UNIVERSITÉ DU QUÉBEC À MONTRÉAL

SERVICES DE GARDE ET OFFRE DE TRAVAIL DES MÈRES : EFFETS DYNAMIQUES SUR LE REVENU DE TRAVAIL AVEC EFFETS DE SÉLECTION ET ENDOGÉNÉITÉ DE L'EXPÉRIENCE

MÉMOIRE
PRÉSENTÉ
COMME EXIGENCE PARTIELLE
DE LA MAÎTRISE EN ÉCONOMIQUE

PAR
NICHOLAS LABELLE ST-PIERRE

NOVEMBRE 2008

REMERCIEMENTS

Je tiens d'abord à remercier tout particulièrement mes directeurs, Pierre Lefebvre et Philip Merrigan, pour la patience et l'intérêt dont ils ont fait preuve durant ce difficile, mais instructif labeur. Le temps et l'énergie qu'ils consacrent à leurs étudiants m'impressionneront toujours.

Ensuite, je n'aurais pu me concentrer à cette tâche à ce point sans le soutien financier de l'équipe du Fonds québécois de la recherche sur la société et la culture (FQRSC) des professeurs Pierre Lefebvre, Sylvana Côté, Christa Japel et Philip Merrigan. De même, les bourses de RBC Banque Royale et du Centre interuniversitaire québécois de statistiques sociales (CIQSS) me furent d'un grand secours. Enfin, je souligne le dévouement exceptionnel des analystes du CIQSS sans qui des technicalités de tout ordre seraient devenues de véritables obstacles à ma recherche. Je mentionnerai entre autres Danielle Forest et Franck Larouche.

REMARQUES

Les estimations de ce travail peuvent être influencées par le plan d'enquête. Les inférences à la population ne sont pas valides puisque les résultats ne sont pas pondérés.

Bien que la recherche et les analyses soient fondées sur des données de Statistique Canada, les opinions exprimées ne représentent pas celles de Statistique Canada.

TABLE DES MATIÈRES

REMARQUES iii

LISTE DES FIGURES vi

LISTE DES TABLEAUX viii

LISTE DES ABRÉVIATIONS x

RÉSUMÉ xii

INTRODUCTION 1

CHAPITRE I
REVUE DE LITTÉRATURE 3
1.1 Impact de la maternité sur le salaire 3

1.2 Impact de la politique de garde du Québec sur l'offre de travail des mères 12

1.3 Expérience et gains salariaux 17

CHAPITRE II
APPROCHE ÉCONOMÉTRIQUE 21
2.1 Problèmes économétriques liés à la régression salariale 21

2.2 Méthodes économétriques d'estimation du rendement de l'expérience 24

2.3 Détection du biais de sélection 35

2.4 Correction du biais de sélection 39

2.5 Utilisation de la forme réduite pour estimer l'effet de la politique d'intérêt 42

2.6 Tests de la faiblesse et de la validité des instruments 44

CHAPITRE III
APPLICATION ET RÉSULTATS 51

3.1 Application des modèles 51
3.2 Analyse descriptive 59
3.3 Résultats 72
3.4 Discussion 94

CONCLUSION 98

ANNEXE A
PRINCIPALES DISPOSITIONS CONCERNANT LES SERVICES DE GARDE ET L'ÉDUCATION PRÉSCOLAIRE AU CANADA 100

ANNEXE B
VARIANCE DE L'ESTIMATEUR DE LA RÉGRESSION INSTRUMENTALE EN DEUX ÉTAPES CORRIGEANT POUR LE BIAIS DE SÉLECTION 103

ANNEXE C
VARIANCE DE L'ESTIMATEUR DE LA RÉGRESSION CORRIGEANT POUR LE BIAIS DE SÉLECTION 106

ANNEXE D
SPÉCIFICATION ET VARIABLES UTILISÉES 108

ANNEXE E
PROGRAMME MATA 110

RÉFÉRENCES 135

LISTE DES FIGURES

Figure		Page
1.1	Écart de salaire entre les sexes en fonction de la maternité pour différents pays.	5
1.2	Effets salariaux potentiels de la maternité	8
1.3	Profils salariaux du secteur public pour des femmes ayant de 15 à 16 années de scolarité	8
1.4	Profils salariaux du secteur privé pour des femmes ayant de 15 à 16 années de scolarité	9
1.5	Salaires prédits d'hommes sans diplôme collégial selon le SIPP de 1996 aux États-Unis	18
3.1	Proportion des femmes de l'échantillon ayant déjà été exposées à la politique de services de garde à partir de 1999	56
3.2	Taux de participation de toutes les femmes de l'échantillon au mois d'avril au Québec, en Ontario et dans les autres provinces (RdC)	62
3.3	Taux de participation de toutes les femmes de l'échantillon au mois d'août au Québec, en Ontario et dans les autres provinces (RdC)	62
3.4	Taux de participation des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au mois d'avril au Québec (Qc) et dans les autres provinces (RdC)	63
3.5	Taux de participation des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au mois d'août au Québec (Qc) et dans les autres provinces (RdC)	63
3.6	Heures annuelles moyennes travaillées par toutes les femmes de l'échantillon au Québec, en Ontario et dans les autres provinces (RdC)	65
3.7	Heures annuelles moyennes travaillées par les mères d'au moins 1 enfant de 0 à 17 ans et par les femmes sans enfant de l'échantillon au Québec (Qc) et dans les autres provinces (RdC)	65

Figure Page

3.8 Semaines annuelles moyennes travaillées par toutes les femmes au Québec, en Ontario et dans les autres provinces (RdC) 66

3.9 Semaines annuelles moyennes travaillées par les mères d'au moins 1 enfant de 0 à 17 ans et par les femmes sans enfant de l'échantillon au Québec (Qc), en Ontario et dans les autres provinces (RdC) 66

3.10 Revenu de travail annuel moyen de toutes les femmes au Québec, en Ontario et dans les autres provinces (RdC) (variable *earng42* de l'EDTR corrigée pour l'inflation) 67

3.11 Revenu de travail annuel moyen des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au Québec (Qc) et dans les autres provinces (RdC) 67

3.12 Expérience moyenne en années des femmes de l'échantillon au Québec, en Ontario et dans les autres provinces (RdC) 70

3.13 Expérience moyenne en années des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au Québec (Qc) et dans les autres provinces (RdC) 70

3.14 Salaire horaire moyen des femmes de l'échantillon au Québec, en Ontario et dans les autres provinces (RdC) 71

3.15 Salaire horaire moyen des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au Québec (Qc) et dans les autres provinces (RdC) 71

A.1 Évolution du nombre de places en services de garde à contribution réduite au Québec 102

LISTE DES TABLEAUX

Tableau		Page
1.1	Effets du mariage et des enfants sur le logarithme du salaire horaire de femmes travaillant à temps plein à partir d'une régression MCO	6
1.2	Effets de la durée d'occupation de l'emploi et de l'expérience	20
3.1	Structure et taille des échantillons de l'EDTR	52
3.2	Structure et taille de l'échantillon de cette étude	53
3.3	Répartition de la variable dichotomique de politique de services de garde au Québec selon l'âge des enfants et l'année d'exposition	56
3.4	Comparaison de statistiques descriptives pour l'échantillon de cette étude et de Semykina et Wooldridge (2006, p. 34)	60
3.5	Estimés du logarithme du salaire horaire réel de Semykina et Woolridge (2006, p. 35)	78
3.6	Estimés du logarithme du salaire horaire réel avec notre échantillon et une spécification semblable à celle Semykina et Wooldridge (2006)	78
3.7	Tests des instruments de nos modèles avec une spécification similaire à celle de Semykina et Wooldridge (2006)	79
3.8	Estimations pour le logarithme du salaire réel horaire avec la spécification incluant les variables de politique	80
3.9	Tests de Hausman	82
3.10	Premières étapes des modèles MCDE et EF-MCDE	83
3.11	Premières étapes du modèle EA-VI et de la procédure #1	85
3.12	Premières étapes de la procédure #2	87
3.13	Tests des instruments de nos modèles avec la spécification contenant les variables de politique	90

Tableau		Page
3.14	Estimations pour le logarithme du salaire horaire réel en forme réduite avec la spécification contenant les variables de politique	91
A.1	Introduction des services de garde au Québec	100
A.2	Services de garde à l'extérieur du Québec	101
D.1	Spécification de l'échantillon	108
D.2	Variables de l'EDTR utilisées pour les modèles	109

LISTE DES ABRÉVIATIONS, SIGLES ET ACRONYMES

CIP	*Consumer Price Index*
CPS	*Current Population Survey*
EA	Effets aléatoires
EA-VI	Effets aléatoires avec variables instrumentales
EDTR	Enquête sur la dynamique du travail et du revenu
EF	Effets fixes
EF-MCDE	Effets fixes des moindres carrés en deux étapes
ELNEJ	Enquête longitudinale nationale sur les enfants et les jeunes
EPA	Enquête sur la population active
HRS	*Health and Retirement Survey*
IABS	*Institut für Arbeitsmarkt und Berufsforschung Sample*
LIS	*Luxembourg Income Study*
MCO	Moindres carrés ordinaires
MCDE	Moindres carrés en deux étapes
NLSY	*National Longitudinal Survey of Young Women*
NLS-YM	*National Longitudinal Surveys of Young Men*
NLS-YW	*National Longitudinal Surveys of Young Women*

P#1	Procédure #1
P#2	Procédure #2
PSID	*Panel Study of Income Dynamics*
RDC	Reste du Canada
SIPP	*Survey of Income and Program Participation*
VI	Variables instrumentales

RÉSUMÉ

Depuis septembre 1997, le gouvernement québécois subventionne la différence entre la contribution parentale et le coût de services de garde universels et publics. Lefebvre et Merrigan (2005a, 2005b, 2007a, 2007b) de même que Baker *et al.* (2005) ont montré que cette politique avait augmenté l'offre de travail des mères. Ce travail tente de vérifier si cette expérience supplémentaire s'est matérialisée par une prime salariale chez les mères québécoises par rapport au reste du Canada.

La procédure de Semykina et Wooldrige (2006) fut choisie pour estimer le rendement de l'expérience sur le salaire à l'aide de l'Enquête sur la dynamique du travail et du revenu (EDTR) de 1993 à 2004. Sous certaines hypothèses, les auteurs contournent plusieurs problèmes économétriques. D'abord, l'expérience est endogène, c'est-à-dire qu'elle découle d'une panoplie d'autres variables. Ensuite, elle peut dépendre de facteurs non observés comme le talent et la motivation au travail, un phénomène désigné par l'hétérogénéité non observée. Puisque l'échantillon ne contient que des femmes qui occupent un emploi, un biais de sélection peut aussi affecter la validité des résultats. La méthode des auteurs détecte et corrige ce biais.

Les instruments contrôlant pour l'endogénéité comptent entre autres la variable de politique de services de garde puisque l'étude fait l'hypothèse qu'elle augmente l'expérience sans être corrélée avec le salaire. De plus, des tests détectent les instruments faibles ou invalides. Enfin, une régression de forme réduite remplace les variables endogènes par leurs instruments afin d'isoler l'effet de la politique sur le salaire.

Les résultats montrent que les femmes du Québec ont profité d'une augmentation de leur expérience ce qui a accru leur salaire, mais des problèmes liés à la conception de la variable d'expérience, à la faiblesse et à l'invalidité des instruments empêchent d'isoler cet effet pour les mères québécoises ayant été exposées à la politique de services de garde.

Mots clés : données de panel, endogénéité, expérience, services de garde.

INTRODUCTION

Depuis l'après-guerre et l'essor phénoménal du taux de participation des femmes sur le marché du travail, l'égalité de traitement entre les hommes et les femmes n'a cessé d'occuper les débats publics. Or, les femmes ont des enfants et même en l'absence de discrimination, la maternité peut avoir de sérieux impacts sur leur capital humain et par conséquent sur leur salaire.

Au Québec, septembre 1997 marque le début de l'implantation des garderies à 5 $ par jour par enfant. Le gouvernement québécois subventionne la différence entre les coûts de fonctionnement de services de garde universels et publics et la contribution parentale. Sachant que cette politique a augmenté l'offre de travail des mères québécoises (Lefebvre et Merrigan, 2005a, 2005b, 2007a, 2007b; Baker *et al.*, 2005), l'expérience supplémentaire ainsi acquise devrait selon la théorie du capital humain se manifester par une prime salariale chez ces femmes par rapport au reste du Canada. Ce travail tente de vérifier cette hypothèse.

Pour ce faire, une régression salariale permettra d'évaluer à la marge l'impact de l'expérience sur le salaire et donc si oui ou non l'offre de travail accrue des mères québécoises a augmenté leur salaire. Cependant, l'expérience présente une nature endogène notoire dans la littérature économique. Autrement dit, elle dépend d'une foule d'autres facteurs dont l'éducation, l'âge ou le nombre d'enfants. Elle peut aussi être corrélée avec plusieurs paramètres non observés dont la motivation au travail et l'habileté. Ce phénomène se nomme hétérogénéité non observée. Il peut aussi y avoir un biais de sélection dans l'échantillon puisque les observations de la régression

ne comportent que des femmes qui travaillent et qui détiennent donc certaines caractéristiques particulières invalidant possiblement une généralisation des résultats pour celles sans emploi.

Ainsi, la procédure de Semykina et Wooldridge (2006) résout sous certaines hypothèses l'ensemble de ces problèmes et sera appliquée à l'Enquête sur la dynamique du travail et du revenu (EDTR) de 1993 à 2004. Cette étude traite de l'endogénéité dans un contexte de données de panel et de l'hétérogénéité non observée. Les auteurs y développent aussi une méthode de détection et de correction du biais de sélection. Les instruments qui corrigent pour l'endogénéité de l'expérience compteront entre autres la politique de services de garde qui augmente l'expérience sans être corrélée avec le terme d'erreur et donc avec le salaire. Divers tests décèleront aussi des instruments faibles ou invalides pour toutes nos estimations. Pour terminer, une régression de forme réduite remplaçant les variables endogènes par leurs instruments isolera l'effet de la politique et validera ou non notre hypothèse.

Par conséquent, le premier chapitre présente le fil conducteur de ce travail par une revue de la littérature économique pertinente. Ensuite, le deuxième chapitre explique en détail la méthode proposée par Semykina et Wooldridge (2006) et les tests d'instruments employés. Enfin, le troisième chapitre expose et commente les résultats obtenus grâce à notre démarche.

CHAPITRE I

REVUE DE LA LITTÉRATURE

Ce chapitre permet de situer notre recherche en passant en revue les études pertinentes. D'abord, la littérature économique relève la présence d'un écart salarial entre les mères et leurs consoeurs n'ayant pas d'enfant *ceteris paribus*. Malgré plusieurs contrôles statistiques, cette différence demeure significative. Toutefois, la perte d'expérience due à l'interruption du travail semble jouer un rôle majeur.

Ensuite, d'autres études montrent que la politique de services de garde du Québec s'est avérée efficace à réduire l'écart de l'offre de travail des mères québécoise par rapport à celles du reste du Canada. Selon la théorie du capital humain, cette expérience supplémentaire devrait entraîner un rattrapage similaire du salaire des mères québécoises. De ce fait, notre étude vise à démontrer si l'expérience supplémentaire des mères imputable à la politique de garde du Québec s'est concrétisée par des augmentations salariales.

1.1 Impact de la maternité sur le salaire

Les économistes reconnaissent généralement l'existence d'un écart salarial entre les mères et les femmes sans enfant (*family gap*) toute chose étant égale par ailleurs. Ainsi, Waldfogel (1997) considère trois hypothèses à l'aide du *National Longitudinal Survey of Young Women* (NLSY) de 1968 à 1988 pour des femmes américaines entre 34 et 44 ans. Premièrement, elle mentionne la théorie du capital humain de Becker (1985) qui considère l'expérience comme la raison principale de cet écart puisque les mères passent moins de temps sur le marché du travail (Waldfogel, 1997, p. 209).

