aussi bien avec la PROC LOGISTIC que la PROC SURVEYLOGISTIC. En revanche, l'estimateur de variance de $\widehat{\theta}^W$ proposé par la PROC LOGISTIC n'est pas convergent (cf. annexe pour davantage de détails). On veillera donc à utiliser la PROC SURVEYLOGISTIC pour estimer la précision des estimateurs¹⁹.

```
proc surveylogistic data=a;
model y (descending)= x; /* L'option link=probit permet d'estimer un probit */
weight w;
run;
```

19. L'intérêt d'utiliser la PROC LOGISTIC lorsque l'on pondère est qu'elle possède des options qui n'existent pas dans la PROC SURVEYLOGISTIC. Par exemple, il n'y a pas d'option permettant de récupérer directement l'estimation de $P(Y = 1|X)$ pour chaque individu dans la PROC SURVEYLOGISTIC.

5 Conclusion : feuille de route du chargé d'études recourant à des données d'enquêtes

Cette conclusion résume les différentes étapes décrites dans la présente note. On se met ici dans la situation d'un chargé d'étude ayant à estimer un modèle donné sur un fichier pondéré. Il connaît donc la (les) variable(s) endogènes Y et les variables exogènes X ainsi que le modèle économétrique qu'il souhaite estimer.

1. Première question : quelles sont les variables \tilde{X} utilisées pour calculer les pondérations ? Pour répondre à cette question consulter la documentation de la source ou discuter avec le service producteur.
2. Deuxième question : quelle est la nature de la sélection ? En d'autres termes, est-on sous l'hypothèse $H0$, sous l'hypothèse $H0'$ ou dans un cas de non réponse non ignorable ? Si $H0$ ou $H0'$, aller en 3. Si non réponse non-ignorable déterminer si on peut utiliser un modèle de sélection ou une méthode de calage généralisé. Dans le premier cas on utilisera un estimateur non pondéré, dans le deuxième cas on utilisera un estimateur pondéré. Si la non-réponse est non-ignorable et que ni le modèle de sélection ni la méthode de calage généralisé ne sont utilisables, la théorie d'une estimation convergente reste à construire !
3. Troisième question : faut-il pondérer ou non ? Pour cela on se référera au tableau 1 p. 13. Pour identifier la ligne du tableau à considérer, il s'agit de savoir si $\tilde{X} \subset X$ et si θ dépend seulement de la loi conditionnelle $F_{Y|X}$ ou de la loi jointe $F_{X,Y}$.
4. Calculer les estimateurs choisis à l'étape 3.
5. Estimer la précision des paramètres estimés à l'étape précédente. S'il s'agit d'un estimateur non pondéré, les procédures usuelles donnent directement des estimateurs corrects de précision. S'il s'agit d'un estimateur pondéré les précisions sont en général mal estimées par les procédures usuelles. Dans ce dernier cas, on peut soit utiliser des procédures *ad hoc* (cf. exemples 1 et 2 p. 15 et 16), soit utiliser la méthode de bootstrap présentée p. 15.
6. Dernière question : les hypothèses précédentes sont-elles raisonnables ? Si les estimateurs pondérés et non pondérés convergent tous les deux, il est possible de tester la compatibilité des hypothèses économétriques du modèle avec les hypothèses concernant la nature du processus de sélection. En cas de rejet de la compatibilité, il faut soit changer de modèle, soit revenir sur les hypothèses $H0$ et/ou $H0'$.

Annexe : preuves et précisions techniques

Résultats de la partie 3

Résultats 1 et 2

Montrons d'abord que l'estimateur pondéré de $F_{X,Y}$ est convergent. Sous H_0 , il tend vers²⁰ :

$$\begin{aligned} E [WD\mathbf{1}\{Y \leq y, X \leq x\}] &= E \left[WE[D|X, Y, \tilde{X}]\mathbf{1}\{Y \leq y, X \leq x\} \right] \\ &= E \left[WE[D|\tilde{X}]\mathbf{1}\{Y \leq y, X \leq x\} \right] \\ &= E [\mathbf{1}\{Y \leq y, X \leq x\}]. \end{aligned}$$

La deuxième égalité provient de H_0 , la troisième de la définition des poids. Ainsi, l'estimateur pondéré de θ sera convergent, que θ dépende de $F_{Y|X}$ seulement ou de la loi jointe de (X, Y) .

Pour montrer que l'estimation non pondérée est valide et "efficace", il suffit de constater que H_0 et $\tilde{X} \subset X$ impliquent H_0' . En effet, pour toute fonction g ,

$$E [g(Y)|D = 1, X] = \frac{E [Dg(Y)|X]}{P(D = 1|X)} = \frac{E \left[E[D|X, Y, \tilde{X}]g(Y)|X \right]}{P(D = 1|\tilde{X})} = E [g(Y)|X].$$

On peut alors utiliser le résultat 3 ci-dessous.

Résultat 3

On a, par l'hypothèse H_0' , $F_{Y|X, D=1} = F_{Y|X}$. Par conséquent, l'inférence menée sans pondération sur les répondants sera valide. De plus, les estimateurs non pondérés qui sont asymptotiquement efficaces en l'absence de sélection le seront aussi ici.

Validité de la méthode de bootstrap proposée

Sous les hypothèses retenues, l'échantillon $(Y_i, D_i, W_i)_{i \in U}$ est i.i.d. Le bootstrap standard effectué sur U est donc valide en général²¹. Ce bootstrap revient à tirer des poids $(U_i^b)_{i \in U}$

20. On suppose ici que les conditions de régularité assurant la convergence des contreparties empiriques sont vérifiées. Pour des détails sur cette question, le lecteur pourra se référer par exemple à van der Vaart (1998).

