



Groupe de Recherche en Économie et Développement International

Cahier de recherche / Working Paper
08-23

MODÈLE DYNAMIQUE EN PRÉSENCE DE BIAIS DE SÉLECTION ET
D'HÉTÉROGÉNÉITÉ INOBSERVÉE : APPLICATION À LA
PARTICIPATION DES FEMMES SUR LE MARCHÉ DU TRAVAIL AU
CANADA

Michael Doyon

Mario Fortin

**MODÈLE DYNAMIQUE EN PRÉSENCE DE BIAIS DE SÉLECTION
ET D'HÉTÉROGÉNÉITÉ INOBSERVÉE :
APPLICATION À LA PARTICIPATION DES FEMMES
SUR LE MARCHÉ DU TRAVAIL AU CANADA**

Michael Doyon¹ et Mario Fortin²

Résumé

Il est bien connu dans la littérature relative à l'offre de travail que négliger les phénomènes d'hétérogénéité individuelle inobservée et de biais de sélection dans la modélisation, peut conduire à une estimation biaisée et non convergente des paramètres d'intérêt. Afin de corriger ces problèmes, Kyriazidou (2001) propose une estimation en deux étapes, tenant compte de la présence de biais de sélection, d'hétérogénéité individuelle non observable et d'une structure autorégressive. Dans un premier temps, une estimation convergente des paramètres inconnus de l'équation de participation au marché du travail est obtenue. Dans un deuxième temps, les coefficients estimés sont intégrés à la construction d'une pondération Kernel appliquée aux instruments de l'estimateur GMM asymptotiquement normal et convergent à un taux près de \sqrt{n} .

Les résultats de l'estimateur GMM proposé sont comparés à un estimateur « naïf » et ceux-ci concluent en la présence significative d'un biais de sélection et d'une structure autorégressive dans l'équation de l'offre de travail des femmes au Canada. De plus, ignorer la présence de l'effet de sélection tend à sous-estimer l'impact des variables liées au ménage, soit le nombre d'enfants et le revenu de travail du conjoint, et à surestimer l'effet positif du salaire sur l'offre de travail.

Mots-clés : Sample selection, semiparametric models, labour supply, Generalized Method of Moments (GMM).

JEL : J21, J22

¹ Bureau de la Surveillance et de la Sécurité Financières du Mouvement des caisses Desjardins, Email : michaeldoyon@yahoo.com

² Département d'économie et GREDI, Université de Sherbrooke, 2500 Boul. de l'Université, Québec, Canada, J1K 2R1, Email : mario.fortin@usherbrooke.ca Cette étude a bénéficié des commentaires des participants au 47^e congrès de la SCSE et plus particulièrement ceux de Dany Brouillette. Nous tenons également à remercier Patrick Richard et Petr Hanel de l'université de Sherbrooke. Cette recherche a reçu le soutien financier du CIQSS.

Introduction

Il est bien connu dans la littérature relative à l'offre de travail que négliger les phénomènes d'hétérogénéité individuelle inobservée et de biais de sélection dans la modélisation, peut conduire à une estimation biaisée et non convergente des paramètres d'intérêts. Une équation définie pour un sous-ensemble non aléatoire d'une population dont les paramètres d'intérêts se réfèrent à la population totale implique la présence d'un biais de sélection. Ainsi, une équation d'offre de travail, dont le nombre d'heures travaillées n'est mesuré que pour un individu participant au marché du travail, conduit à une estimation non convergente des paramètres d'intérêts. Les méthodes usuelles d'estimation en données de panel ne s'appliquent pas à ce contexte et la littérature économétrique s'est considérablement développée à ce sujet au cours des dernières années. Nijman et Verbeek (1992) ainsi que Wooldridge (1995) proposent des méthodes se basant sur la paramétrisation du mécanisme de biais de sélection. Suivant l'approche familière de Heckman (1974), Kyriazidou (1997) propose une méthode d'estimation permettant de relâcher ces hypothèses.

Tout en privilégiant la présence de biais de sélection et d'hétérogénéité individuelle inobservée, une structure dynamique dans un cadre de données transversales et longitudinales est considérée dans notre analyse des déterminants de la participation au marché du travail et de l'offre de travail. D'un intérêt grandissant, des avancées récentes ont été faites dans la direction de modèles dynamiques linéaires en données de panel.¹ Par contre, la littérature théorique et empirique traitant des modèles dynamiques

¹ Voir Ahn et Schmidt (1995), et Arellano et Bover (1995).

linéaires avec biais de sélection demeure relativement pauvre. L'approche de Kyriazidou (2001) ouvre une voie théorique et propose une estimation par la méthode des moments généralisée (GMM) avec poids Kernel, et ce en deux étapes. Dans un premier temps, une estimation convergente des paramètres inconnus de l'équation de sélection est obtenue. Dans un deuxième temps, les coefficients estimés sont intégrés à la pondération Kernel de l'estimateur GMM asymptotiquement normal et convergent à un taux près de \sqrt{n} . L'objectif de cette étude est de démontrer l'impact de l'effet de sélection sur les coefficients d'intérêt estimés et la présence significative d'une structure dynamique dans l'équation d'offre de travail des femmes au Canada.