Pourtant, un écart salarial entre les mères et les femmes sans enfant subsiste même en contrôlant pour l'expérience. Deuxièmement, l'hétérogénéité non observée entre les femmes dont la motivation au travail ne détient pas un pouvoir explicatif important dans les modèles en différence de Waldgofel (1997, p. 213) même s'il réduit l'impact du statut marital soit marié, divorcé et célibataire. Troisièmement, Waldfogel (1997, p. 215) note que le travail à temps partiel joue un rôle non négligeable avec une pénalité supérieure à 10 % par rapport au travail à temps plein. Enfin, malgré tous ces facteurs, un écart inexpliqué de 4 % pour un enfant et de 12 % pour un deuxième enfant persistent. Ces pénalités sont moins fortes pour les afro-américaines et croissent avec le niveau de scolarité (Waldfogel, 1997, p. 216). Elle conclut en soupçonnant la discrimination des employeurs et les parcours plus difficiles des mères sur le marché du travail comme causes plausibles.

Dans un contexte temporel, Waldfogel (1998, p. 137) remarque que l'écart de salaire entre les sexes s'est rétréci au cours des deux dernières décennies alors que celui entre les mères et leurs consoeurs sans enfant s'est agrandi aux États-Unis. Trois bases de données lui servent pour un exemple : le *National Longitudinal Surveys of Young Women* (NLS-YW) et celui pour les jeunes hommes (NLS-YM) de 1980, puis le NLSY de 1991. Ainsi, les femmes sans enfant gagnaient 66,2 % du salaire des hommes en 1980 et 90,1 % en 1991 tandis que les mères obtenaient respectivement 56,3 % et 72,6 %. Ceci porte l'écart salarial maternel de 9,9 % à 17,5 % en 11 ans (Waldfogel, 1998, p. 145). De même, à l'aide d'une régression par moindres carrés ordinaires (MCO) contrôlant pour diverses variables socioéconomiques, la pénalité salariale d'un enfant était de 4,47 % en 1980 et 9,79 % en 1991 (Waldfogel, 1998, p. 146). Ce phénomène laisse croire que ces deux écarts dépendent de dynamiques distinctes.

Au niveau international, l'impact de la maternité varie cependant beaucoup autant au niveau du taux de participation que du salaire. Harkness et Waldfogel (1999) examinent le cas de sept pays industrialisés : l'Australie (1994), le Canada (1994), la Grande-Bretagne (1995), les États-Unis (1994), l'Allemagne (1994), la Finlande (1991) et la Suède (1991). La base de données *Luxembourg Income* Study (LIS) regroupent des microdonnées de divers pays et a servi à leur étude qui inclut des individus entre 24 et 44 ans. Elles trouvent entre autres que plus la maternité diminue le taux de participation des mères pour un pays donné, plus elle réduit leur salaire et plus l'écart général entre les sexes s'accroît (Harkness et Waldfogel, 1999, p. 1). Cette relation s'exprime en partie dans la figure 1.1 construite à partir des données brutes de Harkness et Waldfogel (1999, p. 29).

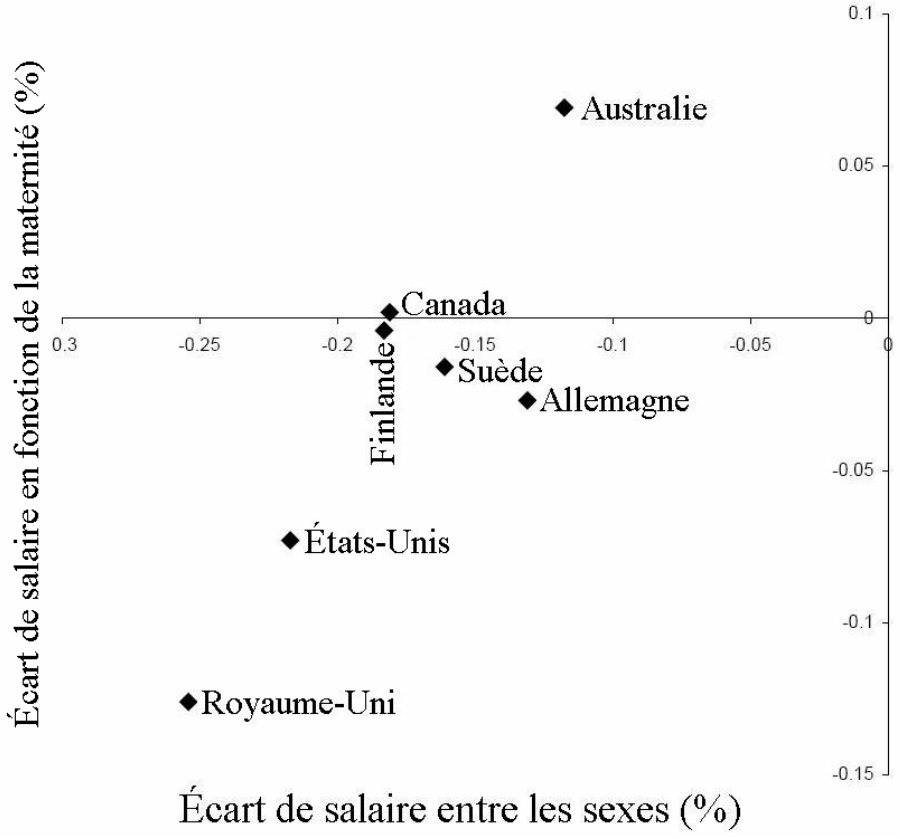


Figure 1.1 Écart de salaire entre les sexes en fonction de la maternité pour différents pays. (Tirée de Harkness et Waldfogel, 1999, p. 29.)

Tableau 1.1 Effets du mariage et des enfants sur le logarithme du salaire horaire de femmes travaillant à temps plein à partir d'une régression MCO.
(Tiré de Harkness et Waldfogel, 1999, p. 23.)

	AU 94	**CN 94**	**RU 95**	**ÉU 94**	**AL 94**	**FI 91**	**SU 91**
Mariée	,028	-,004	,018	,075**	,002	,024	,037**
	(,036)	(,016)	(,030)	(,010)	(,030)	(,015)	(,018)
1 enfant	-,122**	-,058**	,003	-,025**	-,032	-,054**	-,031
	(,046)	(,017)	(,036)	(,012)	(,035)	(,016)	(,021)
2 enfants	-,167**	-,065**	-,176**	-,048**	-,080**	-,034**	-,056**
	(,048)	(,018)	(,037)	(,012)	(,041)	(,017)	(,021)
3 enfants	-,244**	-,203**	-,295**	-,102**	-,134**	-,082**	-,102**
	(,073)	(,025)	(,054)	(,016)	(,078)	(,022)	(,027)
N	1046	7885	1002	11588	1107	3385	685
R^2 ajustée	,0934	,1444	,3125	,2678	,1674	,3037	,1996

* indique un niveau de significativité à p<0,10
** indique un niveau de significativité à p<0,05

Mais, en contrôlant pour divers facteurs liés au salaire tels que l'âge, l'éducation, le nombre d'enfants, la région et l'ethnicité, les résultats deviennent un peu plus probants comme le montre le tableau 1.1. Il faut souligner toutefois que les R^2 sont très faibles. Ni l'expérience ni la durée d'occupation d'emploi ne purent être incluses faute de données. En plus, l'analyse est uniquement transversale soit pour une année donnée (Harkness et Waldfogel, 1999, p. 9). La correction du biais de sélection à la Heckman (1979) qui sera expliquée plus loin n'a pas non plus affecté les résultats (Harkness et Waldfogel, 1999, p. 11). Cependant, cette étude permet quand même de comparer la pénalité salariale maternelle et de vérifier sa présence à travers divers pays. Dans l'ensemble, le Canada se situe entre les deux extrêmes du Royaume-Uni et des pays de la péninsule scandinave.

Par ailleurs, en misant sur le clivage public privé des politiques familiales, Nielsen, Simonsen et Verner (2002) testent l'hypothèse selon laquelle la pénalité salariale de la maternité proviendrait de l'auto sélection des femmes entre les deux secteurs en fonction de leurs anticipations rationnelles. Ils appliquent le modèle de choix endogène de secteur suggéré par Lee (1978) à une base de données transversales danoise de 1997 de femmes entre 20 et 40 ans employées au moins 200 heures par année. De cette façon, une femme prévoyant avoir un enfant choisirait le secteur public alors que l'autre sans enfant préférerait les rémunérations supérieures du secteur privé. Le fossé sectoriel s'est d'autant plus creusé que le salaire réel du secteur public a diminué pour une année sur deux depuis 1985 au Danemark, alors que les conditions de travail se sont constamment améliorées (Nielsen, Simonsen et Verner, 2002, p. 7). Avant d'aborder les résultats de leur étude, Nielsen, Simonsen et Verner (2002, p. 18-19) traitent des différents impacts salariaux possibles pour une mère. Leurs modèles permettent de tenir compte des quatre possibilités (*voir* figure 1.2). La mère peut ne pas connaître de pénalité ou alors simplement subir une perte d'expérience. Par contre, elle peut aussi subir une dépréciation du capital humain qui sera ou non rattrapée plus tard.

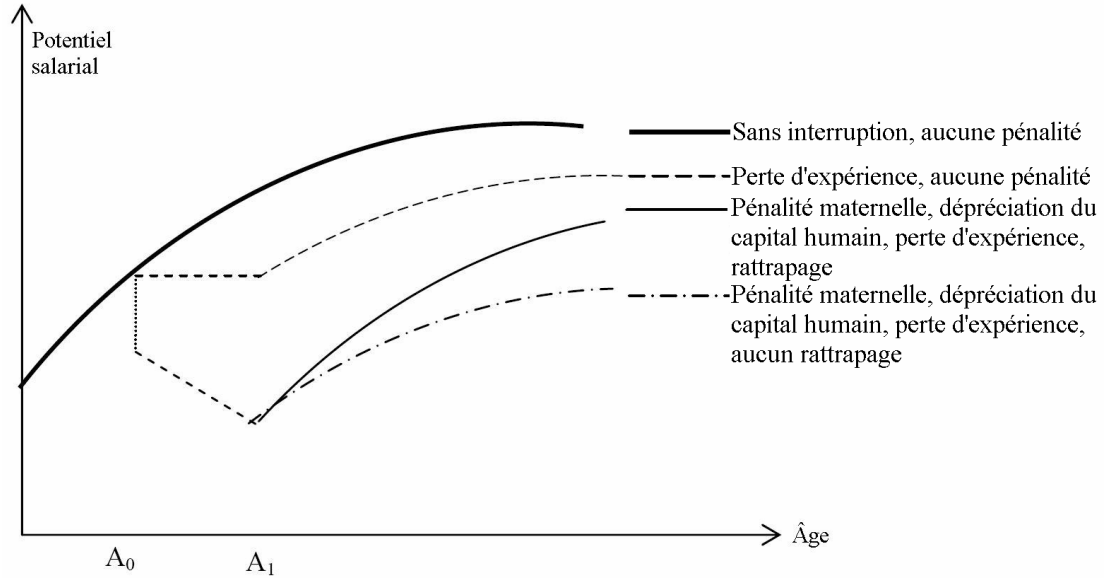


Figure 1.2 Effets salariaux potentiels de la maternité. (Tirée de Nielsen, Simonsen et Verner, 2002, p. 18.)

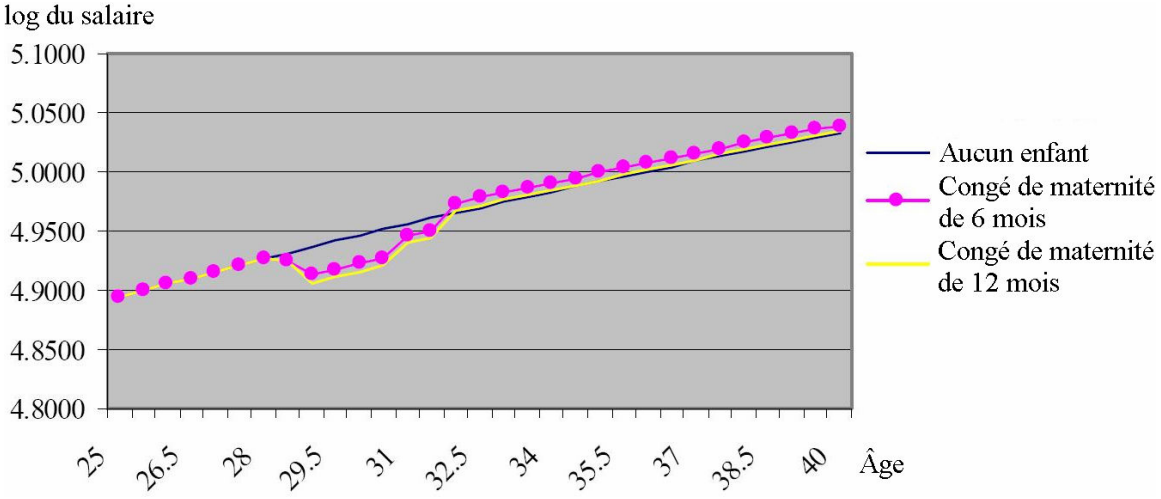


Figure 1.3 Profils salariaux du secteur public pour des femmes ayant de 15 à 16 années de scolarité. (Tirée de Nielsen, Simonsen et Verner, 2002, p. 28.)

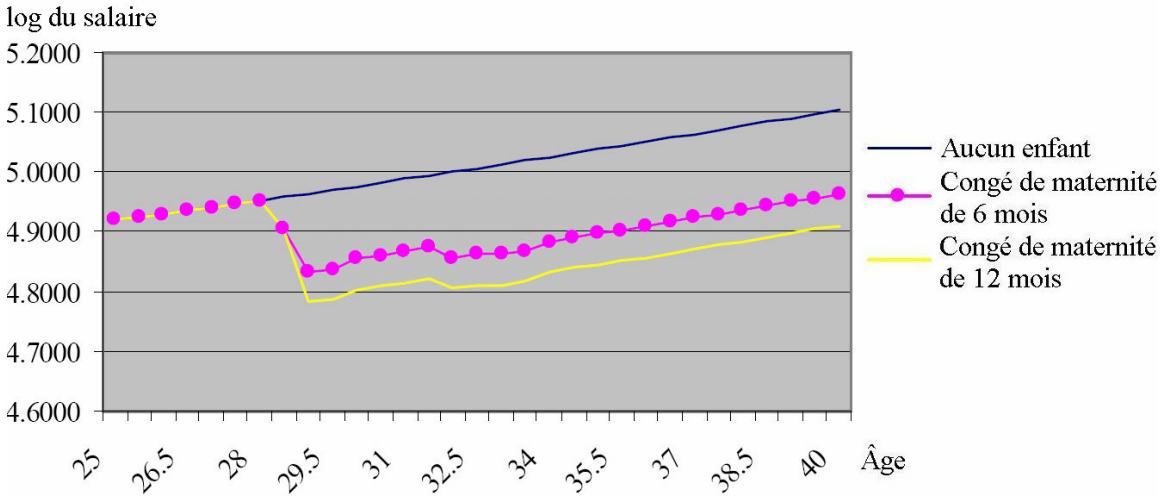


Figure 1.4 Profils salariaux du secteur privé pour des femmes ayant de 15 à 16 années de scolarité. (Tirée de Nielsen, Simonsen et Verner, 2002, p. 29.)