21. Il existe des contre-exemples. Ainsi, la distribution bootstrapée de l'estimateur standard de θ dans un modèle $U[0, \theta]$ est non convergente. De même, Abadie et Imbens (2008) montrent la non convergence du bootstrap dans les procédures d'appariement par le plus proche voisin.

vérifiant :

$$(U_1^b, \dots, U_N^b) \sim \mathcal{M}(N, 1/N, \dots, 1/N),$$

où \mathcal{M} désigne une loi multinomiale. Cependant, il est inutile de tirer tous ces poids car seuls ceux correspondants aux individus répondants seront utilisés. Notons $n_b = \sum_{i \in \mathcal{R}} U_i^b$. On obtient facilement :

$$n_b \sim \text{Bin}(N, n/N).$$

Notons $\mathcal{R} = \{r_1, \dots, r_n\}$. Conditionnellement à n_b , on a, par symétrie :

$$(U_{r_1}^b, \dots, U_{r_n}^b) \sim \mathcal{M}(n_b, 1/n, \dots, 1/n).$$

L'algorithme proposé est donc convergent.

Variance de l'estimateur pondéré dans le modèle linéaire

Dans le modèle (3.1), sous les hypothèses $E(X'\varepsilon) = 0$, $E(|X'X|) < \infty$ et $E(|W^2 X'X \varepsilon^2|) < \infty$, on a le résultat asymptotique suivant :

$$\sqrt{N} (\hat{\beta} - \beta) \longrightarrow \mathcal{N}(0, V)$$

La variance asymptotique de l'estimateur pondéré est

$$V = [E(WX'X|D=1)]^{-1} E(W^2 X' \varepsilon^2 X|D=1) [E(WX'X|D=1)]^{-1}.$$

Cette variance peut être estimée sur l'échantillon par les contreparties empiriques suivantes :

$$\hat{V} = \left(\frac{1}{N} \sum W_i X_i' X_i \right)^{-1} \left(\frac{1}{N} \sum W_i^2 \hat{\varepsilon}_i^2 X_i' X_i \right) \left(\frac{1}{N} \sum W_i X_i' X_i \right)^{-1}$$

Il s'agit de la matrice de White pour le modèle pondéré. On peut donc utiliser l'estimateur robuste de la matrice de variance-covariance, tel que calculé par SAS via l'option ACOV.

Variance de l'estimateur pondéré dans les modèles logit ou probit

On peut montrer que l'estimateur pondéré a comme matrice de variance covariance

$$V_0 = E \left(\frac{F'(X'\theta)^2}{F(X'\theta)(1-F(X'\theta))} X X' \right)^{-1} E \left(W \frac{F'(X'\theta)^2}{F(X'\theta)(1-F(X'\theta))} X X' \right) \\ E \left(\frac{F'(X'\theta)^2}{F(X'\theta)(1-F(X'\theta))} X X' \right)^{-1}$$

La PROC SURVEYLOGISTIC permet d'estimer V_0 de manière convergente²² par

$$\widehat{V}_0 = N \left(\sum_i W_i \frac{F'(X_i'\theta)^2}{F(X_i'\theta)(1-F(X_i'\theta))} X_i X_i' \right)^{-1} \left(\sum_i W_i^2 \frac{F'(X_i'\theta)^2 (Y_i - F(X_i'\theta))^2}{F^2(X_i'\theta)(1-F(X_i'\theta))^2} X_i X_i' \right) \left(\sum_i W_i \frac{F'(X_i'\theta)^2}{F(X_i'\theta)(1-F(X_i'\theta))} X_i X_i' \right)^{-1}.$$

Il s'agit ni plus ni moins que la formule utilisée par les sondeurs pour calculer la variance sous un plan de sondage poissonien dont les probabilités d'ordre 1 sont les $\frac{1}{W_i}$.

Notons que l'estimateur de la variance asymptotique estimée avec la PROC LOGISTIC s'écrit

$$\widehat{V}_1 = \left(\frac{1}{N} \sum_i W_i \frac{F'(X_i'\theta)^2}{F(X_i'\theta)(1-F(X_i'\theta))} X_i X_i' \right)^{-1}$$

On peut remarquer que la variance diminue avec une augmentation des poids et N constant, ce qui est pathologique dans le cadre dans lequel nous sommes ! Cela conduit donc certains utilisateurs à normaliser les poids de telle manière que $\sum_i W_i = N$, dans ce cas

$$\widehat{V}_1 = \left(\frac{1}{\sum_i W_i} \sum_i W_i \frac{F'(X_i'\theta)^2}{F(X_i'\theta)(1-F(X_i'\theta))} X_i X_i' \right)^{-1}.$$

Dans ce dernier cas, l'estimateur de la variance tend vers

$$\lim \widehat{V}_1 = E \left(\frac{F'(X'\theta)^2}{F(X'\theta)(1-F(X'\theta))} X X' \right)^{-1} \neq V_0.$$

En d'autres termes, l'estimateur de la variance proposé par la PROC LOGISTIC n'est jamais convergent, quels que soient les poids utilisés.

22. Le terme central de l'estimateur de la PROC SURVEYLOGISTIC peut sembler étrange, mais c'est un estimateur convergent de $E \left(W \frac{F'(X'\theta)^2}{F(X'\theta)(1-F(X'\theta))} X X' \right)$ car sous les hypothèses du modèle, $E((Y_i - F(X_i'\theta))^2 | X_i) = F(X_i'\theta)(1 - F(X_i'\theta))$.

Références

- Abadie, A. et Imbens, G. W. (2008), ‘On the failure of the bootstrap for matching estimators’, *Econometrica* **76**, 1537–1557.
- Deville, J. C. (2002), La correction de la non-réponse par calage généralisé, *in* ‘Actes des Journées de Méthodologie Statistique 2002’, INSEE, pp. 4–20.
- Deville, J. C. et Sarndal, C. E. (1992), ‘Calibration estimators in survey sampling’, *Journal of the American Statistical Association* **87**, 376–382.
- Imbens, G. W. (1992), ‘An efficient method of moments estimator for discrete choice models with choice-based sampling’, *Econometrica* **60**, 1187–1214.
- Le Guennec, J. et Sautory, O. (2005), La macro calmar2 : redressement d’un échantillon par calage sur marges. Document INSEE.
- Lerman, S. et Manski, C. F. (1977), ‘The estimation of choice probabilities from choice based samples’, *Econometrica* **45**, 1977–1988.
- Sautory, O. (1993), La macro calmar : redressement d’un échantillon par calage sur marges. Document F9310, DSDS, INSEE.
- Tillé, Y. (2001), *Théorie des sondages : Échantillonnage et estimation en populations finies : cours et exercices*, Dunod.
- van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.
- Wooldridge, J. (2002), *Econometrics of Cross Section and Panel Data*, MIT Press.
- Wooldridge, J. (2007), ‘Inverse probability weighted estimation for general missing data problems’, *Journal of Econometrics* **141**, 1281–1301.