La première partie de cette étude présente la base de données retenue pour la création de notre échantillon, les caractéristiques de celui-ci et une description des variables considérées. En deuxième partie, nous définissons le cadre théorique du modèle proposé et les hypothèses sous-jacentes. La troisième section présente les résultats de notre modèle en insistant particulièrement sur les différences marquées entre le modèle dit « naïf », omettant la présence du biais de sélection, et l'estimateur Kernel proposé. Finalement, une conclusion permet un retour sur les résultats obtenus et une discussion des voies futures à considérer.

Base de données, variables et caractéristiques de l'échantillon

Notre échantillon est produit à partir de l'Enquête sur la dynamique du travail et du revenu (EDTR), enquête transversale et longitudinale, où chaque panel de six ans est un sous-ensemble de l'échantillon de l'Enquête sur la population active (EPA). L'EDTR

couvre environ 15 000 ménages et près de 31 000 répondants de 15 ans et plus. L'enquête se base principalement sur sept thèmes de recherches soit : la dynamique de l'emploi et du chômage; les transitions liées au cycle de vie observées sur le marché du travail; la qualité des emplois; la mobilité économique de la famille; la dynamique du faible revenu; les événements de la vie et changements familiaux; et, finalement, la poursuite des études et la combinaison travail-études. Nous travaillons avec le panel 1999-2004 et, afin de réduire les facteurs influant sur l'offre de travail, nous avons restreint notre échantillon aux femmes: (i) en couple et demeurant avec le même époux ou conjoint de fait durant l'enquête; (ii) âgées de 20 à 64 ans; (iii) ne faisant pas partie d'une minorité visible; (iv) n'ayant pas fréquenté un établissement scolaire au cours de l'enquête; (v) n'occupant pas un emploi autonome; (vi) n'étant pas à la retraite; et (vii) ayant des données complètes pour la période de six ans sur les variables présentées au tableau 3. Nous obtenons un échantillon final de 1634 individus pour la période de 1999-2004, totalisant 9 804 observations.

Les tableaux suivants dressent un portrait général de notre échantillon. La troisième colonne du tableau 1 rapporte le nombre de femmes (pourcentage) n'ayant pas déclaré un emploi rémunéré et ce, pour chaque année du panel. Le pourcentage des femmes ne participant pas au marché du travail décline dans le temps passant de 19,5% en 1999 à 15,9% en 2004. Le tableau 2 montre également que 10,6% des femmes déclarent ne pas avoir travaillé au cours de l'enquête. À l'opposé, 74,4% de notre échantillon est composé de femmes participant au marché du travail pour la totalité du panel. Finalement, 14,8% des femmes de notre échantillon ont vécu au moins une transition sur le marché du travail pour la durée totale de l'enquête.

Tableau 1. Participation au marché du travail, échantillon non pondéré

Années	Échantillon	Absentes du marché
1999	1634	319 (19,5)
2000	1634	312 (19,1)
2001	1634	295 (18,1)
2002	1634	277 (16,9)
2003	1634	264 (16,2)
2004	1634	260 (15,9)
Total	9804	1727 (17,6)

Tableau 2. Nombre d'années de travail, échantillon non pondéré

Nbr d'années	Fréquence
0	176 (10,8)
1	49 (3,0)
2	37 (2,3)
3	35 (2,1)
4	52 (3,2)
5	69 (4,2)
6	1216 (74,4)
Total	1634 (100)

L'objectif premier de cette étude étant l'intégration d'une structure dynamique et du phénomène de biais de sélection aux études empiriques existantes, la sélection de l'échantillon et le choix des variables influentes du modèle sont basés sur les travaux de Askildsen *et al.* (2003), de Dustmann et Rochina-Barrachina (2000), et de Miller (1991, 1992, 1993). Les variables explicatives retenues pour notre analyse reflètent les opportunités et les contraintes pouvant affecter l'offre de travail des femmes de notre échantillon. Celles-ci incluent une variable indicatrice de limitation d'activité et les variables affectant la propension au loisir soit, le revenu de travail du conjoint, les autres revenus du ménage, le salaire ainsi que le nombre et l'âge des enfants dans le ménage. Sous l'hypothèse de présence de contrainte de crédit, la stabilité financière du ménage peut être affectée négativement par une diminution du statut financier du ménage, soit une diminution du revenu du conjoint et/ou des autres revenus du ménage. Le revenu supplémentaire potentiel de la femme permettra d'accroître la stabilité financière du ménage, encourageant une augmentation du nombre d'heures travaillées. De plus, il est probable qu'une femme inactive sur le marché du travail réside dans un ménage dans lequel le revenu du conjoint est suffisamment élevé pour combler les besoins associés