Finalement, Nielsen, Simonsen et Verner (2002, p. 30) découvrent que la maternité et le mariage augmentent significativement les chances des Danoises d'évoluer dans le secteur public. Après une correction pour l'endogénéité du choix du secteur, les femmes du secteur public ne voient pas leur expérience se déprécier (*voir* figure 1.3) contrairement au secteur privé (*voir* figure 1.4). Les mères du secteur public obtiennent même après quelques années une prime salariale. Nielsen, Simonsen et Verner (2002, p. 25) éclaircissent cette apparente incongruité par la moindre mobilité des mères qui encourage l'employeur du secteur public à investir davantage dans leur acquisition de capital humain. Au contraire, les femmes du secteur privé subissent une pénalité salariale qu'elles n'arriveront jamais à rattraper. Bref, cette auto sélection du secteur de l'emploi accroît significativement l'écart salarial dû à la famille passant même dans le secteur public à une prime après quelques années. Les chercheurs extrapolent en affirmant que le secteur public attirerait donc des femmes ayant des caractéristiques non observées par les modèles désavantageuses pour les augmentations salariales (Nielsen, Simonsen et Verner, 2002, p. 31)

Toujours en s'attardant aux comportements distincts des femmes, Hotchkiss et Pitts (2007) utilisent les techniques de décomposition standards d'Oaxaca-Blinder (1973) pour quantifier la contribution de l'intermittence au travail dans la différence salariale entre les sexes. À l'aide de l'étude américaine en panel *Health and Retirement Survey* (HRS) qui suit des individus nés entre 1931 ou 1941 et leur conjointe, ils utilisent des données de 1951 à 1991 sur 1852 hommes et 2404 femmes pour décomposer ce différentiel selon des facteurs observés et selon un effet de coefficient dû au traitement différent des hommes et des femmes sur le marché du travail (Hotchkiss et Pitts, 2007, p. 5).

$$\overline{\ln \hat{W}_M} - \overline{\ln \hat{W}_F} = \hat{\beta}_F^{'} \left(\bar{X}_M - \bar{X}_F \right) + \bar{X}_M^{'} \left(\hat{\beta}_M - \hat{\beta}_F \right) \tag{1}$$

Dans l'équation (1), le premier terme identifie des différences dues à des variables observables qui expliqueraient le différentiel salarial alors que le deuxième terme représente une différence de rendement de ces caractéristiques (Greene, 2008, p. 56). Par exemple, si l'expérience augmente davantage le salaire des hommes que des femmes, le deuxième terme $\left(\hat{\beta}_M - \hat{\beta}_F \right)$ sera positif et il faudrait se questionner sur la présence de discrimination.

Bref, Hotchkiss et Pitts (2007, p. 10) concluent que 61 % de l'impact des facteurs observés, soit du terme $\hat{\beta}_F'$, et 19 % du différentiel salarial total s'explique par les comportements d'intermittence différents entre les sexes. En ne contrôlant pas pour l'intermittence, l'inégalité des caractéristiques d'emploi entre les hommes et les femmes constitue presque la totalité de l'effet des facteurs observés. Dans la mesure où les comportements d'intermittence du travail des femmes résultent habituellement d'un arbitrage entre la famille et le travail, Hotchkiss et Pitts (2007) ne s'attendent pas à une parité salariale entre les sexes.

Kunze (2002) explore plus précisément la question de la dépréciation de l'expérience due au chômage, au congé parental, au service militaire et pour d'autres raisons à l'aide d'une base de données ouest-allemandes (*Institut für Arbeitsmarkt und Berufsforschung* Sample - IABS) qui s'étend de 1975 à 1997. La dépréciation de l'expérience se manifeste par la diminution de l'impact salarial d'un même nombre d'années d'expérience à mesure que la période d'arrêt de travail se prolonge. Kunze (2002, p. 19-20) reconnaît que la pénalité de ces interruptions varie beaucoup selon la cause, allant pour une travailleuse de 2-5 % par année pour le chômage à 13-18 % par année pour le congé maternel. Si une femme change d'employeur suite à un congé maternel, elle perdra en moyenne 6 % de plus que celle qui retourne à son employeur d'origine (Kunze, 2002, p. 22). De plus, il semble que la dépréciation de l'expérience due à la maternité soit de 30 % moindre dans des occupations à majorité masculine ou équilibrée (Kunze, 2002, p. 21). Au contraire, pour les hommes, la dépréciation de l'expérience semble de beaucoup inférieure et se situerait à 0,5-2 % (Kunze, 2002, p. 24).

Pourtant, certaines études restent encore incertaines quant au rôle de la maternité. Mentionnons Drolet (2001) qui examine le lien entre l'état matrimonial, la maternité et les salaires. Si la croissance salariale se produit au cours des premières années suivant l'obtention d'un diplôme, alors le moment de l'interruption du travail due au mariage ou à la maternité aurait une importance (Drolet, 2001, p. 2). Par mariage et maternité précoces ou retardés, elle entend une différence de plus d'une année avant ou après l'âge prévu par une régression par MCO. Elle effectue cette analyse à l'aide l'EDTR de 1998 et d'un échantillon de Canadiennes de 18 à 64 ans ayant un emploi rémunéré et n'étudiant pas à temps plein (Drolet, 2001, p. 5). Elle divise son échantillon en trois cohortes selon leur année de naissance soit avant le baby-boom de 1948, durant le baby-boom (1948-1960) et après celui-ci (Drolet, 2001, p. 7). En

somme, elle ne perçoit aucun lien entre le mariage et les salaires alors que la maternité a un rôle mitigé si on ne considère pas le moment où elle se déroule. Toutefois, son analyse présente plusieurs problèmes que notre démarche corrigera. D'abord, elle ne corrige pas pour le biais de sélection de son échantillon du fait de la considération des femmes qui travaillent. Ensuite, elle ne traite pas de l'endogénéité de l'expérience professionnelle qui découle d'une foule de facteurs. Enfin, son analyse contient un net problème d'hétérogénéité non observée, car le choix du moment de ces deux événements pourrait constituer un puissant indicateur de la motivation et du désir d'investir en capital humain d'où un effet sur le salaire. Notre analyse tiendra compte de ces questions. En contrôlant pour le moment de la maternité, Drolet (2001, p. 14) rapporte donc une prime salariale de 5,8 % pour les femmes ayant eu leurs enfants tard dans la vie versus celles à maternité plus précoce. Cette prime persiste dans le temps, mais tend à s'estomper probablement à cause de responsabilités familiales supplémentaires (Drolet, 2001, p. 17). De plus, cet écart se retrouve parmi les trois cohortes, mais s'accroît au fil des générations. Au contraire, le salaire des femmes ayant eu des enfants précocement varie peu selon les années.

Tout compte fait, les recherches résumées ici s'accordent généralement pour affirmer que la maternité pénalise les mères au niveau salarial. Il reste à savoir si des politiques gouvernementales peuvent vraiment contribuer à alléger le fardeau financier supporté par les femmes qui décident d'avoir des enfants.

1.2 Impact de la politique de garde du Québec sur l'offre de travail des mères

Pour atténuer la pénalité salariale due à la maternité par diverses politiques, il faut d'abord identifier ses déterminants. L'expérience semble être un facteur majeur selon les études. Ainsi, un faible coût et une bonne accessibilité des services de garde

permettraient aux mères d'écourter l'interruption de leur carrière et d'augmenter leur offre de travail ce qui améliorerait possiblement leur salaire sur le marché du travail. Les chercheurs se sont donc attardés à préciser le lien entre le coût des services de garde et l'offre de travail des mères. James Heckman (1974, p. 137) fut un des pionniers qui souligna qu'un accroissement du coût des services de garde réduit la participation des femmes au marché du travail. Cette observation explique l'intérêt des économistes pour la participation des mères au marché du travail et de la sensibilité de leur offre de travail au prix des services de garde.

Ainsi, Blau et Kahn (2006) ont produit une étude exhaustive de l'historique de la participation des femmes mariées avec ou sans enfant au marché du travail aux États-Unis entre 1980 et 2000. Grâce au *Current Population Survey* (CPS) de mars, ils analysent trois cohortes de femmes mariées entre 25 et 54 ans au cours de trois années. Il y a donc la cohorte de 1980 (1979-1981), celle de 1990 (1989-1991) et celle de 2000 (1999-2001) (Blau et Kahn, 2006, p. 12). Ils décrivent l'essor phénoménal de la participation des femmes au marché du travail aux États-Unis passant de 31,5 % en 1947 à 60 % en 1999 (Blau et Kahn, 2006, p. 1). À l'aide de leur échantillon, ils trouvent que la décennie des années 90 a toutefois connu un essoufflement à ce titre puisque ce taux n'est passé que de 57 % à 60 % alors qu'il avait augmenté de dix points de pourcentage au cours de la décennie précédente (Blau et Kahn, 2006, p. 13). Leur modèle indique que la courbe de travail des femmes mariées s'est fortement déplacée vers la droite au cours des années 80 alors qu'elle est restée immobile dans les années 90. Ceci explique la stagnation de leur offre de travail dans les années 90 malgré la forte progression de leurs salaires réels durant cette décennie de 17-20 % (Blau et Kahn, 2006, p. 15).

Par ailleurs, les femmes avec ou sans enfant augmentent de moins en moins leur offre de travail suite à un gain salarial. L'élasticité de leur offre de travail par rapport

à leur salaire a diminué de 0,77-0,88 dans la cohorte de 1980 à 0,58-0,64 pour celle de 1990 pour finir à 0,36-0,41 en 2000, soit une diminution de plus de 50 % (Blau et Kahn, 2006, p. 36). Pour les mères uniquement, ces mêmes élasticités sont passées de 0,98-1,04 en 1980 à 0,49-0,54 en 2000 (Blau et Kahn, 2006, p. 27). De même, l'élasticité de l'offre de travail des femmes mariées par rapport au salaire de leur conjoint a aussi décru de 38 % à 47 %. Les mères ont connu une diminution de 29 % pour cette même élasticité de 1980 à 2000. Ces deux phénomènes découlent probablement de meilleures opportunités de carrière pour les femmes, d'une anticipation croissante du divorce et d'un déclin marqué du salaire réel des hommes au cours des années 70 et 80 (Blau et Kahn, 2006, p. 6-7). Par ailleurs, sans contrôler pour l'éducation, chaque enfant de moins d'un an diminuait l'offre de travail d'une mère de 41 % en 1980, de 29 % en 1990 et de 26 % en 2000 (Blau et Kahn, 2006, p. 19). Au fil des décennies, les mères d'enfants en bas âge travaillent de plus en plus. L'élasticité de l'offre de travail des hommes se situant entre 0,05 et 0,10 dans la cohorte de 2000, celle des femmes tend donc à rejoindre celle des hommes. Ces données indiquent que les politiques gouvernementales visant à réduire la pauvreté des mères en augmentant leur salaire deviennent de moins en moins efficaces à mesure que ces élasticités décroissent ce qui concerne entre autres la politique de services de garde du Québec.

Afin de mieux mesurer l'efficacité de politiques gouvernementales, il convient d'évaluer la sensibilité de l'offre de travail des mères en fonction du coût des services de garde. Blau et Currie (2004, p. 8) passent en revue vingt études publiées entre 1992 et 2003 pour le Canada et les États-Unis. Les élasticités du prix des services de garde sur l'offre de travail des mères varient énormément selon les modèles, les bases de données et les spécifications, mais s'échelonnent entre 0,06 et -3,60. Parmi toutes ces études, ils en retiennent trois dont la méthodologie serait plus adéquate principalement à cause du traitement de la garde informelle : Ribar (1995), Tekin

(2002) et Blau et Hagy (1998). Ces trois études arrivent respectivement à une élasticité prix / offre de travail de -0,09, -0,15 et -0,20. Ces chiffres portent à croire que la vraie élasticité qui tient compte de la garde informelle serait faible (Blau et Currie, 2004, p. 11). Ces élasticités inférieures résultent principalement de ce qu'ils appellent un « effet d'éviction » (Blau et Currie, 2004, p. 31). Suite à une subvention d'un mode de garde payant, des mères utilisant un mode non payant vont changer vers le système avec contributions. Par exemple, Baker *et al.* (2005, p. 28-29) estiment avec l'Enquête longitudinale nationale sur les enfants et les jeunes (ELNEJ) qu'une nouvelle utilisatrice sur trois du système à 5 $ québécois avait auparavant un type d'arrangement informel ce qui fait passer leur élasticité prix de -0,58 à -0,28. Bref, en tenant compte des modes de garde informels, une subvention diminuant le coût des services de garde a un impact moindre sur l'offre de travail des mères par rapport à la demande accrue de places en garderies.

Dans un contexte canadien, Cleveland, Gunderson et Hyatt (1996, p. 145) calculent pour le Canada une élasticité du taux d'emploi de -0,388. Pour ce faire, ils utilisent l'Enquête nationale sur la garde des enfants datant de 1988 (1996, p. 142). L'effet de subventions pour des politiques de garde serait donc plus important au Canada. Par exemple, une diminution de 10 % du coût de garde augmenterait l'apport de travail des mères de 3,9 %. D'autres canadiens abondent dans ce sens. Powell (1997, p. 590) trouve une élasticité de la participation à la main d'œuvre par rapport aux coûts des services de garde de -0,38 alors que Michalopoulos et Robins (2000, p. 459) arrivent à une estimation de -0,464 pour les garderies en installation et à -0,156 pour tous les types de garde confondus. Tout compte fait, une politique gouvernementale peut-elle avec de telles élasticités favoriser le travail des mères?

Au Québec, Lefebvre et Merrigan (2007a) examinent si la nouvelle politique de services de garde à 5 $ du Québec se concrétise pour les mères d'enfants de cinq ans

et moins par des taux de participation, des heures travaillées et des revenus annuels croissants par rapport au reste du Canada. En diminuant le coût des services de garde, le gouvernement québécois espérait entre autres favoriser le travail des mères et ainsi améliorer le sort économique de leur famille. Lefebvre et Merrigan (2007a) utilise une approche « quasi-expérimentale » pour déterminer les différences entre un groupe de traitement, les mères québécoises, et un groupe de contrôle, celles du reste du Canada, avant et après l'introduction du nouveau système de subvention des services de garde. Son implantation a débuté en 1997 alors que la politique a véritablement pris son envol en 1999 (*voir* ann. A).

À l'aide de l'EDTR pour les années 1993 à 2002 et d'un modèle économétrique de différence-en-différences, Lefebvre et Merrigan (2007a) décèlent un impact significatif sur tous ces facteurs dont une augmentation du taux de participation de 7,3 %, un prolongement de 4,28 semaines travaillées et une augmentation des revenus de 2 300 $ significative à 10 % (Lefebvre et Merrigan, 2007a, p. 22). En contrôlant pour les années 1999 à 2002, ces effets croissent dans le temps. Par contre, les résultats deviennent moins probants pour les mères ayant un niveau de scolarité secondaire ou moins, celles en fait qui ont expérimenté la moins grande diminution de leurs frais de garde (Lefebvre et Merrigan, 2007a, p. 22). Baker *et al.* (2005, p. 59) arrivent sensiblement aux mêmes constats grâce à l'ELNEJ de 1994 à 2003. En ne conservant que des mères d'enfants de 0 à 4 ans, ils calculent une augmentation du taux de participation de 7,7 % et de 3,53 semaines de travail supplémentaires.

Pour mesurer les effets dynamiques de long terme de cette politique, Lefebvre *et al.* (2007b, p. 4) se tournent vers les mères ayant au moins un enfant de 6 à 11 ans entre 1996 et 2004, mais aucun de moins de 6 ans. Encore une fois, ils notent un taux de participation de 7,4 % en hausse au Québec par rapport au reste du Canada (Lefebvre *et al.*, 2007b, p. 26). Lorsqu'ils désagrègent cet effet de 2002 à 2004, une

tendance à la hausse demeure. Cependant, les résultats pour les mères ayant plus qu'un diplôme secondaire ne sont pas aussi convaincants. Aussi, les résultats pour les revenus annuels de travail ne sont plus significatifs à long terme (Lefebvre *et al.*, 2007b, p. 26-27) contrairement à l'étude précédente qui visait plus le court terme (Lefebvre et Merrigan, 2007a, p. 22). Ce phénomène indiquerait que la politique de services de garde du Québec n'aurait pas eu d'effet dynamique significatif sur le salaire des femmes. Pourtant, les mères québécoises ont acquis plus d'expérience à cause de cette politique, en moyenne un an selon Lefebvre et Merrigan (2007b, p. 32). Elles seraient plus productives et devraient donc recevoir un salaire plus élevé. Néanmoins, Lefebvre et Merrigan (2007b) considèrent l'expérience comme exogène alors que notre démarche l'endogénéise en la rendant dépendante d'une foule de facteurs tels que la politique de services de garde du Québec, le taux de chômage régional, le revenu familial, l'éducation ou encore le nombre d'enfants. Notre étude tentera ainsi de mieux définir le rendement de l'expérience.