G 9001	J. FAYOLLE et M. FLEURBAEY Accumulation, profitabilité et endettement des entreprises		Macro-economic import functions with imperfect competition - An application to the E.C. Trade	G 9311	J. BOURDIEU - B. COLIN-SEDILLOT Les décisions de financement des entreprises françaises : une évaluation empirique des théories de la structure optimale du capital		analyse économique des politiques française et allemande
G 9002	H. ROUSSE Détection et effets de la multicolinéarité dans les modèles linéaires ordinaires - Un prolongement de la réflexion de BELSLEY, KUH et WELSCH	G 9203	I. STAPIC Les échanges internationaux de services de la France dans le cadre des négociations multilatérales du GATT Juin 1992 (1ère version) Novembre 1992 (version finale)	G 9312	L. BLOCH - B. CŒURÉ Q de Tobin marginal et transmission des chocs financiers	G 9412	J. BOURDIEU - B. CŒURÉ - B. COLIN-SEDILLOT Investissement, incertitude et irréversibilité Quelques développements récents de la théorie de l'investissement
G 9003	P. RALLE et J. TOUJAS-BERNATE Indexation des salaires : la rupture de 1983	G 9204	P. SEVESTRE L'économétrie sur données individuelles-temporelles. Une note introductive	G 9313	Equipes Amadeus (INSEE), Banque de France, Méric (DP) Présentation des propriétés des principaux modèles macroéconomiques du Service Public	G 9413	B. DORMONT - M. PAUCHET L'évaluation de l'élasticité emploi-salaire dépendelle des structures de qualification ?
G 9004	D. GUELLEC et P. RALLE Compétitivité, croissance et innovation de produit	G 9205	H. ERKEL-ROUSSE Le commerce extérieur et l'environnement international dans le modèle AMADEUS (réestimation 1992)	G 9314	B. CREPON - E. DUGUET Research & Development, competition and innovation	G 9414	I. KABLA Le Choix de breveter une invention
G 9005	P. RALLE et J. TOUJAS-BERNATE Les conséquences de la désindexation. Analyse dans une maquette prix-salaires	G 9206	N. GREENAN et D. GUELLEC Coordination within the firm and endogenous growth	G 9315	B. DORMONT Quelle est l'influence du coût du travail sur l'emploi ?	G 9501	J. BOURDIEU - B. CŒURÉ - B. SEDILLOT Irreversible Investment and Uncertainty : When is there a Value of Waiting ?
G 9101	Equipe AMADEUS Le modèle AMADEUS - Première partie - Présentation générale	G 9207	A. MAGNIER et J. TOUJAS-BERNATE Technology and trade : empirical evidences for the major five industrialized countries	G 9316	D. BLANCHET - C. BROUSSE Deux études sur l'âge de la retraite	G 9502	L. BLOCH - B. CŒURÉ Imperfections du marché du crédit, investissements des entreprises et cycle économique
G 9102	J.L. BRILLET Le modèle AMADEUS - Deuxième partie - Propriétés variantielles	G 9208	B. CREPON, E. DUGUET, D. ENCAOUA et P. MOHNEN Cooperative, non cooperative R & D and optimal patent life	G 9317	D. BLANCHET Répartition du travail dans une population hétérogène : deux notes	G 9503	D. GOUX - E. MAURIN Les transformations de la demande de travail par qualification en France Une étude sur la période 1970-1993
G 9103	D. GUELLEC et P. RALLE Endogenous growth and product innovation	G 9209	B. CREPON et E. DUGUET Research and development, competition and innovation : an application of pseudo maximum likelihood methods to Poisson models with heterogeneity	G 9318	D. EYSSARTIER - N. PONTY AMADEUS - an annual macro-economic model for the medium and long term	G 9504	N. GREENAN Technologie, changement organisationnel, qualifications et emploi : une étude empirique sur l'industrie manufacturière
G 9104	H. ROUSSE Le modèle AMADEUS - Troisième partie - Le commerce extérieur et l'environnement international	G 9301	J. TOUJAS-BERNATE Commerce international et concurrence imparfaite : développements récents et implications pour la politique commerciale	G 9319	G. CETTE - Ph. CUNÉO - D. EYSSARTIER - J. GAUTIÉ Les effets sur l'emploi d'un abaissement du coût du travail des jeunes	G 9505	D. GOUX - E. MAURIN Persistence des hiérarchies sectorielles de salaires: un réexamen sur données françaises
G 9105	H. ROUSSE Effets de demande et d'offre dans les résultats du commerce extérieur manufacturé de la France au cours des deux dernières décennies	G 9302	Ch. CASES Durées de chômage et comportements d'offre de travail : une revue de la littérature	G 9401	D. BLANCHET Les structures par âge importent-elles ?	G 9505 Bis	D. GOUX - E. MAURIN Persistence of inter-industry wages differentials: a reexamination on matched worker-firm panel data
G 9106	B. CREPON Innovation, taille et concentration : causalités et dynamiques	G 9303	H. ERKEL-ROUSSE Union économique et monétaire : le débat économique	G 9402	J. GAUTIÉ Le chômage des jeunes en France : problème de formation ou phénomène de file d'attente ? Quelques éléments du débat	G 9506	S. JACOBZONE Les liens entre RMI et chômage, une mise en perspective <i>NON PARU - article sorti dans Economie et Prévision n° 122 (1996) - pages 95 à 113</i>
G 9107	B. AMABLE et D. GUELLEC Un panorama des théories de la croissance endogène	G 9304	N. GREENAN - D. GUELLEC / G. BROUSSAUDIER - L. MIOTTI Innovation organisationnelle, dynamisme technologique et performances des entreprises	G 9403	P. QUIRION Les déchets en France : éléments statistiques et économiques	G 9507	G. CETTE - S. MAHFOUZ Le partage primaire du revenu Constat descriptif sur longue période
G 9108	M. GLAUDE et M. MOUTARDIER Une évaluation du coût direct de l'enfant de 1979 à 1989	G 9305	P. JAILLARD Le traité de Maastricht : présentation juridique et historique	G 9404	D. LADIRAY - M. GRUN-REHOMME Lissage par moyennes mobiles - Le problème des extrémités de série	G 9601	Banque de France - CEPREMAP - Direction de la Prévision - Erasme - INSEE - OFCE Structures et propriétés de cinq modèles macro-économiques français
G 9109	P. RALLE et alii France - Allemagne : performances économiques comparées	G 9306	J.L. BRILLET Micro-DMS : présentation et propriétés	G 9405	V. MAILLARD Théorie et pratique de la correction des effets de jours ouvrables	G 9602	Rapport d'activité de la DESE de l'année 1995
G 9110	J.L. BRILLET Micro-DMS NON PARU	G 9307	J.L. BRILLET Micro-DMS - variantes : les tableaux	G 9406	F. ROSENWALD La décision d'investir	G 9603	J. BOURDIEU - A. DRAZNIKS L'octroi de crédit aux PME : une analyse à partir d'informations bancaires
G 9111	A. MAGNIER Effets accélérateur et multiplicateur en France depuis 1970 : quelques résultats empiriques	G 9308	S. JACOBZONE Les grands réseaux publics français dans une perspective européenne	G 9407	S. JACOBZONE Les apports de l'économie industrielle pour définir la stratégie économique de l'hôpital public	G 9604	A. TOPIOL-BENSAÏD Les implantations japonaises en France
G 9112	B. CREPON et G. DUREAU Investissement en recherche-développement : analyse de causalités dans un modèle d'accélérateur généralisé	G 9309	L. BLOCH - B. CŒURE Profitabilité de l'investissement productif et transmission des chocs financiers	G 9408	L. BLOCH, J. BOURDIEU, B. COLIN-SEDILLOT, G. LONGUEVILLE Du défaut de paiement au dépôt de bilan : les banquiers face aux PME en difficulté	G 9605	P. GENIER - S. JACOBZONE Comportements de prévention, consommation d'alcool et tabagie : peut-on parler d'une gestion globale du capital santé ? <i>Une modélisation microéconométrique empirique</i>
G 9113	J.L. BRILLET, H. ERKEL-ROUSSE, J. TOUJAS-BERNATE "France-Allemagne Couplées" - Deux économies vues par une maquette macro-économétrique	G 9310	J. BOURDIEU - B. COLIN-SEDILLOT Les théories sur la structure optimal du capital : quelques points de repère	G 9409	D. EYSSARTIER, P. MAIRE Impacts macro-économiques de mesures d'aide au logement - quelques éléments d'évaluation	G 9606	C. DOZ - F. LENGART Factor analysis and unobserved component models: an application to the study of French business surveys
G 9201	W.J. ADAMS, B. CREPON, D. ENCAOUA Choix technologiques et stratégies de dissuasion d'entrée			G 9410	F. ROSENWALD Suivi conjoncturel de l'investissement	G 9607	N. GREENAN - D. GUELLEC La théorie coopérative de la firme
G 9202	J. OLIVEIRA-MARTINS, J. TOUJAS-BERNATE			G 9411	C. DEFEUILLEY - Ph. QUIRION Les déchets d'emballages ménagers : une		