au niveau de vie désiré.² La présence d'enfants d'âge préscolaire dans le ménage nécessite un investissement intensif en temps, pouvant limiter le nombre d'heures consacrées au travail pour une femme. Par contre, celle-ci peut désirer conserver ses liens avec le marché du travail, orientant ainsi ses choix vers d'autres solutions. Nous incluons la présence d'enfants d'âge préscolaire (0-5 ans) séparément dans notre modèle, car nous faisons l'hypothèse que l'effet négatif de cette variable est plus prononcé que celui de la présence d'enfants d'âge scolaire. De plus, comme le Québec a des politiques de conciliation travail-famille distinctives de celles des autres provinces, notamment par son réseau de garderies subventionnées, nous ajoutons une variable d'interaction entre les enfants d'âge préscolaire et la résidence au Québec. Pour un salaire plus élevé, nous faisons l'hypothèse qu'un individu réduira son loisir au profit du travail. En d'autres termes, nous posons l'hypothèse d'une offre de travail de pente positive. Les variables usuelles de l'éducation et de l'âge ne peuvent être considérées dans notre modèle dû à l'opérateur de différence première du logit conditionnel et du GMM. Finalement, Kyriazidou (2001) démontre que les instruments de l'estimateur GMM Kernel proposé ne peuvent capter les variables agrégées.

Le tableau 3 suivant présente une description des variables considérées dans notre analyse et leurs statistiques descriptives pour l'échantillon non pondéré et [pondéré]. Une pondération en probabilité est utilisée dans l'estimation des paramètres de notre modèle d'offre de travail, assurant ainsi la représentativité de la population canadienne. La pondération appliquée tient compte de l'aspect longitudinal de notre analyse et est construite à même l'EDTR. Les poids sont calculés à partir de l'EPA et sont ajustés de façon à correspondre à la population totale et au sous-échantillonnage de l'EDTR. Les

² Nakamura et Nakamura (1985) suggèrent que ce choix peut expliquer la relation négative entre le revenu du conjoint et l'offre de travail de la femme.

statistiques descriptives diffèrent d'un échantillon à l'autre bien que ces différences ne soient pas marquées. L'âge moyen dans notre échantillon est de 42 [42] ans et l'expérience moyenne est de 14,3 [14,7] années. Le nombre moyen d'heures travaillées rémunérées est de 1319,5 [1310,29] et le taux horaire composite moyen est de \$13,56 [\$14,36]. Le nombre moyen d'enfants d'âge préscolaire (0-5 ans) est de 0,28 [0,30] et le nombre moyen d'enfants d'âge scolaire (6-17 ans) est de 0,82 [0,82]. De plus, 17% [16%] des femmes de notre échantillon ont déclaré une limitation d'activité au cours de l'enquête et le taux d'emploi moyen provincial des femmes de 15 ans et plus est de 55,18% [55,97%]. Cette dernière variable, utilisée comme variable d'exclusion, est extraite de l'EPA et est intégrée à notre échantillon par le biais de la province de résidence des individus. Nous posons l'hypothèse que le taux d'emploi affecte la participation au marché du travail sans toutefois avoir un impact significatif sur le nombre d'heures travaillées par personne employée.³ L'expérience est comptabilisée en années équivalentes à temps plein et le salaire est défini d'après le taux de rémunération horaire implicite de tous les emplois comme travailleur rémunéré, pondéré selon le nombre total d'heures rémunérées pour chacun. Finalement, la variable « autres revenus » résulte de la somme du revenu du marché des deux conjoints moins leur revenu de travail, dont le résultat représente le revenu hors marché du ménage. Une femme est considérée comme participante au marché du travail si elle a déclaré un emploi rémunéré au cours de l'année de référence et l'offre de travail est mesurée par le nombre d'heures de travail rémunérées de tous les emplois.

³ Analogue à Askildsen et *al.* (2003), nous reconnaissons la difficulté de vérifier cette hypothèse en pratique et, par le fait même, la possibilité d'une faible identification.

Tableau 3. Description des variables, moyenne et (écart type)