1.3 Expérience et gains salariaux

Si les mères québécoises travaillent plus à cause de la politique de services de garde du Québec, elles devraient acquérir plus d'expérience et par conséquent, obtenir des salaires plus élevés. Les études démontrant le lien entre l'expérience et les gains salariaux nous portent à le croire. Dans les années 1950, des économistes tels que Theodore Schultz, Jacob Mincer et Gary Becker ont redécouvert les intuitions d'Adam Smith sur les liens entre la productivité et les salaires et développé la théorie du capital humain (Kaufman, 1991, p. 291). Le capital humain se définit comme suit :

> L'ensemble des capacités productives d'un individu provenant de l'acquisition préalable de connaissances générales ou spécifiques ainsi que de l'expérience passée notamment sur des systèmes de production. (Duthil, 2004, p. 66)

Ces conceptions ont encouragé une foule de chercheurs à établir une corrélation positive entre l'expérience et le salaire. Pour souligner l'importance de l'expérience, la figure 1.4 montre jusqu'à quel point le salaire progresse avec l'âge. À l'aide du *Survey of Income and Program Participation* (SIPP) de 1984 à 2003, French *et al.* (2006) se penchent sur la question du début de carrière d'hommes entre 18 et 28 ans, période où le salaire croît généralement le plus rapidement soit entre 3 % et 6 %. De leur côté, Goldsmith et Veum (2002, p. 434) concluent à un accroissement salarial un peu moindre de 2,8 % pour des hommes plus âgés de 31 à 39 ans du NLSY ayant en moyenne 8 année d'expérience. French *et al.* (2006) décomposent ensuite cette croissance salariale en fonction des changements d'emploi, des cycles d'affaire et du rendement de l'expérience. Parmi ces trois facteurs, le rendement de l'expérience estimé à 4 % demeure de loin le plus important peu importe l'éducation ou le sexe même s'il varie beaucoup dans le temps (French *et al.*, 2006, p. 26).

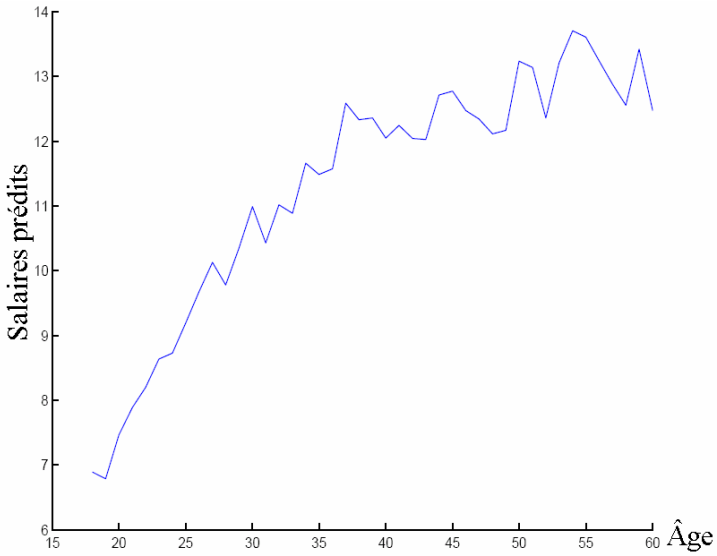


Figure 1.5 Salaires prédits d'hommes sans diplôme collégial selon le SIPP de 1996 aux États-Unis. (Tirée de French *et al.*, 2006, p. 32.)

Gladden et Taber (2006) s'attardent uniquement au cas des travailleurs peu qualifiés dans leurs premières années d'expérience grâce au NLSY79. L'échantillon comprend des jeunes de 14 à 22 ans avec moins d'une année de collège et dix ans d'expérience ou moins sur le marché du travail (Gladden et Taber, 2006, p. 10). Avec un modèle à effet aléatoire corrélé $\left(w_{it} = \theta_i + \beta_i E_{it} + u_{it}\right)$, Gladden et Taber (2006, p. 1) découvrent que les habiletés non observées du travailleur $\left(\theta_i\right)$ sont en fait non corrélées avec les habiletés d'apprentissage liées à l'expérience $\left(\beta_i\right)$. Bref, la croissance salariale serait aussi élevée pour les travailleurs peu ou moyennement spécialisés (Gladden et Taber, 2006, p. 18). Ainsi, les mères québécoises de notre étude auraient toutes un avantage salarial lié à l'amélioration de leur capital humain par leur expérience accrue sur le marché du travail.

Enfin, Zhang (2002) affirme aussi que l'acquisition d'expérience par les femmes accroît leur salaire et ce peu importe leur qualification. Avec l'EDTR pour les années 1993 à 1998 et un échantillon composé de jeunes travailleurs canadiens des deux sexes âgés de 16 à 30 ans en 1993, il décompose l'effet de la durée d'occupation d'emploi et de l'expérience entre les sexes et deux degrés de qualification des travailleurs. Les travailleurs peu qualifiés ont obtenu un diplôme d'études secondaires ou moins (Zhang, 2002, p. 2). Le tableau 1.2 résume ses résultats et permet de comparer pour les deux sexes l'impact de la durée d'occupation d'un emploi par rapport à l'expérience pour un haut et un bas degré de qualification. Il découvre entre autres pour les femmes non qualifiées que la durée d'occupation de l'emploi compte autant que l'expérience avec des incidences moyennes sur 6 ans de 2,2 % contre 2,4 % (Zhang, 2002, p. 11). Par exemple, une femme peu qualifiée verrait son salaire augmenter dans un même emploi de 4,5 % à 6,6 % en trois ans, 9,9 % à 11,9 % en cinq ans et de 8,3 % à 10,8 % en 10 ans (Zhang, 2002, p. 14-15). La hausse cesse vers la seizième année alors qu'au total, le salaire aura crû de 17 %. Une femme qualifiée profiterait environ de la même courbure avec des gains supplémentaires

maximums de 14 % à 20 % entre la neuvième et la dixième année d'emploi. Par conséquent, autant les travailleuses qualifiées que non qualifiées profiteraient de l'expérience additionnelle accumulée dans la mesure où la politique de garde du Québec incite plus de mères à travailler.

Tableau 1.2 Effets de la durée d'occupation de l'emploi et de l'expérience.
(Tiré de Zhang, 2002, p. 11.)

	Hommes peu qualifiés		Hommes hautement qualifiés		Femmes peu qualifiées		Femmes hautement qualifiées	
	Durée	Exp.	Durée	Exp.	Durée	Exp.	Durée	Exp.
Moyenne 1993-1998	0,0301 (0,0042)	0,0246 (0,0047)	0,0054 (0,0041)	0,0499 (0,0047)	0,0221 (0,0047)	0,0242 (0,0055)	0,0255 (0,0034)	0,0413 (0,0041)

En substance, notre étude contribue à la littérature existante en tentant de découvrir si l'expérience supplémentaire des mères québécoises suscitée par la nouvelle politique de garde du Québec s'est traduite par des salaires plus élevés. Un tel effet allègerait ou annulerait même la pénalité due à la maternité.

CHAPITRE II

APPROCHE ÉCONOMÉTRIQUE

La littérature économique constate de sérieux problèmes économétriques lors de l'analyse du processus salarial (Blau et Kahn, 2006, p.8; Gladden et Taber, 2006, p. 1; Semykina et Wooldridge, 2006, p. 21). Ce chapitre vise à relever les difficultés majeures généralement mentionnées par les chercheurs. De plus, nous expliquons la méthode proposée par Semykina et Wooldridge (2006) pour contourner le biais de sélection et l'endogénéité des variables dans un modèle de données de panel avec effets fixes. Nous utilisons ensuite une forme réduite pour revenir à la démarche antérieure de Wooldridge (1995b) sans variable instrumentale afin d'isoler l'effet marginal de la politique de services de garde du Québec. Enfin, nous présentons différentes méthodes pour évaluer la faiblesse et la validité des instruments.

2.1 Problèmes économétriques liés à la régression salariale

Trois difficultés majeures se posent généralement dans les équations de salaires en données de panel : l'hétérogénéité, l'endogénéité et le biais de sélection (Semykina et Wooldridge, 2006, p. 21).

D'abord, l'hétérogénéité concerne les caractéristiques des individus difficilement observables comme les habiletés, les préférences ou la motivation (Greene, 2008, p. 182). Ces caractéristiques étant souvent exclues et corrélées avec des variables explicatives du modèle, les régressions standard par MCO ne produiront pas alors des estimateurs convergents. Par exemple, la motivation à travailler pourrait être corrélée

avec le niveau de scolarité (Blau et Kahn, 2006, p. 11-12). Dans ce cas, une estimation incluant une constante différente pour chaque individu appelée effet fixe produira des estimés convergents pour les coefficients. Il faut toutefois disposer de données longitudinales, c'est-à-dire des informations sur N individus indicés i et sur T périodes indicées t.

$y_{it} = x_{it}\beta + c_i + u_{it}$ où x_{it} inclut les K régresseurs, i=1,…, N et t=1,…,T (2)

Dans l'équation précédente, c_i serait l'effet fixe attribué à chaque individu i en supposant qu'il demeure constant pour toutes les t périodes observées (Greene, 2008, p. 182).

Ensuite, l'endogénéité découle d'une corrélation non nulle entre un régresseur et l'erreur du modèle.

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_k x_k + u \quad (3)$$

Dans l'équation précédente, $y, x_1, x_2, ..., x_k$ sont des scalaires observables, u est l'erreur non observable aléatoire et $\beta_0, \beta_1, \beta_2, ..., \beta_k$ sont les paramètres à estimer. Dans ce cas-ci, l'endogénéité se manifesterait par une corrélation non nulle entre x_j et u soit $Cov(x_j, u) \neq 0$, j=1,2,…,k (Wooldridge, 2002, p. 49-50).

Wooldridge (2002, p. 50-51) propose trois explications principales à l'endogénéité. Premièrement, une erreur de mesure survient lorsqu'il faut s'astreindre à des données imparfaites x_j par rapport à la vraie valeur x_j^* (Wooldridge, 2002, p. 51). Tout dépendant de la relation entre x_j et x_j^*, l'erreur de mesure se retrouvera dans l'erreur du modèle u ce qui pourrait engendrer une corrélation non nulle entre x_j et u. Deuxièmement, la simultanéité se présente lorsqu'une variable explicative x_j dépend aussi de la variable expliquée y. Si x_j

dépend en partie de y, alors x_j et u seront généralement corrélés (Wooldridge, 2002, p. 51). Par exemple, de hauts salaires pourraient avoir incité des individus à travailler davantage d'où une hausse de l'expérience accumulée au cours des années. Troisièmement, le biais de variables omises résulte simplement de l'omission de variables pertinentes corrélées avec une ou des variables explicatives du modèle. Greene (2008, p. 133) définit le biais de l'omission du régresseur X_2 dans la régression de X de $K = K_1 + K_2$ colonnes ainsi :

$$y = X\beta + u = X_1\beta_1 + X_2\beta_2 + u \tag{4}$$

$$\Rightarrow b_1 = \left(X_1'X_1\right)^{-1}X_1'y = \beta_1 + \left(X_1'X_1\right)^{-1}X_1'X_2\beta_2 + \left(X_1'X_1\right)^{-1}X_1'u\,. \tag{5}$$

Alors, le biais de variables omises serait $\left(X_1'X_1\right)^{-1}X_1'X_2\beta_2$ ce qui rend l'estimateur b_1 non convergent. Le rôle de l'habileté dans une équation salariale illustre bien ce type de biais. Dans ce cas, les années d'éducation sont probablement corrélées avec les habiletés non observées d'une personne. Plus le nombre d'années d'éducation augmente, plus il est vraisemblable qu'elle dispose de meilleures habiletés ce qui devrait favoriser son salaire. Sans contrôler pour ces caractéristiques non observées, l'impact à la marge d'une année d'éducation additionnelle sur le salaire sera surévalué parce que les habiletés augmentent aussi en moyenne. Le coefficient captera donc deux effets distincts. Ainsi, le contrôle des effets fixes dans une équation salariale diminuera probablement le paramètre de l'éducation.

Par conséquent, l'expérience pose de sérieux problèmes d'endogénéité. Hormis la corrélation entre l'expérience et les habiletés non observées, si la décision de participer au marché du travail dépend de l'offre salariale, alors un choc exogène sur les salaires dans le passé sera corrélé avec le nombre d'années d'expérience qu'on observe par la suite (Semykina et Wooldridge, 2006, p. 21). De même, les décisions d'intermittence au travail des mères qui affectent leur expérience pourraient aussi provenir d'anticipations rationnelles sur les conséquences salariales présentes et

futures. La décision d'avoir un enfant, de prendre un congé maternel payé ou non et de freiner l'accumulation d'expérience dépend sûrement de l'offre salariale présente et anticipée par ces femmes (Hotchkiss et Pitts, 2007, p. 4). Des salaires élevés favoriseraient peut-être le travail des femmes et donc leur expérience (Blau et Kahn, 2006, p. 15). Bref, l'hétérogénéité, l'endogénéité et le biais d'omission de variables relèvent de la validité interne, soit de la consistance des paramètres estimés.

La validité externe concerne plutôt la généralisation des résultats de l'étude à une véritable population. Est-il raisonnable d'étendre les conclusions de l'étude à la population d'intérêt? Le biais de sélection constitue une facette importante de cette question. Il se manifeste lorsque la disponibilité des observations d'un échantillon dépend de la valeur de la variable dépendante. Par exemple, le salaire des mères est observé seulement pour celles qui décident d'occuper un emploi. Les mères ayant reçu des chocs positifs sur leur salaire seront plus enclines à participer au marché du travail ce qui biaise les erreurs de l'individu i au temps t (u_{it}) vers le haut (Gladden et Taber, 2006, p. 16). D'ailleurs, ce biais peut évoluer à travers le temps pour des données longitudinales puisque le taux de participation des femmes avec de jeunes enfants a augmenté au cours des dernières décennies et que l'élasticité de leur offre de travail a diminué. Les raisons motivant le retrait des femmes du marché du travail ont donc changé ce qui a modifié la nature du biais de sélection. En somme, notre analyse économétrique exige une méthode permettant de résoudre ces divers problèmes ce que propose justement Semykina et Wooldridge (2006).

2.2 Méthodes économétriques d'estimation du rendement de l'expérience

L'étude de l'impact de l'expérience sur le salaire exige une approche économétrique permettant de contourner les problèmes engendrés par l'hétérogénéité,

l'endogénéité et le biais de sélection. La méthode proposée par Semykina et Wooldridge (2006) tient compte de ces problèmes. Dans un contexte de données longitudinales, les auteurs contrôlent pour l'hétérogénéité à l'aide d'effets fixes et corrigent pour l'endogénéité par l'inclusion de variables instrumentales. Les méthodes économétriques employées dans ce travail introduisent graduellement ces corrections. L'évolution du rendement de l'expérience à travers ces différentes méthodes fournit une meilleure idée de l'ampleur du biais découlant de l'hétérogénéité, de l'endogénéité et du biais de sélection.