G 9608	N. GREENAN - D. GUELLEC Technological innovation and employment reallocation
G 9609	Ph. COUR - F. RUPPRECHT L'intégration asymétrique au sein du continent américain : un essai de modélisation
G 9610	S. DUCHENE - G. FORGEOT - A. JACQUOT Analyse des évolutions récentes de la productivité apparente du travail
G 9611	X. BONNET - S. MAHFOUZ The influence of different specifications of wages-prices spirals on the measure of the NAIRU : the case of France
G 9612	PH. COUR - E. DUBOIS, S. MAHFOUZ, J. PISANI-FERRY The cost of fiscal retrenchment revisited: how strong is the evidence ?
G 9613	A. JACQUOT Les flexions des taux d'activité sont-elles seulement conjoncturelles ?
G 9614	ZHANG Yingxiang - SONG Xueqing Lexique macroéconomique Français-Chinois
G 9701	J.L. SCHNEIDER La taxe professionnelle : éléments de cadrage économique
G 9702	J.L. SCHNEIDER Transition et stabilité politique d'un système redistributif
G 9703	D. GOUX - E. MAURIN Train or Pay: Does it Reduce Inequalities to Encourage Firms to Train their Workers?
G 9704	P. GENIER Deux contributions sur dépendance et équité
G 9705	E. DUGUET - N. IUNG R & D Investment, Patent Life and Patent Value An Econometric Analysis at the Firm Level
G 9706	M. HOUEBINE - A. TOPIOL-BENSAÏD Les entreprises internationales en France : une analyse à partir de données individuelles
G 9707	M. HOUEBINE Polarisation des activités et spécialisation des départements en France
G 9708	E. DUGUET - N. GREENAN Le biais technologique : une analyse sur données individuelles
G 9709	J.L. BRILLET Analyzing a small French ECM Model
G 9710	J.L. BRILLET Formalizing the transition process : scenarios for capital accumulation
G 9711	G. FORGEOT - J. GAUTÉ Insertion professionnelle des jeunes et processus de déclassement
G 9712	E. DUBOIS High Real Interest Rates: the Consequence of a Saving Investment Disequilibrium or of an insufficient Credibility of Monetary Authorities?
G 9713	Bilan des activités de la Direction des Etudes et Synthèses Economiques - 1996