Variable	Description	Échantillon non pondéré	Échantillon pondéré
Heures	Total d'heures rémunérées de tous les emplois	1319,51 (820,66)	1310,29 (816,69)
Travail	Indicateur; 1 si la personne a occupé un emploi rémunéré, 0 sinon	0,82 (0,38)	0,83 (0,38)
Salaire	Taux horaire composite de rémunération	13,56 (10,34)	14,36 (11,07)
Taux_emploi_15	Taux d'emploi; Femmes; 15 ans et plus; par provinces	55,18 (4,45)	55,97 (3,98)
Exp	Nombre d'années d'expérience de travail	14,28 (9,54)	14,71 (9,54)
Autres_revenus	Revenus hors marché du ménage	2373,07 (8682,53)	2407,44 (8763,14)
Revenu_conjoint	Revenu de travail du conjoint / 1000	45,05 (33,61)	49,79 (40,22)
Enfant_0_5	Nombre d'enfants d'âge préscolaire (0-5 ans) dans le ménage	0,28 (0,61)	0,30 (0,63)
Enfant_6_17	Nombre d'enfants d'âge scolaire (6-17 ans) dans le ménage	0,82 (1,01)	0,82 (1,03)
Age	Âge de la personne	42,74 (8,28)	42,80 (8,15)
Limitation	Indicateur; 1 si incapacité, 0 sinon	0,17 (0,37)	0,16 (0,37)
d99 - d04	Variables muettes temporelles; 1999-2004		
Enfant_0_5_Qc	Nombre d'enfants (0-5 ans) résidence au Québec, 0 sinon		
Age2	Âge de la personne au carré		
Exp2	Nombre d'années d'expérience de travail au carré		

Modèle économétrique

Nous considérons le modèle dynamique suivant en présence d'hétérogénéité inobservée et de biais de sélection :

$$y_{it}^* = \rho_0 y_{it-1}^* + x_{it} \beta_0 + \alpha_i + v_{it} + \lambda_{1it}; \quad i = 1, \dots, N; \quad t = 1, \dots, T \quad (1)$$

$$d_{it}^* = z_{it} \gamma_0 + \eta_i + u_{it} \quad (2)$$

$$d_{it} = 1 \text{ si } d_{it}^* > 0, \quad 0 \text{ sinon} \quad (3)$$

$$y_{it} = d_{it} y_{it}^* \quad (4)$$

L'équation (1) est le modèle d'offre de travail. Le nombre d'heures travaillées y_{it}^* par l'individu i à la période t est une variable latente observée si et seulement si la fonction d'indication $d_{it} = 1$, soit que l'individu i participe au marché du travail au temps t .

Les paramètres d'intérêts à estimer sont ρ_0, β_0 et γ_0 . Les termes $(\alpha_i$ et $\eta_i)$ représentent les effets fixes individuels inobservables. Pour leur part, x_{it} et z_{it} sont les vecteurs des variables explicatives strictement exogènes pouvant contenir des éléments communs,

$v_{it} \equiv \varepsilon_{it} - \lambda_{1it}$ ainsi que l'effet de sélection

$$\begin{aligned} \lambda_{1it} &\equiv E[\varepsilon_{it} \mid d_{it} = 1, \zeta_i] \\ &= E[\varepsilon_{it} \mid u_{it} \leq z_{it} \gamma_0 + \eta_i, \zeta_i] \\ &= \Lambda_{1it}(z_{it} \gamma_0 + \eta_i, \zeta_i), \end{aligned}$$

où $\Lambda_{1it}(\cdot)$ est une fonction inconnue et $\zeta_i \equiv (z_i, x_i, \alpha_i, \eta_i, y_{i0}^*, d_{i0})$ où $z_i = (z_{i1}, \dots, z_{iT})$ et $x_i = (x_{i1}, \dots, x_{iT})$. L'approche générale de Kyriazidou (2001) consiste à utiliser l'opération de différence première, sur la base de chaque individu, afin d'éliminer les termes α_i et λ_{1it} . L'auteur pose l'hypothèse H1 suivante :

$$\{(\varepsilon_{it}, u_{it})\}_{t=1}^T \text{ est i.i.d dans le temps } \forall i | \zeta_i,$$

une hypothèse de stricte stationnarité sur $(\varepsilon_{it}, u_{it})$. Pour un individu i avec $z_{it}\gamma_0 = z_{it-1}\gamma_0$ et $d_{it} = d_{it-1} = 1$, l'effet de sélection λ_{1it} sera constant pour ces deux périodes et éliminé par l'opération de différence première. Par H1 nous avons:

$$\lambda_{1it} = \lambda_{1it-1} \Rightarrow z_{it}\gamma_0 = z_{it-1}\gamma_0 \Rightarrow z_{it}\gamma_0 + \eta_i = z_{it-1}\gamma_0 + \eta_i.$$

Cette approche permet donc à l'effet du biais de sélection de varier d'un individu à l'autre tout en étant constant dans le temps. Considérant ainsi l'aspect dynamique du modèle et la présence de variables continues, Kyriazidou (2001) propose un estimateur GMM Kernel avec opérateur de différence première construit en deux étapes. En premier lieu, les paramètres de l'équation (2) sont estimés par un logit conditionnel.⁴ En second lieu, ces estimations sont utilisées dans la construction des poids Kernel appliqués aux instruments de l'estimateur GMM proposé.

⁴ Voir Chamberlain (1984) pour une description détaillée du logit conditionnel.