D'abord, la régression par MCO nous sert de référence à cause de sa simplicité et de son usage répandu. Elle consiste à minimiser la somme des carrés des erreurs du modèle. En regroupant l'erreur idiosyncrasique u_{it} et l'effet fixe c_i de l'équation (2), on obtient le terme d'erreur v_{it} dans l'équation suivante :

$$y_{it} = x_{it}\beta + v_{it}\text{, t=1,…,T} \tag{6}$$

où x_{it} est un vecteur $1 \times K$ contenant les variables explicatives, β est un vecteur de paramètres $K \times 1$ et v_{it} est le vecteur des erreurs contenant l'effet non observé c_i. Normalement, l'estimateur MCO se présente sous cette forme :

$$\hat{\beta}_{MCO} = \left(\frac{1}{n}\sum_{j=1}^{n} x_j' x_j\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n} x_j' y_j\right)\text{, où n est le nombre total d'observations.} \tag{7}$$

Toutefois, une notation différente devra être employée pour des données de panel, c'est-à-dire un échantillon de personnes observées sur plusieurs périodes. Si chaque personne est observée un même nombre de fois, le panel est équilibré. Dans le cas contraire, le panel est déséquilibré. L'indicateur de sélection s_{it} permet alors de généraliser l'équation d'un estimateur à un panel déséquilibré. La variable

dichotomique s_{it} prend la valeur de 1 si les variables pertinentes sont observées pour la personne i au temps t, soient (y_{it}, x_{it}) dans une régression MCO. Autrement, elle annule l'élément dans la sommation en prenant la valeur de 0. Par exemple, si la variable dépendante y_{it} est disponible, mais non le vecteur x_{it}, l'élément sera annulé. La notation suivante convient donc à la méthode MCO dans un contexte de panel déséquilibré sur N individus et pour un nombre maximal de périodes observées T :

$$\hat{\beta}_{MCO} = \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} x'_{it} x_{it} \right)^{-1} \left(\frac{1}{N} \sum_{t=1}^{N} \sum_{t=1}^{T} s_{it} x'_{it} y_{it} \right) \tag{8}$$

$$\hat{\beta}_{MCO} = \beta + \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} x'_{it} x_{it} \right)^{-1} \left(\frac{1}{N} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} x'_{it} u_{it} \right). \tag{9}$$

Afin que l'estimateur MCO soit convergent et asymptotiquement normal, plusieurs hypothèses sont nécessaires (Greene, 2008, p. 72-73). Comme le font Semykina et Wooldridge (2006, p. 7) pour la méthode MCDE, nous conservons la notation en panel pour la suite des explications. Ainsi, sachant que $s_{it} = 1$ lorsque (y_{it}, x_{it}) sont observés, la méthode MCO suppose (Wooldridge, 2002, p. 553-554) :

i- $E(v_{it} \mid x_{it}, s_{it}) = 0$, t = 1,…T ;
ii- rang $E\left(\sum_{t=1}^{T} s_{it} x'_{it} x_{it}\right) = K$;
iii- Conditions de régularité standard : $E(v_{it}^2) < \infty$ et $E(x_{itk}^2) < \infty$, k = 1,…,K.

L'hypothèse (i) affirme que le biais de variable omise est inexistant ou qu'il n'y a plus aucun pouvoir explicatif restant dans le terme d'erreur v_{it} en contrôlant pour les variables explicatives et l'indicateur de sélection. L'hypothèse (ii) de plein rang interdit la multicollinéarité ou la présence d'une relation linéaire exacte entre les K régresseurs ou colonnes de x_{it}. L'hypothèse (iii) rassemble les conditions nécessaires à ce qu'un théorème central limite s'applique tel que l'estimateur $\hat{\beta}_{MCO}$ soit

convergent et asymptotiquement normal (Wooldridge, 2002, p 554). En plus, en présence d'homoscédasticité $\left(E\left(v_{it}^2 \middle| x_{it}, s_{it}\right) = \sigma^2\right)$, la variance asymptotique usuelle est valide :

$$Var\left(\hat{\beta}_{MCO} \middle| x_{it}, s_{it}\right) = \frac{1}{n-K}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} u_{it}^2\right)\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} x_{it}' x_{it}\right) \text{ où } n = \sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} . \quad (10)$$

Par contre, si l'hypothèse (i) est violée, alors un ou des régresseurs endogènes sont corrélés avec le terme d'erreur v_{it} de l'équation (6) et il y a présence d'endogénéité. Ces régresseurs endogènes rendent chaque élément du vecteur de coefficient $\hat{\beta}_{MCO}$ non convergent, c'est-à-dire qu'il comporte un biais et ne converge pas asymptotiquement vers la valeur de la population β_{MCO}. La correction pour l'endogénéité nécessite un ensemble d'instruments corrélés avec les variables endogènes, mais non avec les erreurs du modèle. Cette propriété permet ensuite d'estimer par MCO les K combinaisons linéaires des instruments les plus hautement corrélés avec les régresseurs. Ces valeurs prédites sont orthogonales aux erreurs v_{it}, c'est-à-dire non corrélées. La maximisation de la corrélation entre les régresseurs et les K combinaisons linéaires des instruments visent à maximiser l'efficience de l'estimateur. Il ne reste qu'à utiliser ces prédictions afin de régresser par MCO la variable expliquée. Ces deux étapes expliquent le nom de cette méthode : moindres carrés en deux étapes ou MCDE.

Les hypothèses sous-jacentes ressemblent beaucoup à celles de la méthode MCO, mais en remplaçant le vecteur de variables explicatives x_{it} par le vecteur z_{it} de dimension $1 \times \mathrm{L}$. Le vecteur z_{it} contient les L instruments non corrélés aux erreurs v_{it} de l'équation (6) : $E\left(v_{it} \mid z_{it}\right) = 0$, t = 1,…T. Dès lors, l'indicateur de sélection s_{it}

prend la valeur de 1 uniquement lorsque (y_{it}, x_{it}, z_{it}) sont observés (Wooldridge, 2002, p. 553). Les hypothèses suivantes doivent donc s'appliquer pour avoir un estimateur MCDE convergent et asymptotiquement normal (Wooldridge, 2002, p. 553-554) :

i- $E(v_{it} \mid z_{it}, s_{it}) = 0$, t = 1,…,T;

ii- rang $E\left(\sum_{t=1}^{T} s_{it} z_{it}' x_{it}\right) = K$;

iii- rang $E\left(\sum_{t=1}^{T} s_{it} z_{it}' z_{it}\right) = L$ où $L \geq K$;

iv- Conditions de régularité standard : $E(v_{it}^2) < \infty$, $E(x_{itk}^2) < \infty$ $E(z_{itl}^2) < \infty$, k = 1,…,K et l = 1,…,L.

L'hypothèse (i) pose que les instruments z_{it} doivent être exogènes, mais peuvent être corrélés avec l'indicateur de sélection s_{it}. La relation entre v_{it} et s_{ir}, r≠t, n'est pas précisée si bien que la sélection au cours des périodes précédentes et suivantes n'a aucune importance au niveau de la consistance des estimateurs (Semykina et Wooldridge, 2006, p. 9). Les hypothèses (ii) et (iii) proscrivent une relation linéaire exacte entre les colonnes des vecteurs x_{it} et z_{it}. Ensuite, l'estimateur $\hat{\beta}_{MCDE}$ dans un contexte de données de panel déséquilibré est de la forme :

$$\hat{\beta}_{MCDE} = \left[\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} x_{it}' z_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} z_{it}' z_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} z_{it}' x_{it}\right)\right]^{-1} \times \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} x_{it}' z_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} z_{it}' z_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} z_{it}' y_{it}\right) \quad (11)$$

$$\hat{\beta}_{MCDE} = \beta + \left[\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} x_{it}' z_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} z_{it}' z_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} z_{it}' x_{it}\right)\right]^{-1} \times \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} x_{it}' z_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} z_{it}' z_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} z_{it}' v_{it}\right). \quad (12)$$

Par ailleurs, les instruments se divisent en deux classes : les régresseurs exogènes inclus (z_{1it}) et les régresseurs exogènes exclus (z_{2it}). Le régresseurs exogènes inclus se trouvent aussi dans le vecteur de variables explicatives x_{it} si bien qu'ils ne sont pas considérés comme endogènes lorsque le modèle calcule le vecteur de coefficients $\hat{\beta}_{MCDE}$. Au contraire, les régresseurs exogènes exclus ne figurent pas dans le vecteur x_{it}. En regroupant dans z_{1it} tous les régresseurs de x_{it} qui ne sont pas endogènes y compris la constante $x_{it1}=1$, le vecteur z_{2it} doit comporter au moins autant de variables exogènes exclues que de variables explicatives endogènes plus une autre variable pour tenir compte de la sélection (Semykina et Wooldridge, 2006, p. 20).

Par ailleurs, l'hétérogénéité constitue un autre problème majeur qu'une régression salariale rencontre. Dans l'équation (2), l'hétérogénéité implique alors une corrélation positive entre l'effet fixe non observé c_i et les variables explicatives x_{it}. Puisque c_i n'est pas indicé de t, la constante est différente pour chaque individu, mais demeure constante dans le temps. Une technique d'estimation consiste à soustraire aux variables explicatives et à la variable expliquée la moyenne des périodes observées pour chaque personne. En posant que t=1 pour la première période observée, que t=T pour le maximum de périodes possiblement observées et que $T_i=\sum_{t=1}^{T} s_{it}$ définit le nombre de périodes observées pour la personne i, la soustraction des moyennes des personnes s'effectue de la manière suivante :

$$\ddot{x}_{it} \equiv x_{it} - T_i^{-1}\sum_{r=1}^{T} s_{ir}x_{ir} \tag{13}$$

$$\ddot{y}_{it} \equiv y_{it} - T_i^{-1}\sum_{r=1}^{T} s_{ir}y_{ir} \tag{14}$$

$$\ddot{z}_{it} \equiv z_{it} - T_i^{-1}\sum_{r=1}^{T} s_{ir}z_{ir}\ . \tag{15}$$

Les personnes observées qu'une seule fois ne peuvent faire partie de l'échantillon pour les méthodes en effets fixes. L'indicateur de sélection prend la valeur de 1 si

(y_{it}, x_{it}) ou si (y_{it}, x_{it}, z_{it}) sont observés selon que z_{it} est utilisé ou non dans l'estimateur. Cette transformation annule le terme c_i. Puisque la moyenne pour chaque individu est soustraite des observations, l'effet des caractéristiques individuelles non observées supposé fixe à travers le temps disparaît. Ainsi, sa corrélation avec les variables explicatives n'a plus d'importance. Les estimateurs se calculent ensuite d'une manière comparable aux méthodes MCO et MCDE.

De cette façon, l'estimateur avec effets fixes (EF) ne contrôle que pour l'hétérogénéité inobservée. Les hypothèses suivantes garantissent un estimateur EF convergent et asymptotiquement normal (Wooldridge, 2002, p. 579) :

i- $E(u_{it} \mid x_i, s_i, c_i) = 0$, t = 1,…,T;

ii- rang $E\left(\sum_{t=1}^{T} s_{it} \ddot{x}'_{it} \ddot{x}_{it}\right) = K$;

iii- Conditions de régularité standard : $E(u_{it}^2) < \infty$, $E(\ddot{x}_{itk}^2) < \infty$ k = 1,…,K.

L'hypothèse (i) implique que $E(s_{it} \ddot{x}'_{it} u_{it}) = 0$ et donc que le biais de variable omise est inexistant. Elle n'exclut aucune relation entre s_i et (x_i, c_i) qui peut être aléatoire ou déterminée. Toutefois, les erreurs idiosyncrasiques u_{it} ne peuvent être corrélées avec le vecteur des indicateurs de sélection s_i étant donné (x_i, c_i) pour tout t (Wooldridge, 2002, p. 579-580). Ainsi, l'estimateur EF se calcule d'une manière similaire à l'estimateur MCO :

$$\hat{\beta}_{EF} = \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} \ddot{x}'_{it} \ddot{x}_{it}\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} \ddot{x}'_{it} \ddot{y}_{it}\right) \tag{16}$$

$$\hat{\beta}_{EF} = \beta + \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} \ddot{x}'_{it} \ddot{x}_{it}\right)^{-1} \left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} \ddot{x}'_{it} u_{it}\right). \tag{17}$$

Par ailleurs, l'estimateur avec effets fixes des moindres carrés en deux étapes (EF-MCDE) combine une correction pour l'hétérogénéité et l'endogénéité. S'inspirant de la forme de l'estimateur MCDE et de la soustraction des moyennes individuelles de l'estimateur EF, la méthode EF-MCDE permet d'obtenir à partir de l'équation (2) un estimateur de la forme :

$$\hat{\beta}_{EF-MCDE} = \left[\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\ddot{z}_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{z}_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{x}_{it}\right)\right]^{-1} \times\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\ddot{z}_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{z}_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{y}_{it}\right) \tag{18}$$

$$\hat{\beta}_{EF-MCDE} = \beta + \left[\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\ddot{z}_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{z}_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{x}_{it}\right)\right]^{-1} \times\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\ddot{z}_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{z}_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}u_{it}\right). \tag{19}$$

Affirmer que cet estimateur est convergent et se conforme à la loi normale asymptotiquement nécessite toutefois quelques hypothèses. Ici, $z_i = (z_{i1},...,z_{iT})$ et $s_i = (s_{i1},...,s_{iT})$ (Semykina et Wooldridge, 2006, p. 11) :

i- $E(u_{it} \mid z_i, s_i, c_i) = 0$, t = 1,…T;
ii- rang $E\left(\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\ddot{z}_{it}\right) = K$;
iii- rang $E\left(\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{z}_{it}\right) = L$;
iv- Conditions de régularité standard : $E(u_{it}^2) < \infty$, $E(\ddot{x}_{itk}^2) < \infty$ $E(\ddot{z}_{itl}^2) < \infty$, k = 1,…,K et l = 1,…,L.

La première hypothèse est de loin la plus cruciale. Elle impose que les instruments soient exogènes conditionnellement aux effets fixes. Autrement dit, les instruments z_{it} ne peuvent être corrélés avec les erreurs idiosyncrasiques u_{it} étant

donné l'effet fixe c_i, soit que $E\left(\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}u_{it}\right)=0$ (Semykina et Wooldridge, 2006, p. 11). Ainsi, puisque les effets fixes disparaissent du modèle en enlevant la moyenne, aucune corrélation n'est exigée pour l'indicateur de sélection s_{it} et les effets fixes c_i. Toutefois, comme pour l'estimateur EF, le vecteur des indicateurs de sélection s_i ne peut être corrélé avec l'erreur idiosyncrasique u_{it} pour tout t.

De plus, aucune hypothèse de distribution n'est faite pour les erreurs u_{it} de l'équation primaire (2). Les termes d'erreur peuvent être corrélés en série au niveau individuel et distribués de façon hétérogène à condition de corriger les écarts types. L'hétéroscédasticité désigne une variance des erreurs du modèle qui varie selon la valeur des variables explicatives. Aucune condition ne restreint la relation entre la variance conditionnelle $\text{var}\left(v_{it}\middle|x_{it}\right)$ et le vecteur x_{it}. Il peut aussi y avoir une forme de corrélation au niveau individuel entre les erreurs du modèle à différentes périodes : $\text{cov}\left(v_{ir},v_{is}\middle|X_i\right)=\sigma_{irs}$ pour $r\neq s$. Les écarts types rapportés dans ce travail corrigent pour ces deux problèmes, mais n'ajustent pas pour une corrélation possible entre individus. Ils proviennent de la racine carrée des éléments diagonaux des matrices de variance covariance asymptotique estimées de la manière suivante :

$$\hat{V}_a\left(\hat{\beta}_{MCO}\right)=\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}x'_{it}x_{it}\right)^{-1}\left\{\sum_{i=1}^{N}\left[\left(\sum_{t=1}^{T} s_{it}x'_{it}\hat{v}_{it}\right)\left(\sum_{t=1}^{T} s_{it}x'_{it}\hat{v}_{it}\right)'\right]\right\}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}x'_{it}x_{it}\right)^{-1} \quad (20)$$

$$\text{où } \hat{v}_{it}=y_{it}-x_{it}\hat{\beta}_{MCO}$$

$$\hat{V}_a\left(\hat{\beta}_{MCDE}\right)=\left(C'D^{-1}C\right)^{-1}C'D^{-1}BD^{-1}C\left(C'D^{-1}C\right)^{-1} \quad (21)$$

$$\text{où } C=\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}x'_{it}z_{it} \quad (22)$$

$$D=\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}z'_{it}z_{it} \quad (23)$$

$$B = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{t=1}^{T} s_{it} z'_{it}\hat{v}_{it}\right)\left(\sum_{t=1}^{T} s_{it} z'_{it}\hat{v}_{it}\right)'\right] \text{ où } \hat{v}_{it} = y_{it} - x_{it}\hat{\beta}_{MCDE} \tag{24}$$

$$\hat{V}_a\left(\hat{\beta}_{EF}\right) = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\ddot{x}_{it}\right)^{-1}\left\{\sum_{i=1}^{N}\left[\left(\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\hat{u}_{it}\right)\left(\sum_{t=1}^{T} s_{it}x'_{it}\hat{u}_{it}\right)'\right]\right\}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}x'_{it}x_{it}\right)^{-1} \tag{25}$$

$$\text{où } \hat{u}_{it} = y_{it} - x_{it}\hat{\beta}_{EF}$$

$$\hat{V}_a\left(\hat{\beta}_{EF-MCDE}\right) = \left(\ddot{C}'\ddot{D}^{-1}\ddot{C}\right)^{-1}\ddot{C}'\ddot{D}^{-1}\ddot{B}\ddot{D}^{-1}\ddot{C}\left(\ddot{C}'\ddot{D}^{-1}\ddot{C}\right)^{-1} \tag{26}$$

$$\text{où } \ddot{C} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{x}'_{it}\ddot{z}_{it} \tag{27}$$

$$\ddot{D} = \frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\ddot{z}_{it} \tag{28}$$

$$\ddot{B} = \frac{1}{N}\sum_{i=1}^{N}\left[\left(\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\hat{u}_{it}\right)\left(\sum_{t=1}^{T} s_{it}\ddot{z}'_{it}\hat{u}_{it}\right)'\right] \text{ où } \hat{u}_{it} = \ddot{y}_{it} - \ddot{x}_{it}\hat{\beta}_{EF-MCDE}\,. \tag{29}$$

Aussi, un terme de correction pour les degrés de liberté multiplie la matrice de variance covariance des estimateurs :

$$d = \frac{n-1}{n-K}\cdot\frac{N}{N-1} \text{ où } n = \sum_{i=1}^{N}\sum_{t=1}^{T} s_{it} \tag{30}$$

et N correspond au nombre de personnes de l'échantillon.