G 9714	F. LEQUILLER Does the French Consumer Price Index Overstate Inflation?
G 9715	X. BONNET Peut-on mettre en évidence les rigidités à la baisse des salaires nominaux ? Une étude sur quelques grands pays de l'OCDE
G 9716	N. IUNG - F. RUPPRECHT Productivité de la recherche et rendements d'échelle dans le secteur pharmaceutique français
G 9717	E. DUGUET - I. KABLA Appropriation strategy and the motivations to use the patent system in France - An econometric analysis at the firm level
G 9718	L.P. PELÉ - P. RALLE Âge de la retraite : les aspects incitatifs du régime général
G 9719	ZHANG Yingxiang - SONG Xueqing Lexique macroéconomique français-chinois, chinois-français
G 9720	M. HOUEBINE - J.L. SCHNEIDER Mesurer l'influence de la fiscalité sur la localisation des entreprises
G 9721	A. MOURougANE Crédibilité, indépendance et politique monétaire Une revue de la littérature
G 9722	P. AUGERAUD - L. BRIOT Les données comptables d'entreprises Le système intermédiaire d'entreprises Passage des données individuelles aux données sectorielles
G 9723	P. AUGERAUD - J.E. CHAPRON Using Business Accounts for Compiling National Accounts: the French Experience
G 9724	P. AUGERAUD Les comptes d'entreprise par activités - Le passage aux comptes - De la comptabilité d'entreprise à la comptabilité nationale - A <i>paraître</i>
G 9801	H. MICHAUDON - C. PRIGENT Présentation du modèle AMADEUS
G 9802	J. ACCARDO Une étude de comptabilité générationnelle pour la France en 1996
G 9803	X. BONNET - S. DUCHÊNE Apports et limites de la modélisation « Real Business Cycles »
G 9804	C. BARLET - C. DUGUET - D. ENCAOUA - J. PRADEL The Commercial Success of Innovations An econometric analysis at the firm level in French manufacturing
G 9805	P. CAHUC - Ch. GIANELLA - D. GOUX - A. ZILBERBERG Equalizing Wage Differences and Bargaining Power - Evidence from a Panel of French Firms
G 9806	J. ACCARDO - M. JLASSI La productivité globale des facteurs entre 1975 et 1996

G 9807	Bilan des activités de la Direction des Etudes et Synthèses Economiques - 1997	Bis	Une estimation de l'élasticité de l'emploi peu qualifié à son coût
G 9808	A. MOURougANE Can a Conservative Governor Conduct an Accomodative Monetary Policy ?	G 9913	Division « Redistribution et Politiques Sociales » Le modèle de microsimulation dynamique DESTINIE
G 9809	X. BONNET - E. DUBOIS - L. FAUVET Asymétrie des inflations relatives et menus costs : tests sur l'inflation française	G 9914	E. DUGUET Macro-commandes SAS pour l'économétrie des panels et des variables qualitatives
G 9810	E. DUGUET - N. IUNG Sales and Advertising with Spillovers at the firm level: Estimation of a Dynamic Structural Model on Panel Data	G 9915	R. DUHAUTOIS Evolution des flux d'emplois en France entre 1990 et 1996 : une étude empirique à partir du fichier des bénéficiaires réels normaux (BRN)
G 9811	J.P. BERTHIER Congestion urbaine : un modèle de trafic de pointe à courbe débit-vitesse et demande élastique	G 9916	J.Y. FOURNIER Extraction du cycle des affaires : la méthode de Baxter et King
G 9812	C. PRIGENT La part des salaires dans la valeur ajoutée : une approche macroéconomique	G 9917	B. CRÉPON - R. DESPLATZ - J. MAIRESSE Estimating price cost margins, scale economies and workers' bargaining power at the firm level
G 9813	A.Th. AERTS L'évolution de la part des salaires dans la valeur ajoutée en France reflète-t-elle les évolutions individuelles sur la période 1979-1994 ?	G 9918	Ch. GIANELLA - Ph. LAGARDE Productivity of hours in the aggregate production function: an evaluation on a panel of French firms from the manufacturing sector
G 9814	B. SALANIÉ Guide pratique des séries non-stationnaires	G 9919	S. AUDRIC - P. GIVORD - C. PROST Evolution de l'emploi et des coûts par qualification entre 1982 et 1996
G 9901	S. DUCHÊNE - A. JACQUOT Une croissance plus riche en emplois depuis le début de la décennie ? Une analyse en comparaison internationale	G 2000/01	R. MAHIEU Les déterminants des dépenses de santé : une approche macroéconomique
G 9902	Ch. COLIN Modélisation des carrières dans Destinie	G 2000/02	C. ALLARD-PRIGENT - H. GUILMEAU - A. QUINET The real exchange rate as the relative price of nontradables in terms of tradables: theoretical investigation and empirical study on French data
G 9903	Ch. COLIN Evolution de la dispersion des salaires : un essai de prospective par microsimulation	G 2000/03	J.-Y. FOURNIER L'approximation du filtre passe-bande proposée par Christiano et Fitzgerald
G 9904	B. CREPON - N. IUNG Innovation, emploi et performances	G 2000/04	Bilan des activités de la DESE - 1999
G 9905	B. CREPON - Ch. GIANELLA Wages inequalities in France 1969-1992 An application of quantile regression techniques	G 2000/05	B. CREPON - F. ROSENWALD Investissement et contraintes de financement : le poids du cycle Une estimation sur données françaises
G 9906	C. BONNET - R. MAHIEU Microsimulation techniques applied to inter-generational transfers - Pensions in a dynamic framework: the case of France	G 2000/06	A. FLIPO Les comportements matrimoniaux de fait
G 9907	F. ROSENWALD L'impact des contraintes financières dans la décision d'investissement	G 2000/07	R. MAHIEU - B. SÉDILLOT Microsimulations of the retirement decision: a supply side approach
G 9908	Bilan des activités de la DESE - 1998	G 2000/08	C. AUDENIS - C. PROST Déficit conjoncturel : une prise en compte des conjonctures passées
G 9909	J.P. ZOYEM Contrat d'insertion et sortie du RMI Evaluation des effets d'une politique sociale	G 2000/09	R. MAHIEU - B. SÉDILLOT Equivalent patrimonial de la rente et souscription de retraite complémentaire
G 9910	Ch. COLIN - FI. LEGROS - R. MAHIEU Bilans contributifs comparés des régimes de retraite du secteur privé et de la fonction publique	G 2000/10	R. DUHAUTOIS Ralentissement de l'investissement : petites ou grandes entreprises ? industrie ou tertiaire ?
G 9911	G. LAROQUE - B. SALANIÉ Une décomposition du non-emploi en France	G 2000/11	G. LAROQUE - B. SALANIÉ Temps partiel féminin et incitations financières à l'emploi
G 9912	B. SALANIÉ Une maquette analytique de long terme du marché du travail	G2000/12	Ch. GIANELLA Local unemployment and wages
G 9912	Ch. GIANELLA		