Par l'hypothèse H1, Kyriazidou (2001) démontre que les conditions d'orthogonalité de Ahn et Schmidt (1995) s'adaptent au modèle (1)-(4). Soit $\theta_0 \equiv (\rho_0, \beta_0')'$, $w_{it} = (x_{it}, y_{it}^*)$ et $\hat{\gamma}_n$ un estimateur convergent de γ_0 . Posons :

$$m_{1it,j}(\theta_0) \equiv d_{it}d_{it-1}d_{it-2}d_{it-j}y_{it-j}^* (\Delta y_{it}^* - \Delta w_{it}\theta_0), \quad t = 2, \dots, T; j = 2, \dots, t$$

$$m_{2it,a}(\theta_0) \equiv d_{it}d_{it-1}d_{it-2}x_{is,a} (\Delta y_{it}^* - \Delta w_{it}\theta_0), \quad t = 2, \dots, T; s = 1, \dots, T; a = 1, \dots, k$$

$$m_{3it}(\theta_0) \equiv d_{it}d_{it-1}d_{it-2}d_{it-1}d_{it-2} (y_{it}^* - w_{it}\theta_0) (\Delta y_{it}^* - \Delta w_{it}\theta_0), \quad t = 2, \dots, T-1$$

$$m_{4it}(\theta_0) \equiv d_{it}d_{it-1}d_{it-2} \left((y_{it}^* - w_{it}\theta_0)^2 - (y_{it-1}^* - w_{it-1}\theta_0)^2 \right), \quad t = 2, \dots, T.$$

Soit $m_{1it}(\theta_0)$ et $m_{2it}(\theta_0)$ les vecteurs des fonctions $m_{1it,j}(\theta_0)$ et $m_{2it,a}(\theta_0)$ respectivement. H1 implique les conditions d'orthogonalité suivantes :

$$E[m_{lit}(\theta_0) | \Delta z_{it}\gamma_0 = 0] = 0, \quad l = 1, \dots, 4 \quad (5)$$

et requiert des individus ayant $z_{it}\gamma_0 = z_{it-1}\gamma_0$ avec une probabilité de un, ce qui n'est pas le cas en présence de variables continues. Une stratégie possible est la pondération de chaque observation et ce inversement proportionnelle à la distance $|z_{it}\gamma_0 - z_{it-1}\gamma_0|$. Assumant une suridentification, l'estimateur proposé résout

$$\hat{\theta}_n = \arg \min \hat{G}_n(\theta)' A_n' A_n \hat{G}_n(\theta) \quad (6)$$

où A_n est une matrice stochastique $\xrightarrow{p} A_0$ une limite non stochastique finie, $A_n' A_n$ est une matrice de pondération définie positive, $\theta \equiv (\rho, \beta)$ et $\hat{G}_n(\theta)$ est le vecteur des moments dont les rangées sont de la forme

$$\frac{1}{n} \sum_{i=1}^n \frac{1}{h_n} K \left(\frac{\Delta z_{it} \hat{\gamma}_n}{h_n} \right) m_{it}(\theta), \quad l = 1, \dots, 4$$

où $h_n = hn^{\frac{-1}{2(r+1)+1}}$ avec r degré de lissage, $0 < h < \infty$, $h_n \rightarrow 0$ avec $n \rightarrow \infty$, et $K(\cdot)$ est la fonction de densité Kernel, satisfaisant les conditions de convergence et de normalité asymptotique de l'estimateur proposé.⁵ La fenêtre h_n est déterminée par la méthode « plug-in » de Horowitz (1992), sélectionnant la valeur de h qui minimise l'erreur quadratique moyenne asymptotique (AMSE) de l'estimateur. L'approche de Hansen (1982) est utilisée dans le choix de la matrice de pondération de l'estimateur GMM et la forme matricielle est employée.⁶ Cet estimateur asymptotiquement normal et convergent à un taux près de \sqrt{n} , permet une corrélation possible entre les effets fixes inobservés et les variables explicatives ainsi qu'une corrélation entre les effets fixes inobservés et les termes d'erreur. De plus, cet estimateur élimine le problème potentiel d'endogénéité dû aux erreurs de mesure. La nature non-paramétrique de cet estimateur requiert une restriction d'exclusion, impliquant qu'au moins une variable explicative présente dans l'équation de participation (2) soit exclue de l'équation d'offre de travail (1) et des instruments.

⁵ Voir Kyriazidou (1997, 2001) pour plus de détails.

⁶ Kyriazidou (2001) cite que le choix de la matrice de pondération n'a pas nécessairement les propriétés optimales du GMM standard, la variance asymptotique de l'estimateur étant affectée par le choix de h_n et de la fonction Kernel.

Résultats

Notre analyse des résultats porte une attention particulière sur la comparaison des paramètres estimés entre l'estimateur GMM Kernel proposé et l'estimateur GMM naïf, omettant la présence de biais de sélection. Nous commentons les résultats présentés au tableau 4 suivant, soit les paramètres estimés de l'équation de participation et les résultats obtenus de l'équation principale.

Tableau 4. Équation de participation.