Enfin, l'estimateur avec effets aléatoires (EA) contrôle pour l'hétérogénéité inobservée tout en minimisant la variance (Wooldridge, 2002, p. 261-262 et 287) :

$$\hat{\beta}_{EA} = \left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\breve{x}'_{it}\breve{x}_{it}\right)^{-1}\left(\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it}\breve{x}'_{it}\breve{y}_{it}\right) \tag{31}$$

où $\breve{x}_{it} = x_{it} - \hat{\lambda}\bar{x}_i$, $\breve{y}_{it} = y_{it} - \hat{\lambda}\bar{y}_i$ pour t = 1,…,T et i = 1,…,N

et $\hat{\lambda} = 1 - \left[1\Big/\left(1 + T\left(\frac{\hat{\sigma}_c}{\hat{\sigma}_u}\right)\right)\right]^{1/2}$

$\hat{\sigma}_v^2 = \dfrac{1}{(NT-K)}\sum_{i=1}^{N}\sum_{t=1}^{T}\hat{\hat{v}}_{it}^2$ $\hat{\hat{v}}_{it}$ correspond aux résidus MCO de l'équation (6)

$\hat{\sigma}_c^2 = \dfrac{1}{[NT(T-1)/2-K]}\sum_{i=1}^{N}\sum_{t=1}^{T-1}\sum_{s=t+1}^{T}\hat{\hat{v}}_{it}\hat{\hat{v}}_{is}$

$\hat{\sigma}_u^2 = \hat{\sigma}_v^2 - \hat{\sigma}_c^2$.

Par contre, l'estimateur EA requiert deux hypothèses fortes (Wooldridge, 2002, p. 257) :

i- $E(u_{it}|x_i, c_i) = 0$, t = 1,…,T et où $x_i = (s_{i1}x_{i1},...,s_{iT}x_{iT})$;
ii- $E(c_i|x_i) = E(c_i|) = 0$.

L'hypothèse d'exogénéité stricte des variables explicatives conditionnelles à c_i (i) implique que $E(y_{it}|x_i, c_i) = x_{it}\beta_{EA} + c_i$. Donc, les valeurs observées des autres périodes x_{is} pour $s \neq r$ n'ont pas d'impact sur la variable expliquée (Wooldridge, 2002, p. 253). Dans une équation salariale, l'hypothèse (i) signifierait que l'expérience des autres périodes n'aurait pas d'impact sur le salaire d'une période t. L'hypothèse (ii) annule toute influence de c_i sur les variables explicatives. Dans notre exemple précédent, les habiletés non observées ne pourraient augmenter le rendement de l'expérience sur le salaire. Deux autres hypothèses sont aussi nécessaires (Wooldridge, 2002, p. 258-259) :

iii- rang $E(x_i'\Omega x_i) = K$ où Ω est une matrice K × K dont les éléments diagonaux sont $\sigma_c^2 + \sigma_u^2$ et les autres éléments sont σ_c^2 ;
iv- $E(u_i u_i'|x_i, c_i) = \sigma_u^2 I_{T_i}$ et $E(c_i^2|x_i) = \sigma_c^2$.

Cette procédure peu plausible nous permettra tout de même de comparer le rendement de l'expérience en effets fixes et aléatoires à l'aide du test de Hausman :

$$H = \frac{\left(\hat{\beta}_{\exp EF} - \hat{\beta}_{\exp EA}\right)^2}{\left(\text{var}\left(\hat{\beta}_{\exp EF}\right) - \text{var}\left(\hat{\beta}_{\exp EA}\right)\right)} \overset{a}{\sim} \chi_1^2 . \qquad (32)$$

Si l'hypothèse nulle est rejetée et que les deux coefficients sont statistiquement différents, alors seul l'estimateur $\hat{\beta}_{\exp EF}$ est convergent puisque c_i est probablement corrélé avec u_{it}. Si c_i n'est pas corrélé avec u_{it} tel que le prescrit l'hypothèse (i) ci-dessus, l'écart entre la méthode avec effets fixes et effets aléatoires s'amenuise à mesure que le nombre d'observations augmente. La même statistique sera calculée pour comparer l'estimateur $\hat{\beta}_{\exp EF-MCDE}$ et $\hat{\beta}_{\exp EA-VI}$. La méthode à effets aléatoires avec variables instrumentales (EA-VI) de Baltagi et Chang (2000) servira pour cette comparaison. Elle demeure complexe et dépasse les objectifs de ce travail. En somme, ces différentes méthodes d'estimation permettront d'examiner les écarts entre les rendements de l'expérience en fonction des corrections pour l'hétérogénéité inobservée et l'endogénéité. Il reste toutefois à détecter la présence d'un biais de sélection et s'il y a lieu, de le corriger.

2.3 Détection du biais de sélection

Il y a présence d'un biais de sélection lorsque u_{it} est corrélé avec les indicateurs de sélection s_{ir} pour tous les t et les r. Le biais de sélection se divise en un biais inter temporel entre différentes périodes et en un biais contemporain qui concerne l'erreur et l'indicateur de sélection d'une même période. Pour détecter le biais inter temporel, Semykina et Wooldrige (2006, p. 12) proposent d'ajouter un régresseur à l'équation (2) qui varie selon le temps : $s_{i,t-1}$, $s_{i,t+1}$, $\sum_{r=1}^{t-1} s_{ir}$ ou $\sum_{r=t+1}^{T} s_{ir}$. Par exemple, l'indicateur de sélection $s_{i,t-1}$ prend la valeur de 1 si ($y_{i,t-1}$, $x_{i,t-1}$, $z_{i,t-1}$) sont observés. Si le test statistique confirme la significativité du coefficient de $s_{i,t-1}$, alors le vecteur de

paramètres $\hat{\beta}_{EF-MCDE}$ de l'équation (18) n'est pas convergent à cause d'un biais de sélection inter temporel.

Par contre, pour le biais de sélection contemporain, on ne peut ajouter s_{it} au temps t et procéder au même test, car $s_{it}=1$ pour tous les t dans l'échantillon (Semykina et Wooldridge, 2006, p. 13). Alors, Semykina et Wooldridge reviennent à l'équation d'intérêt :

$$y_{it} = x_{it}\beta + v_{it} . \tag{6}$$

Dans le cas de données de panel et de variables explicatives exogènes, Verbeek et Nijman (1992) modélisent $E(v_{it} \mid z_{it}, s_{it})$ dans l'équation (6) et testent la significativité du coefficient (Semykina et Wooldridge, 2006, p. 13). Cependant, si un effet non observé se trouve dans l'erreur v_{it}, nous pourrions conclure par ces tests qu'il y a un biais de sélection seulement parce que l'effet non observé est corrélé avec une variable explicative. Il faut donc tenir compte de l'effet fixe.

À partir de la méthode de Wooldridge (1995b) qui ne considérait pas l'endogénéité, Semykina et Wooldrige (2006, p. 13) élaborent un test de biais de sélection contemporain pour le cas de régressions instrumentales à effets fixes. Contrairement aux régressions par effets aléatoires, cette méthode par effets fixes permet une corrélation entre la sélection et l'effet non observé, soit entre s_{it} et c_i (Semykina et Wooldridge, 2006, p. 13). L'équation (6) devient en ajoutant les effets fixes :

$$y_{it1} = x_{it1}\beta_1 + c_{i1} + u_{it1}, \text{ t} = 1,\ldots,\text{T}, \tag{33}$$

où x_{it1} est un vecteur $1 \times K$ contenant les variables explicatives dont quelques-unes endogènes, β_1 est un vecteur de paramètres $K \times 1$, z_{it} est le vecteur des instruments

$1 \times L$, c_{i1} est l'effet non observé et u_{it1} est l'erreur idiosyncrasique. L'ajout du « 1 » souligne l'absence d'hypothèse quant à la forme fonctionnelle de c_i et u_{it}.

Par conséquent, l'objectif est de déterminer si l'indicateur de sélection s_{it1} est corrélé avec l'erreur idiosyncrasique u_{it1}. Avant de poursuivre, Semykina et Wooldridge (2006, p. 14) supposent que les instruments z_{it} sont toujours observés et qu'un nouvel indicateur de sélection modélisé ou hypothétique s_{it2} prend la valeur de 1 si (y_{it1},x_{it1}) sont observés. Afin d'isoler l'effet fixe c_{i1} à l'intérieur du terme d'erreur v_{it}, Semykina et Wooldridge (2006, p. 14) doivent modéliser la relation entre c_{i1} et les instruments z_{it}. S'inspirant de Mundlak (1978), ils posent :

$$c_{i2} = \eta_2 + \bar{z}_i \xi_2 + a_{i2}. \tag{34}$$

Bref, par hypothèse, l'effet fixe modélisé c_{i2} dépend linéairement des moyennes inter temporelles des instruments exogènes $\bar{z}_i$.

Outre l'exogénéité des instruments et la forme fonctionnelle de l'effet fixe, Semykina et Wooldridge (2006, p. 14) font deux hypothèses supplémentaires. Quatre hypothèses principales sous-tendent ainsi leur méthode. D'une part, ils définissent la fonction permettant de générer un indicateur de sélection latent s^*_{it2}. Dès lors, l'indicateur de sélection hypothétique s_{it2} prend la valeur de 1 lorsque l'indicateur de sélection latente s^*_{it2} dépassera 0. D'autre part, le terme d'erreur de la fonction de l'indicateur de sélection latent une fois l'effet fixe c_{i2} substitué donne : $v_{it2} = a_{i2} + u_{it2}$. Semykina et Wooldridge (2006, p. 15) postulent que v_{it2} est distribué selon la loi normale conditionnellement aux instruments z_i. Ils arrivent donc aux conclusions suivantes :

$$s^*_{it2} = z_{it}\delta_2 + c_{i2} + u_{it2}, \text{ t} = 1,\ldots,\text{T}, \tag{35}$$

$$s_{it2} = 1\left[s_{it2}^{*} \rangle 0\right] = 1\left[z_{it}\delta_2 + c_{i2} + u_{it2} \rangle 0\right] = 1\left[\eta_{t2} + z_{it}\delta_{t2} + \bar{z}_i\xi_{t2} + v_{it2} \rangle 0\right], \quad (36)$$

$$v_{it2} \mid z_i \sim \text{Normal}(0,1),\ t = 1,\ldots,T. \quad (37)$$

Ici, Semykina et Wooldridge (2006, p. 16) supposent en plus que (u_{it1}, v_{i2}) est indépendant de (z_i, c_{i1}), où $v_{i2} = (v_{i12},...,v_{iT2})'$ et (u_{it1}, v_{it2}) est indépendant de $(v_{i12},...,v_{i,t-1,2}, v_{i,t+1,2},...,v_{iT2})$.

Ensuite, les termes de correction proviennent d'un probit effectué pour chaque période soit (Semykina et Wooldridge, 2006, p. 17) :

$$P(s_{it2} = 1 \mid z_i) = \Phi(\eta_{t2} + z_{it}\delta_{t2} + \bar{z}_i\xi_{t2}). \quad (38)$$

Semykina et Wooldridge s'inspirent alors d'Heckman (1976, 1979) et utilisent les prédictions du probit pour calculer le ratio inverse de Mill. Ce ratio est une fonction monotone décroissante de la probabilité qu'une observation se trouve dans l'échantillon (Heckman, 1979, p. 156). Les termes de correction estimés $\hat{\lambda}_{it2}$ sont donc obtenus ainsi :

$$E(v_{it2} \mid z_i, s_{it2} = 1) = \lambda(\eta_{t2} + z_{it}\delta_{t2} + \bar{z}_i\xi_{t2}), \quad (39)$$

$$\hat{\lambda}_{it2} = \lambda\left(\hat{\eta}_{t2} + z_{it}\hat{\delta}_{t2} + \bar{z}_i\hat{\xi}_{t2}\right) = \frac{\phi\left(\hat{\eta}_{t2} + z_{it}\hat{\delta}_{t2} + \bar{z}_i\hat{\xi}_{t2}\right)}{1 - \Phi\left(\hat{\eta}_{t2} + z_{it}\hat{\delta}_{t2} + \bar{z}_i\hat{\xi}_{t2}\right)} = \frac{\phi\left(\hat{\eta}_{t2} + z_{it}\hat{\delta}_{t2} + \bar{z}_i\hat{\xi}_{t2}\right)}{\Phi\left(-\hat{\eta}_{t2} - z_{it}\hat{\delta}_{t2} - \bar{z}_i\hat{\xi}_{t2}\right)}, \quad (40)$$

où φ représente la fonction de densité d'une distribution normale centrée réduite alors que Φ fait référence à la fonction cumulative. Préférablement, l'interaction entre une variable dichotomique de temps et les termes $\hat{\lambda}_{it2}$ permettra d'estimer un facteur de correction pour le biais de sélection contemporain différent à chaque période. De ce fait, l'équation de départ (33) peut se réécrire comme une équation primaire estimée à l'aide de l'équation (18) :

$$y_{it1} = x_{it}\beta + c_{i1} + \rho_1 E(v_{it2} \mid z_i, s_{it2}) + e_{it1},\ \text{t=1,…,T.} \tag{41}$$

Puisque le calcul de l'estimateur $\hat{\beta}_{EF-MCDE}$ rejette les observations avec des données manquantes pour lesquelles s_{it2}=0, l'échantillon ne contient que des observations où s_{it2}=1. Nous pouvons alors réécrire :

$$y_{it1} = x_{it}\beta + c_{i1} + \rho_1 E(v_{it2} \mid z_i, s_{it2} = 1) + e_{it1},\ \text{t=1,…,T,} \tag{42}$$

$$y_{it1} = x_{it}\beta + c_{i1} + \rho_t \hat{\lambda}_{it2} + e_{it1},\ \text{t=1,…,T.} \tag{43}$$

Enfin, il suffit d'estimer le $\hat{\beta}_{EF-MCDE}$ de l'équation (18) en incluant les termes $\hat{\lambda}_{it2}$ dans les variables explicatives comme dans les variables instrumentales (Semykina et Wooldridge, 2006, p. 17). Si le test joint sur les paramètres calculés $\hat{\rho}_t$ rejette l'hypothèse nulle $H_0 : \rho_1 = 0, \rho_2 = 0, ..., \rho_t = 0$, alors il y a présence d'un biais de sélection contemporain ce qui nécessite une procédure de correction. En somme, cette section complète la description de la procédure #1 qui détecte le biais de sélection contemporain, c'est-à-dire dans une même période.