G2000/13	B. CREPON - Th. HECKEL - Informatisation en France : une évaluation à partir de données individuelles - Computerization in France: an evaluation based on individual company data
G2001/01	F. LEQUILLER - La nouvelle économie et la mesure de la croissance du PIB - The new economy and the measurement of GDP growth
G2001/02	S. AUDRIC La reprise de la croissance de l'emploi profite-t-elle aussi aux non-diplômés ?
G2001/03	I. BRAUN-LEMAIRE Evolution et répartition du surplus de productivité
G2001/04	A. BEAUDU - Th. HECKEL Le canal du crédit fonctionne-t-il en Europe ? Une étude de l'hétérogénéité des comportements d'investissement à partir de données de bilan agrégées
G2001/05	C. AUDENIS - P. BISCOURP - N. FOURCADE - O. LOISEL Testing the augmented Solow growth model : An empirical reassessment using panel data
G2001/06	R. MAHIEU - B. SÉDILLOT Départ à la retraite, irréversibilité et incertitude
G2001/07	Bilan des activités de la DESE - 2000
G2001/08	J. Ph. GAUDEMET Les dispositifs d'acquisition à titre facultatif d'annuités viagères de retraite
G2001/09	B. CRÉPON - Ch. GIANELLA Fiscalité, coût d'usage du capital et demande de facteurs : une analyse sur données individuelles
G2001/10	B. CRÉPON - R. DESPLATZ Evaluation des effets des dispositifs d'allègements de charges sociales sur les bas salaires
G2001/11	J.-Y. FOURNIER Comparaison des salaires des secteurs public et privé
G2001/12	J.-P. BERTHIER - C. JAULENT R. CONVENEVOLE - S. PISANI Une méthodologie de comparaison entre consommations intermédiaires de source fiscale et de comptabilité nationale
G2001/13	P. BISCOURP - Ch. GIANELLA Substitution and complementarity between capital, skilled and less skilled workers: an analysis at the firm level in the French manufacturing industry
G2001/14	I. ROBERT-BOBEE Modelling demographic behaviours in the French microsimulation model Destinie: An analysis of future change in completed fertility
G2001/15	J.-P. ZOYEM Diagnostic sur la pauvreté et calendrier de revenus : le cas du "Panel européen des ménages"
G2001/16	J.-Y. FOURNIER - P. GIVORD La réduction des taux d'activité aux âges extrêmes, une spécificité française ?

G2001/17	C. AUDENIS - P. BISCOURP - N. RIEDINGER Existe-t-il une asymétrie dans la transmission du prix du brut aux prix des carburants ?
G2002/01	F. MAGNIEN - J.-L. TAVERNIER - D. THESMAR Les statistiques internationales de PIB par habitant en standard de pouvoir d'achat : une analyse des résultats
G2002/02	Bilan des activités de la DESE - 2001
G2002/03	B. SÉDILLOT - E. WALRAET La cessation d'activité au sein des couples : y a-t-il interdépendance des choix ?
G2002/04	G. BRILHAULT - Rétropolation des séries de FBCF et calcul du capital fixe en SEC-95 dans les comptes nationaux français - Retropolation of the investment series (GFCF) and estimation of fixed capital stocks on the ESA-95 basis for the French balance sheets
G2002/05	P. BISCOURP - B. CRÉPON - T. HECKEL - N. RIEDINGER How do firms respond to cheaper computers? Microeconomic evidence for France based on a production function approach
G2002/06	C. AUDENIS - J. DERUYON - N. FOURCADE L'impact des nouvelles technologies de l'information et de la communication sur l'économie française - un bouclage macro-économique
G2002/07	J. BARDAJI - B. SÉDILLOT - E. WALRAET Évaluation de trois réformes du Régime Général d'assurance vieillesse à l'aide du modèle de microsimulation DESTINIE
G2002/08	J.-P. BERTHIER Réflexions sur les différentes notions de volume dans les comptes nationaux : comptes aux prix d'une année fixe ou aux prix de l'année précédente, séries chaînées
G2002/09	F. HILD Les soldes d'opinion résumant-ils au mieux les réponses des entreprises aux enquêtes de conjoncture ?
G2002/10	I. ROBERT-BOBÉE Les comportements démographiques dans le modèle de microsimulation Destinie - Une comparaison des estimations issues des enquêtes Jeunes et Carrières 1997 et Histoire Familiale 1999
G2002/11	J.-P. ZOYEM La dynamique des bas revenus : une analyse des entrées-sorties de pauvreté
G2002/12	F. HILD Prévisions d'inflation pour la France
G2002/13	M. LECLAIR Réduction du temps de travail et tensions sur les facteurs de production
G2002/14	E. WALRAET - A. VINCENT - Analyse de la redistribution intragénérationnelle dans le système de retraite des salariés du privé - Une approche par microsimulation - Intragenerational distributional analysis in the french private sector pension scheme - A microsimulation approach