Age2	-0,0083 ***	(0,0024)
Enfant_0_5	-1,4141 ***	(0,2271)
Enfant_6_17	-0,9107 ***	(0,2125)
Revenu_conjoint	-0,0071 *	(0,0040)
Limitation	-0,6671 ***	(0,2295)
Taux_emploi_15	0,1732 **	(0,0785)
Exp2	0,0230 ***	(0,0047)
Enfant_0_5_Qc	0,3413	(0,5555)
Autres_revenus	-6,3e-06	(9,5e-06)
d00	0,4973 *	(0,2758)
d01	1,3230 ***	(0,4400)
d02	1,9568 ***	(0,6330)
d03	2,6328 ***	(0,8342)
d04	3,1018 ***	(1,0298)
Log likelihood	-480,1369	
Nbr d'observations	1452	

Écart type entre parenthèses. *** et ** et * sont statistiquement différents de zéro au seuil de 1%, 5% et 10%, respectivement.

Le contexte théorique du logit conditionnel réduit notre échantillon à 1452 observations, soit 242 individus ayant déclaré au moins une transition sur le marché du travail pour la durée totale de l'enquête. L'échantillon restreint ne nous permet pas de travailler avec l'expérience décalée et le phénomène d'endogénéité de cette variable contemporaine est

réduit par la comptabilisation de l'expérience en équivalence années temps plein. Les paramètres d'intérêts estimés affectent la probabilité de participation dans le sens intuitif et la convexité de l'effet de l'âge est conforme à la littérature empirique, de même que la présence d'un effet temporel. Le contre effet de la présence d'enfants pour un ménage québécois n'est pas significatif, de même que les autres revenus du ménage. L'amplitude $\Delta z_{it} \hat{\gamma}_n$ est par la suite intégrée à l'estimateur GMM Kernel avec opérateur de différence première. Nous avons utilisé une fonction Kernel gaussienne et une fenêtre $h_n = hn^{-1/5} = 6$. Plusieurs expérimentations nous ont révélé que les paramètres estimés de l'équation principale (1) se stabilisaient pour $h_n \geq 6$, valeur inférieure à la fenêtre optimale calculée par la méthode « plug-in ». Par l'utilisation d'une fenêtre minimale, nous minimisons ainsi l'importance du biais possible en échantillon fini de l'estimateur GMM.⁷ La deuxième colonne du tableau 5 rapporte les paramètres de l'estimateur GMM naïf. Les paramètres estimés de l'estimateur GMM Kernel sont présentés dans la troisième colonne de ce même tableau.⁸

⁷ Kyriazidou (1997, 2001) et Horowitz (1992) suggèrent la possibilité d'un biais important en échantillon fini avec h_n qui augmente.

⁸ Les conditions d'orthogonalité $E[m_{it}(\theta_0) | \Delta z_{it} \gamma_0 = 0] = 0$ pour $l = 1, \dots, 3$ sont considérées. Les instruments ont été construits en tenant compte de l'endogénéité et de la censure du salaire. Les instruments passent le test de suridentification ($\chi^2(112) = 112,71$). Le test de suridentification (J-Test) de l'estimateur avec pondération Kernel suit une χ^2 non centrée causé par le biais asymptotique de l'estimateur. Kyriazidou (2001) suggère que ce biais est généralement petit et que le paramètre de non-centralité (NCP) est petit. L'utilisation de la distribution χ^2 centrée en présence d'un NCP grand biaiserait, dans le pire des cas, les résultats vers le rejet de l'hypothèse nulle. Voir Guvenen (2005).

Tableau 5. Équation de l'offre de travail

	GMM naïf	GMM Kernel
Heures ($t-1$)	0,10509 *** (0,0272)	0,08489 *** (0,0291)
Ln salaire	86,0314 *** (9,2531)	67,1676 *** (19,2027)
Enfant_0_5	-150,1803 *** (24,6639)	-199,3629 *** (27,7687)
Enfant_6_17	-38,4751 *** (14,7767)	-56,6438 *** (17,8693)
Limitation	-42,407 *** (16,2063)	-49,2346 *** (17,6545)
Enfant_0_5_Qc	89,8182 * (46,9768)	91,5901 ** (42,8870)
Revenu_conjoint	-0,2512 *** (0,1153)	-0,8242 *** (0,1902)
Nbr d'observations	6536	6536

Écart type entre parenthèses. *** et ** et * sont statistiquement différents de zéro au seuil de 1%, 5% et 10%, respectivement.