2.4 Correction du biais de sélection

Afin de corriger le biais de sélection, Semykina et Wooldridge (2006, p. 18) vont faire une suite d'hypothèses leur permettant d'estimer les paramètres précédents par une régression instrumentale ordinaire et ce en tenant aussi compte des effets fixes, donc de l'hétérogénéité. Celles-ci sont (Semykina et Wooldridge, 2006, p. 18-19) :

i- z_{it} est toujours observé alors que (x_{it1},y_{it1}) est observé quand s_{it2}=1;
ii- la sélection se conforme aux équations précédentes (36) et (37);
iii- $E(u_{it1} \mid z_i, v_{it2}) = E(u_{it1} \mid v_{it2}) = \rho_{t1} v_{it2}$, t = 1,…,T;
iv- $c_{i1} = \eta_1 + \bar{z}_i \xi_1 + a_{i1}$ où $E(a_{i1} \mid z_i, v_{it2}) = E(a_{i1} \mid v_{it2}) = \phi_{t1} v_{it2}$, t = 1,…,T.

En substance, après avoir formalisé le processus de sélection (ii), supposé que les instruments étaient non corrélés aux erreurs idiosyncrasiques u_{it1} (iii) et défini la forme fonctionnelle que prend l'effet fixe non observé (iv), Semykina et Wooldridge (2006, p. 20) arrivent à une équation de départ (33) transformée, applicable sans problème à une régression MCDE :

$$y_{it1} = x_{it1}\beta_1 + c_{i1} + u_{it1} \qquad t = 1,\ldots,T, \qquad (33)$$

$$y_{it1} = x_{it1}\beta_1 + \left[\eta_1 + \bar{z}_i\xi_1 + a_{i1}\right] + u_{it1} = x_{it1}\beta_1 + \eta_1 + \bar{z}_i\xi_1 + v_{it1} \qquad t = 1,\ldots,T, \qquad (44)$$

$$y_{it1} = x_{it1}\beta_1 + \eta_1 + \bar{z}_i\xi_1 + \gamma_{t1}E\left(v_{it2} \mid z_i, s_{it2}\right) + e_{it1} \qquad t = 1,\ldots,T, \qquad (45)$$

$$y_{it1} = x_{it1}\beta_1 + \eta_1 + \bar{z}_i\xi_1 + \gamma_{t1}\lambda_{it2} + e_{it1} \qquad t = 1,\ldots,T. \qquad (46)$$

La notation utilise le « 1 » probablement à cause que la procédure n'est plus un test de biais, mais fournira des résultats valides cette fois. Avec une telle approche, nous pouvons ignorer sans problème la corrélation de l'erreur idiosyncrasique e_{it1} avec les indicateurs de sélection des autres périodes s_{ir2}, $r \neq t$ (Semykina et Wooldridge, 2006, p. 19). Par conséquent, il ne faut pas prétendre que $E\left(e_{it1} \mid z_i, s_{i2}\right) = 0$. Plutôt, les hypothèses précédentes impliquent seulement que $E\left(e_{it1} \mid z_i, s_{it2}\right) = 0$, t = 1,…,T (Semykina et Wooldridge, 2006, p. 19). En d'autres mots, la procédure #2 qui corrige pour le biais de sélection contemporain tolère en plus l'existence d'un biais de sélection inter temporel. Aussi, aucune hypothèse n'est faite sur la distribution de cette erreur (Semykina et Wooldridge, 2006, p. 6).

Ainsi, selon les hypothèses (i-iv) de l'estimateur $\hat{\beta}_{MCDE}$ (11) et les hypothèses (i-iv) précédentes pour la correction du biais de sélection, l'estimateur $\hat{\beta}_{MCDE}$ de la procédure #2 sera convergent et asymptotiquement normal. Dès lors, la correction du biais de sélection s'obtient en définissant d'une part un vecteur $\hat{w}_{it} = \left(1, x_{it1}, \bar{z}_i, 0,\ldots,0, \hat{\lambda}_{it2}, 0,\ldots,0\right)$ contenant une constante, les variables explicatives,

les moyennes des variables instrumentales et les termes de correction $\hat{\lambda}_{it2}$. D'autre part, le vecteur corrigeant l'endogénéité est $\hat{h}_{it} = \left(1, z_{it}, \bar{z}_i, 0, ..., 0, \hat{\lambda}_{it2}, 0, ..., 0\right)$ qui comporte une constante, les variables instrumentales, leur moyenne et les termes de correction $\hat{\lambda}_{it2}$. Le vecteur estimé des paramètres $\theta = \left(\eta_1, \beta_1', \xi_1', \gamma_1, ..., \gamma_T\right)'$ de dimension $\left(1 + K + L + T\right) \times 1$ devient donc (Semykina et Wooldridge, 2006, p. 28) :

$$\hat{\theta} = \left[\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{w}_{it}'\hat{h}_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{h}_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{w}_{it}\right)\right]^{-1}$$
$$\times\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{w}_{it}'\hat{h}_{it}\right)\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}'\hat{h}_{it}\right)^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{h}_{it}' y_{it}\right). \quad (47)$$

Le calcul de l'écart type exige beaucoup plus de calculs (*voir* ann. B).

Enfin, à l'aide des quelques hypothèses précédentes, la procédure de Semykina et Wooldridge (2006) sert à contourner les problèmes d'hétérogénéité, d'endogénéité et de biais de sélection. La procédure #1 détecte le biais de sélection contemporain alors que la procédure #2 le corrige et demeure robuste au biais de sélection inter temporel. Pour ce faire, Semykina et Wooldridge (2006) posent l'hypothèse cruciale et non vérifiable dans l'équation (34) que l'effet fixe c_{i2} dépend linéairement des moyennes inter temporelles des instruments exogènes $\bar{z}_i$. Dans les tests comme dans les deux estimateurs précédents, les erreurs de l'équation de sélection (v_{it2}) doivent être distribuées selon la loi normale asymptotiquement, mais peuvent contenir un effet fixe et être corrélées entre les périodes (Semykina et Wooldridge, 2006, p. 6). Aussi, aucune hypothèse ne porte sur la distribution des erreurs de l'équation primaire (2) alors qu'elles peuvent être hétéroscédastiques et corrélées dans le temps. Cependant, pour véritablement isoler les effets des instruments dont la variable dichotomique de notre politique d'intérêt, il faudra revenir à la procédure de

Wooldridge (1995b) qui ne traite que des variables explicatives exogènes, c'est-à-dire sans instrument.

2.5 Utilisation de la forme réduite pour estimer l'effet de la politique d'intérêt

Une équation de forme réduite est obtenue lorsque les variables endogènes sont remplacées par les variables instrumentales exogènes du modèle (Wooldridge, 2002, p. 213). Cette équation s'applique alors tout à fait à la procédure de Wooldridge (1995b) qui ressemble à celle de Semykina et Wooldridge (2006) à l'exception qu'il n'y a plus d'instrument.

Ainsi, les équations s'apparentent aux sections 2.2 et 2.3. La régression à effets fixes s'écrit donc (Wooldridge, 1995b, p. 118) :

$$\hat{\beta}_{EF} = \left(N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} \ddot{x}'_{it} \ddot{x}_{it} \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} \ddot{x}'_{it} \ddot{y}_{it} \right) \tag{48}$$

où $\ddot{x}_{it} = x_{it} - T_i \sum_{r=1}^{T} s_{ir} x_{ir}$ et $T_i = \sum_{t=1}^{T} s_{it}$.

De même, le probit de l'équation de sélection ne s'effectue que sur les variables explicatives exogènes du modèle et les termes de correction $\hat{\lambda}_{it}$ se calculent en conséquence (Wooldridge, 1995b, p. 123) :

$$P(s_{it} = 1 \mid x_i) = \Phi(x_i \delta_t) \tag{49}$$

$$\text{où } \hat{\lambda}_{it} = \lambda(x_i \hat{\delta}_t) = \frac{\phi(x_i \hat{\delta}_t)}{1 - \Phi(x_i \hat{\delta}_t)} = \frac{\phi(x_i \hat{\delta}_t)}{\Phi(-x_i \hat{\delta}_t)}. \tag{50}$$

Les termes de correction $\hat{\lambda}_{it}$ seront inclus dans la régression à effets fixes de la même manière et le test sera le même soit $H_0 : \rho_1 = 0, \rho_2 = 0, ..., \rho_t = 0$ (Wooldridge, 1995b, p. 123).

Enfin, la correction s'effectue presque de la même manière (*voir* ann. C pour le calcul de l'écart type) (Wooldridge, 1995b, p. 129-130) :

$$\hat{\theta} = \left(N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} \hat{w}'_{it} \hat{w}_{it} \right)^{-1} \left(N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it} \hat{w}'_{it} y_{it} \right) \tag{51}$$

où $\hat{w}_{it} = \left(1, x_{i1}, ..., x_{iT}, x_{it}, 0, ..., 0, \hat{\lambda}_{it}, 0, ..., 0\right)$.

La méthode de Wooldridge (1995b) qui s'applique à des variables strictement exogènes servira à calculer l'effet marginal des divers instruments sur le salaire. L'impact et la significativité de la variable dichotomique de la politique de garde du Québec pourront donc être étudiés. Ainsi, la procédure #3 (48) inclut les termes de correction $\hat{\lambda}_{it}$ et détecte le biais de sélection contemporain. La procédure #4 (51) contient les termes de correction $\hat{\lambda}_{it}$ et les moyennes inter temporelles des variables explicatives ce qui corrige pour ce problème.

Pour finir, le langage matriciel Mata de Stata 10 a servi à la mise au point de notre programme permettant d'obtenir des résultats conformes à la démarche de Semykina et Wooldridge (2006) (*voir* ann. E). Appliqué sur de petites bases de données, il fournit les mêmes estimations que le programme que Mme Semykina a bien voulu nous faire parvenir. Seulement, son programme utilise le langage matriciel standard de Stata et n'est pas assez puissant pour traiter les dizaines de milliers d'observations de notre échantillon de 1993 à 2004.

2.6 Tests de la faiblesse et de la validité des instruments

Les résultats empiriques des procédures de Semykina et Wooldridge (2006) ne comprennent que les rendements de l'expérience et de l'éducation (*voir* tableau 3.5). Pourtant, les nombreux régresseurs exogènes exclus utilisés comme instruments dans la procédure #1 et #2 doivent être suffisamment corrélés avec les régresseurs endogènes, mais non corrélés avec le terme d'erreur du modèle. À la différence de Semykina et Wooldridge (2006), notre travail aborde ces deux facettes de l'usage d'instruments. Les tests de faiblesse et de validité des instruments proviennent des nouvelles commandes *estat firststage* et *estat overid* de Stata 10, mais notre propre programme (*voir* ann. E p. 129-133) réplique parfaitement leurs résultats (StataCorp, 2007).

Au cours des dernières décennies, les instruments faiblement corrélés avec les régresseurs endogènes ont attiré l'attention de maints chercheurs. Dans ce cas, les estimateurs MCDE ou MMG sont biaisés vers l'estimateur MCO et les inférences à l'aide des écarts-types deviennent erronées. Stock *et al.* (2002) présentent une intéressante compilation d'articles sur le sujet. En effet, la complexité du problème suscite beaucoup d'interrogations et il ne suffit pas d'ajouter plus d'instruments pour le solutionner. Au contraire, le biais de l'estimateur MCDE ou MMG augmente avec le nombre d'instruments (Hahn et Hausman, 2003, p. 119). Il faut donc savoir non seulement détecter les instruments faibles parmi ceux utilisés, mais les remplacer par de meilleurs. Notre travail se limite à évaluer la liste d'instruments suggérés par Semykina et Wooldridge (2006). Les méthodes suivantes ne sont pas spécifiquement conçues pour des données de panel et la notation cesse donc d'utiliser les indices i et t.

D'abord, nous présentons pour chaque régresseur endogène pris séparément le R^2, le R^2 ajusté, le R^2 partiel, la statistique F et la p-value associée de la première étape de la régression MCDE. Dans l'application empirique de Semykina et Wooldridge (2006, p. 24), les deux variables explicatives endogènes comprennent l'expérience et l'expérience au carré. Le R^2 exprime le pourcentage de la variance d'un des régresseurs endogènes expliqué par les instruments dans la régression MCO. Le R^2 ajusté corrige pour l'augmentation du R^2 dû au simple fait d'inclure davantage de régresseurs. Ainsi, l'équation suivante illustre la première étape de l'estimateur MCDE :

$$x_{endo} = Z_1\pi_1 + Z_2\pi_2 + e_1 \tag{52}$$

où x_{endo} est un vecteur $n \times 1$ d'observations sur un seul régresseur endogène, Z_1 est une matrice $n \times l_1$ de régresseurs exogènes inclus aussi dans les variables explicatives, Z_2 est une matrice $n \times l_2$ de régresseurs exogènes exclus des variables explicatives et e_1 est un terme d'erreur. Les deux formules suivantes calculent le R^2 et le R^2 ajusté (Greene, 2008, p. 34-36) :

$$R^2 = \frac{\sum_{j=1}^{n}\left(\hat{x}_j - \bar{x}\right)^2}{\sum_{j=1}^{n}\left(x_j - \bar{x}\right)^2} \text{ où } \hat{x}_j = Z_{1j}\hat{\pi}_1 + Z_{2j}\hat{\pi}_2 \text{ et } \bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j \tag{53}$$

$$R^2 ajusté = 1 - \frac{n-1}{n-\left(l_1 + l_2\right)}\left(1 - R^2\right). \tag{54}$$

De hautes valeurs de ces deux statistiques indiquent que les instruments expliquent une grande partie de la variable endogène. La statistique F du test de Wald vérifie l'hypothèse jointe $H_0 : \pi_2 = 0$ ou alors que les paramètres des régresseurs exogènes exclus $\left(\hat{\pi}_2\right)$ sont statistiquement significatifs. Normalement, Stock *et al.* (2002, p. 522) suggèrent une statistique F supérieure à 10 pour faire de l'inférence avec la

méthode MCDE. La p-value associée à ce test rapporte la probabilité de rejetée l'hypothèse nulle (H_0) alors qu'elle est vraie. Une valeur sous les 10 %, 5 % ou 1 % confirme la pertinence des régresseurs exogènes exclus (Z_2).

Toutefois, observer uniquement ces valeurs ne suffit pas. Si le régresseur endogène (x_{endo}) est fortement corrélé avec les régresseurs exogènes inclus (Z_1), mais faiblement avec les régresseurs exogènes exclus (Z_2), alors la faiblesse des instruments comptant réellement pour corriger l'endogénéité passera inaperçue. Le R^2 partiel reprend le calcul du R^2, mais ne tient compte que des composantes orthogonales aux régresseurs exogènes inclus (Z_1) (StataCorp, 2007, p. 54) :

$$M_{Z_1} x_{endo} = M_{Z_1} Z_2 \xi + e_2 \tag{55}$$

où $M_{Z_1} = I - Z_1 (Z_1' Z_1)^{-1} Z_1'$, ξ est un vecteur de paramètres $l_2 \times 1$ et e_2 est un terme d'erreur. La matrice M_{Z_1} permet d'éliminer l'effet des régresseurs exogènes inclus (Z_1) en ne conservant que les résidus d'une régression MCO du terme multiplié par M_{Z_1}.

$$R^2 \, partiel = 1 - \frac{\sum_{j=1}^{n} \hat{e}_{2j}^2}{\sum_{j=1}^{n} \left(\left[M_{Z_1} x_{endo} \right]_j - \bar{x}_{M_{Z_1}} \right)^2} \tag{56}$$

$$\text{où } \hat{e}_2 = M_{Z_1} x_{endo} - M_{Z_1} Z_2 \hat{\xi} \text{ et } \bar{x}_{M_{Z_1}} = \frac{1}{n} \sum_{j=1}^{n} \left[M_{Z_1} x_{endo} \right]_j .$$

En somme, le R^2 partiel représente le pourcentage de la variance du régresseur endogène expliqué par les régresseurs exogènes exclus (Z_2). Tous les tests précédents sont robustes à l'hétéroscédasticité et à la corrélation en série au niveau individuel.