G2002/15	P. CHONE - D. LE BLANC - I. ROBERT-BOBEE Offre de travail féminine et garde des jeunes enfants
G2002/16	F. MAUREL - S. GREGOIR Les indices de compétitivité des pays : interprétation et limites
G2003/01	N. RIEDINGER - E. HAUVY Le coût de dépollution atmosphérique pour les entreprises françaises : Une estimation à partir de données individuelles
G2003/02	P. BISCOURP et F. KRAMARZ Création d'emplois, destruction d'emplois et internationalisation des entreprises industrielles françaises : une analyse sur la période 1986-1992
G2003/03	Bilan des activités de la DESE - 2002
G2003/04	P.-O. BEFFY - J. DERUYON - N. FOURCADE - S. GREGOIR - N. LAÏB - B. MONFORT Évolutions démographiques et croissance : une projection macro-économique à l'horizon 2020
G2003/05	P. AUBERT La situation des salariés de plus de cinquante ans dans le secteur privé
G2003/06	P. AUBERT - B. CRÉPON Age, salaire et productivité La productivité des salariés décline-t-elle en fin de carrière ?
G2003/07	H. BARON - P.O. BEFFY - N. FOURCADE - R. MAHIEU Le ralentissement de la productivité du travail au cours des années 1990
G2003/08	P.-O. BEFFY - B. MONFORT Patrimoine des ménages, dynamique d'allocation et comportement de consommation
G2003/09	P. BISCOURP - N. FOURCADE Peut-on mettre en évidence l'existence de rigidités à la baisse des salaires à partir de données individuelles ? Le cas de la France à la fin des années 90
G2003/10	M. LECLAIR - P. PETIT Présence syndicale dans les firmes : quel impact sur les inégalités salariales entre les hommes et les femmes ?
G2003/11	P.-O. BEFFY - X. BONNET - M. DARRACQ-PARIES - B. MONFORT MZE: a small macro-model for the euro area
G2004/01	P. AUBERT - M. LECLAIR La compétitivité exprimée dans les enquêtes trimestrielles sur la situation et les perspectives dans l'industrie
G2004/02	M. DUÉE - C. REBILLARD La dépendance des personnes âgées : une projection à long terme
G2004/03	S. RASPILLER - N. RIEDINGER Régulation environnementale et choix de localisation des groupes français
G2004/04	A. NABOULET - S. RASPILLER Les déterminants de la décision d'investir : une approche par les perceptions subjectives des firmes

G2004/05	N. RAGACHE La déclaration des enfants par les couples non mariés est-elle fiscalement optimale ?
G2004/06	M. DUÉE L'impact du chômage des parents sur le devenir scolaire des enfants
G2004/07	P. AUBERT - E. CAROLI - M. ROGER New Technologies, Workplace Organisation and the Age Structure of the Workforce: Firm-Level Evidence
G2004/08	E. DUGUET - C. LELARGE Les brevets accroissent-ils les incitations privées à innover ? Un examen microéconométrique
G2004/09	S. RASPILLER - P. SILLARD Affiliating versus Subcontracting: the Case of Multinationals
G2004/10	J. BOISSINOT - C. L'ANGEVIN - B. MONFORT Public Debt Sustainability: Some Results on the French Case
G2004/11	S. ANANIAN - P. AUBERT Travailleurs âgés, nouvelles technologies et changements organisationnels : un réexamen à partir de l'enquête « REPONSE »
G2004/12	X. BONNET - H. PONCET Structures de revenus et propensions différentes à consommer - Vers une équation de consommation des ménages plus robuste en prévision pour la France
G2004/13	C. PICART Évaluer la rentabilité des sociétés non financières
G2004/14	J. BARDAJI - B. SÉDILLOT - E. WALRAET Les retraites du secteur public : projections à l'horizon 2040 à l'aide du modèle de microsimulation DESTINIE
G2005/01	S. BUFFETEAU - P. GODEFROY Conditions de départ en retraite selon l'âge de fin d'études : analyse prospective pour les générations 1945 à 1974
G2005/02	C. AFSA - S. BUFFETEAU L'évolution de l'activité féminine en France : une approche par pseudo-panel
G2005/03	P. AUBERT - P. SILLARD Délocalisations et réductions d'effectifs dans l'industrie française
G2005/04	M. LECLAIR - S. ROUX Mesure et utilisation des emplois instables dans les entreprises
G2005/05	C. L'ANGEVIN - S. SERRAVALLE Performances à l'exportation de la France et de l'Allemagne - Une analyse par secteur et destination géographique
G2005/06	Bilan des activités de la Direction des Études et Synthèses Économiques - 2004
G2005/07	S. RASPILLER La concurrence fiscale : principaux enseignements de l'analyse économique
G2005/08	C. L'ANGEVIN - N. LAÏB Éducation et croissance en France et dans un panel de 21 pays de l'OCDE
G2005/09	N. FERRARI Prévoir l'investissement des entreprises