Le nombre d'heures travaillées au temps $t-1$ a un effet positif et significatif pour les deux estimateurs analysés, de même que l'effet du salaire⁹. Le revenu de travail du conjoint a un impact négatif sur le nombre d'heures travaillées des femmes, de même qu'une limitation d'activité et le nombre d'enfants d'âge scolaire et préscolaire dans le ménage. Ce dernier effet est toutefois diminué pour les ménages résidant au Québec, dû principalement, croyons nous, aux différentes politiques travail-famille, et en particulier aux garderies subventionnées par l'état qui réduisent le coût du travail des femmes.¹⁰ La présence du biais de sélection est vérifiée par un test d'Hausman, comparant les deux estimateurs. Celui-ci est concluant, nous avons un test statistique rejetant l'hypothèse nulle d'absence de biais de sélection ($\chi^2(7) = 21,13$). Sur la base de ce test,

⁹ Kyriazidou (2001) observe une détérioration des propriétés en échantillon fini de l'estimateur pour des paramètres autorégressifs s'approchant de l'unité.

¹⁰ La variable âge² s'est révélée instable dans l'estimation du GMM. Les effets temporels et autres revenus du ménage ne sont pas significatifs.

l'estimateur GMM naïf est biaisé et la prise en compte du biais de sélection a un impact significatif sur les paramètres estimés. L'effet positif du salaire diminue de 86,03 à 67,17 en présence de biais de sélection et l'impact de la variable dépendante décalée diminue de 0,11 à 0,09. L'impact négatif du nombre d'enfants d'âge préscolaire (-99,36) et d'âge scolaire (-56,64) est supérieur pour le GMM Kernel et le contre effet pour les ménages québécois demeure relativement stable pour les deux estimateurs. La différence entre les deux estimateurs est plus marquée pour la variable du revenu de travail du conjoint alors que l'impact de celle-ci est de trois fois supérieur pour le GMM Kernel, soit -0,83 comparativement à -0,25 pour le GMM naïf. L'impact d'une limitation d'activité va dans le sens espéré et est plus important pour l'estimateur Kernel.

Conclusion

Cette étude propose une application empirique de l'estimateur GMM Kernel proposé par Kyriazidou (2001), permettant de contrôler les phénomènes d'hétérogénéité individuelle inobservée et de biais de sélection. Les résultats obtenus concluent en la présence significative d'un biais de sélection et d'une structure dynamique dans l'équation de l'offre de travail des femmes au Canada. De plus, ignorer la présence de l'effet de sélection tend à sous-estimer l'impact des variables liées au ménage, soit le nombre d'enfants et le revenu du conjoint, et à surestimer l'effet positif du salaire.

Certaines améliorations peuvent être apportées à notre analyse, ouvrant ainsi la voie à des études empiriques futures. Tout d'abord, considérer l'impact du statut sur le marché du travail au temps $t-1$ sur la probabilité de participation au temps t contribuerait à

perfectionner notre modèle. Honoré et Kyriazidou (1997) proposent un estimateur allant dans ce sens. Ensuite, la méthode « plug-in » employée pour la construction des poids Kernel ne semble pas optimale dans le cas présent, dû au biais asymptotique très faible de notre estimateur. Une expérimentation des différentes méthodes proposées en littérature pourrait éventuellement améliorer la performance de notre modèle. Par ailleurs, les données de l'EDTR étant sélectionnées à partir d'un plan d'échantillonnage complexe, des variances bootstrap devraient être considérées dans les tests d'hypothèses des paramètres d'intérêts. Finalement, l'homoscédasticité des termes d'erreurs, imposée par l'hypothèse H1 de l'estimateur Kernel, n'est pas vérifiée dans notre analyse. Les applications empiriques de l'estimateur proposé par Kyriazidou (1997, 2001) évitent généralement cet aspect. Néanmoins, Dustmann et Rochina-Barrachina (2000) proposent l'estimateur du « maximum score » afin de vérifier cette hypothèse sur l'erreur de l'équation de sélection.

Bibliographie

- AHN, Seung C. and SCHMIDT, Peter, "Efficient Estimation of model for Dynamic Panel data", Journal of Econometrics, vol. 68, (1995), pp. 5-27.
- AHN, Seung C. and SCHMIDT, Peter, "Efficient Estimation of model for Dynamic Panel data: Alternative Assumptions and Simplified Estimation", Journal of Econometrics, vol. 76, (1997), pp. 309-321.
- AMEMIYA, Takeshi, Advanced Econometrics, Harvard University Press, Cambridge, Massachusetts, (1985).
- ARELLANO, Manuel and BOND, Stephen, "Some Tests of Specification for Panel Data: Monte Carlo Evidence and an Application to Employment Equations", Reviews of Economics Studies, vol. 58, (1991), pp. 277-297.
- ARELLANO, Manuel and BOVER, Olympia, "Another Look at the Instrumental Variables estimation of Error Components Models", Journal of Econometrics, vol. 68, (1995), pp. 29-51.
- ASKILDSEN, Jan Erik, BALTAGI, Badi H and HOLMAS, Tor Helge, "Wage Policy in the Health Care Sector: a Panel Data Analysis of Nurses' Labour Supply", Health Economics, vol. 12, (2003), pp.705-719.
- BALESTRA, Pietro and NERLOVE, Marc, "Pooling Cross section and Times-Series Data in the Estimation of a Dynamic Model : The Demand for Natural Gaz. ", Econometrica, vol. 34, (1966), pp. 585-612.
- BENITEZ-SILVA, Hugo, Labor Supply Flexibility and Portfolio Choice: An Empirical Analysis, Working Paper No. 2003-056, Michigan Retirement Research Center, University of Michigan, (2003).

- BLUNDELL, Richard and BOND, Stephen, "Initial Condition and Moments Restrictions in Dynamic Panel Data Models", Journal of Econometrics, vol. 87, (1998), pp. 115-143.
- CHAMBERLAIN, Gary, "Analysis of Covariance with Qualitative Data", Review of Economic Studies, vol. 47, (1980), pp. 225-238.
- CHAMBERLAIN, Gary, "Panel Data", Handbook of Econometrics, Volume II, Z. Griliches and M. Intriligator, Amsterdam : North Holland, chap.22, pp. 1247-1318, (1984).
- DUSTMANN, Christian and ROCHINA-BARRACHINA, Maria Engracia, Selection Correction in Panel Data Models: An Application to Labour Supply and Wages, Discussion Paper No. 163, Institut for the study of Labor, IZA, (2000).
- GREENE, William H., Econometric Analysis, Prentice Hall, New Jersey, (2003).
- GUVENEN, Fatih, Do Stockholders Share Risk More Effectively Than Non-stockholders?, Working Paper, Rochester Center for Economic Research, University of Rochester, (2005).
- HANSEN, Lars. Peter, "Large Sample Properties of Generalized Method of Moments Estimators", Econometrica, Vol. 50, (1982), pp. 1029-1054.
- HECKMAN, James J., "Shadow Prices, Market Wages, and Labour Supply", Econometrica, vol. 42, (1974), pp. 679-694.
- HONORÉ, Bo E., "Trimmed LAD and Least Squares Estimation of Truncated and Censored Regression Models with Fixed Effects", Econometrica, Vol. 60, (1992), pp. 533-565.

- HONORÉ, Bo E. and KYRIAZIDOU, Ekaterini, "Panel Data Discrete Choice Models with Lagged Dependent Variables", Econometrica, Vol. 68 No.4, (2000), pp. 839-874.
- HOROWITZ, Joel L., "A Smoothed Maximum Score Estimator for the Binary Response Model", Econometrica, Vol. 60, (1992), pp. 505-531.
- KIVIET, Jan. F, "On Bias Inconsistency and Efficiency of Some Estimators in Dynamic Panel Models", Journal of Econometrics, Vol. 68, (1995), pp. 53-78.
- KYRIAZIDOU, Ekaterini, "Estimation of Panel Data Sample Selection Models", Econometrica, Vol. 65, (1997), pp. 1335-1364.
- KYRIAZIDOU, Ekaterini, "Estimation of Dynamic Panel Data Sample Selection Models", Review of Economic Studies, Vol. 68, (2001), pp. 543-572.
- MARRON, J.S., "Automatic Smoothing Parameter Selection", Empirical Economics, Vol. 13, (1988), pp. 187-208.
- MILLER, Carole F., , "Married Womens' Tastes and thhe Decision to Participate in the Labour Market : Result from a Fixed Effect Model", Applied Economics, Vol. 23, (1991), pp. 1499-1509.
- MILLER, Carole F., , "Experience and the Decision to Participate in the Labour Market : Result from a Conditional Logit Model", Journal of Economic Studies, Vol. 19, (1992), pp. 3-21.
- MILLER, Carole F., , "Part-time Participation Over the Life Cycle Among Married Women who Work in the Market", Applied Economics, Vol. 25, (1993), pp. 91-99.
- NAKAMURA, A. and NAKAMURA, M., "The Second Paycheck : A Socioeconomic Analysis of Earning", Academic Press, New York, (1985)

- NERLOVE, Marc, "Further Evidence on the Estimation of Dynamics Economic Relations from a Time Series of Cross-Sections", Econometrica, Vol. 39, (1971), pp. 359-382.
- NICKELL, Stephen, "Biases in Dynamic Model with Fixed Effects", Econometrica, Vol. 49, (1981), pp. 1417-1426.
- NIJMAN, Theo E. and VERBEEK, Marno, "Testing for Selectivity Bias in Panel Data Models", International Economic Reviews, vol. 33, (1992), pp. 681-703.
- POWELL, James L., Semiparametric Estimation of Bivariate Latent Variable Models, Working Paper No. 8704, Social Systems Research Institute, University of Wisconsin-Madison, (1987).
- WOOLDRIDGE, Jeffrey M., "Selection Correction for Panel Data Model under Conditional Mean Independence Assumption", Journal of Econometrics, pp. 115-132, (1995).
- WOOLDRIDGE, Jeffrey M., Econometric Analysis of Cross Section and Panel Data, The MIT Press, Cambridge, Massachusetts, (2002).