	Un indicateur des révisions dans l'enquête de conjoncture sur les investissements dans l'industrie.	G2006/10	C. AFSA L'estimation d'un coût implicite de la pénibilité du travail chez les travailleurs âgés	G2008/02	P. BISCOURP - X. BOUTIN - T. VERGÉ The Effects of Retail Regulations on Prices Evidence from the Loi Galland
G2005/10	P.-O. BEFFY - C. L'ANGEVIN Chômage et boucle prix-salaires : apport d'un modèle « qualifiés/peu qualifiés »	G2006/11	C. LELARGE Les entreprises (industrielles) françaises sont-elles à la frontière technologique ?	G2008/03	Y. BARBESOL - A. BRIANT Économies d'agglomération et productivité des entreprises : estimation sur données individuelles françaises
G2005/11	B. HEITZ A two-states Markov-switching model of inflation in France and the USA: credible target VS inflation spiral	G2006/12	Q. BIAU - N. FERRARI Théorie de l'opinion Faut-il pondérer les réponses individuelles ?	G2008/04	D. BLANCHET - F. LE GALLO Les projections démographiques : principaux mécanismes et retour sur l'expérience française
G2005/12	O. BIAU - H. ERKEL-ROUSSE - N. FERRARI Réponses individuelles aux enquêtes de conjoncture et prévision macroéconomiques : Exemple de la prévision de la production manufacturière	G2006/13	A. KOUBI - S. ROUX Une réinterprétation de la relation entre productivité et inégalités salariales dans les entreprises	G2008/05	D. BLANCHET - F. TOUTLEMONDE Évolutions démographiques et déformation du cycle de vie active : quelles relations ?
G2005/13	P. AUBERT - D. BLANCHET - D. BLAU The labour market after age 50: some elements of a Franco-American comparison	G2006/14	R. RATHELOT - P. SILLARD The impact of local taxes on plants location decision	G2008/06	M. BARLET - D. BLANCHET - L. CRUSSON Internationalisation et flux d'emplois : que dit une approche comptable ?
G2005/14	D. BLANCHET - T. DEBRAND - P. DOURGNON - P. POLLET L'enquête SHARE : présentation et premiers résultats de l'édition française	G2006/15	L. GONZALEZ - C. PICART Diversification, recentrage et poids des activités de support dans les groupes (1993-2000)	G2008/07	C. LELARGE - D. SRAER - D. THESMAR Entrepreneurship and Credit Constraints - Evidence from a French Loan Guarantee Program
G2005/15	M. DUÉE La modélisation des comportements démographiques dans le modèle de microsimulation DESTINIE	G2007/01	D. SRAER Allègements de cotisations patronales et dynamique salariale	G2008/08	X. BOUTIN - L. JANIN Are Prices Really Affected by Mergers?
G2005/16	H. RAOUI - S. ROUX Étude de simulation sur la participation versée aux salariés par les entreprises	G2007/02	V. ALBOUY - L. LEQUIEN Les rendements non monétaires de l'éducation : le cas de la santé	G2008/09	M. BARLET - A. BRIANT - L. CRUSSON Concentration géographique dans l'industrie manufacturière et dans les services en France : une approche par un indicateur en continu
G2006/01	C. BONNET - S. BUFFETEAU - P. GODEFROY Disparités de retraite de droit direct entre hommes et femmes : quelles évolutions ?	G2007/03	D. BLANCHET - T. DEBRAND Aspiration à la retraite, santé et satisfaction au travail : une comparaison européenne	G2008/10	M. BEFFY - É. COUDIN - R. RATHELOT Who is confronted to insecure labor market histories? Some evidence based on the French labor market transition
G2006/02	C. PICART Les gazelles en France	G2007/04	M. BARLET - L. CRUSSON Quel impact des variations du prix du pétrole sur la croissance française ?	G2008/11	M. ROGER - E. WALRAET Social Security and Well-Being of the Elderly: the Case of France
G2006/03	P. AUBERT - B. CRÉPON - P. ZAMORA Le rendement apparent de la formation continue dans les entreprises : effets sur la productivité et les salaires	G2007/05	C. PICART Flux d'emploi et de main-d'œuvre en France : un réexamen	G2008/12	C. AFSA Analyser les composantes du bien-être et de son évolution Une approche empirique sur données individuelles
G2006/04	J.-F. OUVRARD - R. RATHELOT Demographic change and unemployment: what do macroeconomic models predict?	G2007/06	V. ALBOUY - C. TAVAN Massification et démocratisation de l'enseignement supérieur en France	G2008/13	M. BARLET - D. BLANCHET - T. LE BARBANCHON Microsimuler le marché du travail : un prototype
G2006/05	D. BLANCHET - J.-F. OUVRARD Indicateurs d'engagements implicites des systèmes de retraite : chiffrages, propriétés analytiques et réactions à des chocs démographiques types	G2007/07	T. LE BARBANCHON The Changing response to oil price shocks in France : a DSGE type approach	G2009/01	P.-A. PIONNIER Le partage de la valeur ajoutée en France, 1949-2007
G2006/06	G. BIAU - O. BIAU - L. ROUVIERE Nonparametric Forecasting of the Manufacturing Output Growth with Firm-level Survey Data	G2007/08	T. CHANEY - D. SRAER - D. THESMAR Collateral Value and Corporate Investment Evidence from the French Real Estate Market	G2009/02	Laurent CLAVEL - Christelle MINODIER A Monthly Indicator of the French Business Climate
G2006/07	C. AFSA - P. GIVORD Le rôle des conditions de travail dans les absences pour maladie	G2007/09	J. BOISSINOT Consumption over the Life Cycle: Facts for France	G2009/03	H. ERKEL-ROUSSE - C. MINODIER Do Business Tendency Surveys in Industry and Services Help in Forecasting GDP Growth? A Real-Time Analysis on French Data
G2006/08	P. SILLARD - C. L'ANGEVIN - S. SERRAVALLE Performances comparées à l'exportation de la France et de ses principaux partenaires Une analyse structurelle sur 12 ans	G2007/10	C. AFSA Interpréter les variables de satisfaction : l'exemple de la durée du travail	G2009/04	P. GIVORD - L. WILNER Les contrats temporaires : trappe ou marchepied vers l'emploi stable ?
G2006/09	X. BOUTIN - S. QUANTIN Une méthodologie d'évaluation comptable du coût du capital des entreprises françaises : 1984-2002	G2007/11	R. RATHELOT - P. SILLARD Zones Franches Urbaines : quels effets sur l'emploi salarié et les créations d'établissements ?	G2009/05	G. LALANNE - P.-A. PIONNIER - O. SIMON Le partage des fruits de la croissance de 150 à 2008 : une approche par les comptes de surplus
		G2007/12	V. ALBOUY - B. CRÉPON Aléa moral en santé : une évaluation dans le cadre du modèle causal de Rubin	G2009/06	L. DAVEZIES - X. D'HAULTFOEUILLE Faut-il pondérer ?... Ou l'éternelle question de l'économètre confronté à des données d'enquête
		G2008/01	C. PICART Les PME françaises : rentables mais peu dynamiques		