Lorsqu'un modèle contient plus d'un régresseur endogène, les statistiques précédentes peuvent générer de mauvaises conclusions. Si un régresseur exogène exclus z_1 est fortement corrélé avec deux régresseurs endogènes x_{endo1} et x_{endo2} et qu'un autre z_2 l'est peu, alors la régression MCO de x_{endo1} et de x_{endo2} sur Z [z_1 z_2] produira un R^2 élevé même si les paramètres ne sont pas identifiés à cause de la faiblesse de z_2 (Shea, 1997, p. 348). Même si le modèle est suridentifié, l'inclusion d'instruments faibles affectera aussi les propriétés des paramètres puisque le biais de variables instrumentales augmente avec leur nombre. Soient Z la matrice $n \times L$ contenant l'ensemble des instruments inclus et exclus (Z_1 et Z_2), x_{endo1} le régresseur endogène dont la statistique est calculée et X_{endo0} les autres régresseurs endogènes, alors le R^2 de Shea $\left(R_S^2\right)$ se calcule ainsi (StataCorp, 2007, p. 54 et Shea, 1997, p. 348-349) :

$$\widetilde{x}_{endo1} = M_H x_{endo1} = x_{endo1} - H\left(H'H\right)^{-1}\left(H'x_{endo1}\right) \text{ où } H = \left[X_{endo0}\ Z_1\right] \tag{57}$$

$$\hat{x}_{endo1} = P_Z x_{endo1} = Z\left(Z'Z\right)^{-1}\left(Z'x_{endo1}\right) \tag{58}$$

$$\hat{X}_{endo0k} = P_Z X_{endo0k} = Z\left(Z'Z\right)^{-1}\left(Z'X_{endo0k}\right) \text{ où k = 1,…,K-1} \tag{59}$$

$$\hat{X}_{endo0} = \left[\hat{X}_{endo1}, \hat{X}_{endo2}, ..., \hat{X}_{endoK-1}\right] \tag{60}$$

$$\widetilde{\hat{x}}_{endo1} = M_{\hat{H}}\hat{x}_{endo1} = \hat{x}_{endo1} - \hat{H}\left(\hat{H}'\hat{H}\right)^{-1}\left(\hat{H}'\hat{x}_{endo1}\right) \text{ où } \hat{H} = \left[\hat{X}_{endo0}\ Z_1\right] \tag{61}$$

$$\hat{\varepsilon} = M_{\widetilde{\hat{x}}_{endo1}}\widetilde{x}_{endo1} = \widetilde{x}_{endo1} - \widetilde{\hat{x}}_{endo1}\left(\widetilde{\hat{x}}'_{endo1}\widetilde{\hat{x}}_{endo1}\right)^{-1}\left(\widetilde{\hat{x}}'_{endo1}\widetilde{x}_{endo1}\right) \tag{62}$$

$$R_S^2 = 1 - \frac{\sum_{j=1}^{n}\hat{\varepsilon}_j^2}{\sum_{j=1}^{n}\left(\widetilde{x}_{endo1\,j} - \overline{\widetilde{x}}_{endo1}\right)} \text{ où } \overline{\widetilde{x}}_{endo1} = \frac{1}{n}\sum_{j=1}^{n}\widetilde{x}_{endo1\,j} \tag{63}$$

$$R_S^2 ajusté = 1 - \frac{n-1}{n-L+1}\left(1 - R_S^2\right). \tag{64}$$

Essentiellement, Shea (1997, p. 348) souligne que la composante de la valeur prédite d'un régresseur endogène qui est orthogonale aux valeurs prédites des autres

régresseurs $\left(\tilde{\hat{x}}_{endo1}\right)$ devrait pouvoir expliquer la composante des observations du régresseur endogène qui est orthogonale aux autres régresseurs $\left(\tilde{x}_{endo1}\right)$. Il faut préciser que les valeurs prédites des autres régresseurs $\left(\hat{H}\right)$ comprennent en fait les valeurs prédites des autres régresseurs endogènes $\left(\hat{X}_{endo0}\right)$ et les valeurs observées des régresseurs exogènes inclus $\left(Z_1\right)$. Les résidus de cette régression par MCO $\left(\hat{\varepsilon}\right)$ servent donc au calcul du R^2 de Shea $\left(R_S^2\right)$. Puisque le biais de variables instrumentales s'accroît à mesure que le nombre d'instrument augmente, le R^2 de Shea ajusté opère la même correction que le R^2 ajusté standard. La valeur idéale pour ces deux dernières statistiques dépend du type de modèle et du genre de données (StataCorp, 2007, p. 51). Shea n'offre donc aucune définition formelle d'instrument faible. Enfin, ces deux statistiques sont robustes à l'hétéroscédasticité et à la corrélation en série au niveau individuel.

Ensuite, le test de Stock et Yogo (2005, p. 84) pose l'hypothèse nulle que les instruments sont faibles. Elle est rejetée si la valeur Eigen minimale $\left(g_{\min}\right)$ proposée par Cragg et Donald (1993) est supérieure à une valeur critique donnée.

$$G = \frac{1}{L}\hat{\Sigma}_{VV}^{-1/2} X'_{endo} M'_{Z_1} Z_2 \left(Z'_2 M_{Z_1} Z_2\right)^{-1} Z'_2 M_{Z_1} X_{endo} \hat{\Sigma}_{VV}^{-1/2} \tag{65}$$

$$\text{où } \hat{\Sigma}_{VV}^{-1/2} = \frac{1}{n-L} X'_{endo} M_Z X_{endo} \text{ et } M_Z = I - Z\left(Z'Z\right)^{-1} Z'$$

$$g_{\min} = \min eval\left(G\right) \tag{66}$$

Stock et Yogo (2005, p. 100-101) fournissent les valeurs critiques dans les tableaux 5.1 et 5.2. Elles peuvent dépendre de l'une des deux caractérisations d'instruments faibles qu'ils proposent (Stock et Yogo, 2005, p. 81). D'une part, des instruments sont faibles si le biais relatif de l'estimateur VI par rapport à l'estimateur MCO dépasse un certain seuil en pourcentage (b). Si la valeur Eigen minimale dépasse la

valeur critique du biais relatif à b = 5 %, alors les instruments ne sont pas faibles et causent un biais d'au plus 5 % par rapport à l'estimateur MCO. D'autre part, Stock et Yogo (2005, p. 87-88) considèrent les instruments faibles si un test de Wald à un niveau de $\alpha = 5\%$ a un taux de rejet supérieur à r = 10 %, 15 %, 20 % ou 25 %. Stock et Yogo (2005, p. 87) estiment le taux de rejet à l'aide d'une fonction convergente et de la corrélation entre les erreurs de la première étape et de l'équation structurelle (ρ). La statistique de la valeur Eigen minimale fonctionne avec plus d'un régresseur endogène, mais n'est pas robuste à l'hétéroscédasticité. Elle offre par contre un second avis non négligeable.

Par ailleurs, les instruments doivent non seulement bien expliquer les régresseurs endogènes, mais aussi être exogènes, c'est-à-dire non corrélés avec le terme d'erreur structurel du modèle. Si le nombre de régresseurs exogènes exclus (l_2) dépassent le nombre de régresseurs endogènes (g), alors une procédure peut tester si ces $m = l_2 - g$ instruments supplémentaires sont statistiquement corrélés avec les erreurs. À partir de cette idée générale, Wooldridge (1995a) a développé un test des restrictions de suridentification robuste à l'hétéroscédasticité (StataCorp, 2007, p. 55) :

$\hat{X}_{endo} = P_Z X_{endo} = Z(Z'Z)^{-1}(Z'X_{endo})$ de dimension N × g (67)

$Q = [z_{21}\ z_{22} \cdots z_{2m}]$ où m est le nombre de conditions du suridentification (68)

$\hat{q}_j = M_P q_j = q_j - H(H'H)^{-1}(H'q_j)$ où j = 1,…,m et $H = [\hat{X}_{endo}\ Z_1]$ (69)

$k_{ij} = \hat{q}_{ij} u_i$ où i = 1,…,N et j = 1,…,M (70)

$\hat{k}_j = [k_{1j}\ k_{2j} \cdots k_{Nj}]'$ où j = 1,…,m (71)

$1 = \theta_1 \hat{k}_1 + \ldots + \theta_m \hat{k}_m + \varepsilon$ (72)

$W = N - \sum_{i=1}^{N} \varepsilon_i^2 \sim \chi^2(m).$ (73)

Le résultat du test ne dépend pas du choix des m régresseurs exogènes exclus sélectionnés parmi les l_2 régresseurs exogènes exclus disponibles pour former la matrice Q. Si la p-value associée à la statistique de test W est supérieure à un certain seuil normalement de 1 %, 5 % ou 10 %, alors l'hypothèse nulle selon laquelle les instruments sont valides n'est pas rejetée. Des instruments invalides découlent soit d'une corrélation positive entre un ou des instruments avec le terme d'erreur ou d'une forme structurelle mal spécifiée (StataCorp, 2007, p. 52). Autrement dit, un ou des régresseurs exogènes exclus devraient alors faire partie de l'équation structurelle puisqu'ils sont endogènes.

Finalement, le résumé de ces quelques procédures d'évaluation d'instruments nous permettra de mieux comprendre l'évolution du rendement de l'expérience sur le salaire fourni par nos méthodes d'estimation en plus d'examiner la robustesse de la correction pour l'endogénéité.

CHAPITRE III

APPLICATION ET RÉSULTATS

Ce chapitre présente les diverses facettes de l'application de la procédure de Semykina et Wooldridge (2006). L'analyse descriptive discute de l'évolution des variables d'intérêt de l'échantillon de 1993 à 2004. La présentation et la discussion des résultats terminent le chapitre.

3.1 Application des modèles

L'application de la procédure de Semykina et Wooldridge (2006) vise à déterminer l'impact sur le salaire de la politique de services de garde du Québec. L'intuition de départ découle simplement de l'observation par Lefebvre et Merrigan (2007b) que cette politique a influencé l'expérience, mais non le salaire. De ce fait, elle constitue probablement un bon instrument pour l'expérience, une variable typiquement endogène. La démarche qui suit tente de vérifier cette hypothèse.

Notre étude utilise l'EDTR, une enquête longitudinale de Statistique Canada instaurée en 1993. Elle permet depuis de suivre dans le temps des ménages, des familles et des individus canadiens et ce grâce à plusieurs variables liées au capital humain, à la démographie et au marché du travail. Statistique Canada sélectionne les participants de l'EDTR à partir de l'Enquête sur la population active (EPA) qui couvre toutes les provinces canadiennes à l'exception du Nunavut, des Territoires du

Nord-Ouest, du Yukon, des bases militaires, des réserves amérindiennes et des institutions. Les personnes sont interrogées au cours de six années consécutives. Le tableau 3.1 permet de mieux comprendre la structure longitudinale des données de l'EDTR qui sont disponibles jusqu'à maintenant :

Tableau 3.1 Structure et taille des échantillons[1] de l'EDTR.

Panel 1	93	94	95	96	97	98						
Panel 2				96	97	98	99	00	01			
Panel 3							99	00	01	02	03	04
Panel 4										02	03	04
Panel 5												
Personne	30237	30197	29819	61786	62118	62497	62335	61641	63475	60309	61320	59260
Famille	17255	17287	17147	35146	35547	35790	35610	35355	36414	34611	35170	34100
Unité	22,2	22,5	22,8	23,1	23,4	23,7	23,9	24,3	24,6	24,9	25,2	25,6

1. Nombre de personnes âgées de 15 ans et plus et de familles de recensement sans poids; nombre d'unités avec poids âgées de 15 ans et plus en millions.
Source : calculs de l'auteur selon l'EDTR de 1993 à 2004.

L'échantillon contient 84 968 observations sur 25 665 femmes âgées de 18 à 65 ans inclusivement sélectionnées à partir de l'EDTR pour les années 1993 à 2004. Le tableau 3.2 offre plus de détails sur la répartition des observations de l'échantillon selon les années et les panels. L'inclusion de femmes sans critère quant au nombre d'enfants a permis d'élargir l'échantillon et d'obtenir plus de variabilité pour les modèles à effets fixes.

Ensuite, Semykina et Wooldridge (2006, p. 24) se sont inspirés de Dustmann et Rochina-Barrachina (2000) pour construire leur spécification. Elle a servi à estimer le rendement de l'expérience dans une régression salariale avec l'étude américaine *Panel Study of Income Dynamics* (PSID) pour des femmes de 1980 à 1992. Elle s'applique donc bien à notre cadre d'analyse. Leur échantillon comporte 864 femmes et 11 232 observations dont 8 254 rapportent un salaire. Conformément à leur spécification, les professionnelles de l'agriculture et les femmes obtenant leur

principale source de revenu d'un emploi autonome sont aussi écartées de notre échantillon à cause de l'atypie de leur dynamique familiale. Puisque les employés en milieu de garde ont connu des gains salariaux allant de 38 % à 40 % entre 1999 et 2005 à cause de la nouvelle politique de garde du Québec (Baker *et al.*, 2005 p. 14), les femmes occupant ces postes sont aussi exclues de notre échantillon. En effet, la politique a eu un effet direct sur leur salaire ce qui en ferait dans leur cas une variable explicative au lieu d'un instrument de leur expérience de travail.

Tableau 3.2 Structure et taille de l'échantillon de cette étude.

	Canada		Québec		
	Personnes	Observations	Personnes	Observations	%
Panel 1	6 504	23 505	1 521	5 487	23,4
Panel 2	7 506	27 275	1 640	5 994	21,8
Panel 3	6 469	22 533	1 352	4 493	20,9
Panel 4	5 186	11 655	1 046	2 388	20,2
Total	25 665	84 968	5 559	18 362	21,7
1993	4 318		1035		24,0
1994	4 100		965		23,5
1995	3 853		894		23,2
1996	9 264		2 074		22,4
1997	8 704		1 965		22,6
1998	8 335		1 920		23,0
1999	9 425		2 020		21,4
2000	8 185		1 710		20,9
2001	7 618		1 538		20,2
2002	7 566		1 519		20,1
2003	7 114		1 423		20,0
2004	6 486		1 299		20,0
Total	84 968		18 362		21,6

Source : calcul de l'auteur à partir de son échantillon de l'EDTR de 1993 à 2004.

Pour le choix des variables, nous reprenons exactement la spécification de Semykina et Wooldridge (2006, p. 24-25). Le logarithme du salaire en dollars constants de 2002 pour chaque province constitue la variable dépendante (série 326-0021 de CANSIM). Semykina et Wooldridge (2006, p. 24) utilisent plutôt des dollars constants de 1983 en fonction du *Consumer Price Index* (CPI). Les variables explicatives comprennent l'expérience, l'expérience au carré, les années de scolarité depuis le primaire et les variables dichotomiques d'années. Ainsi, onze années équivalent à un diplôme d'étude secondaire au Québec alors qu'il en faut généralement 12 dans le reste du Canada.

Mentionnons aussi que l'expérience représente dans notre étude le nombre d'années de travail équivalentes en années complètes à temps plein comptabilisées seulement à partir du premier emploi à temps plein. La variable d'expérience *yrxfte11* dans l'EDTR ne mesure donc pas le travail à temps partiel avant le premier emploi à temps plein et ne comprend que des nombres entiers. Ainsi, une personne n'ayant jamais occupé un emploi à temps plein, mais occupant un emploi à temps partiel depuis des années obtiendrait à chaque année une valeur nulle pour la variable d'expérience *yrxfte11*. Parmi toutes les variables considérées, aucune ne comptabilisait l'expérience à temps partiel pour une personne n'ayant jamais occupé d'emploi à temps plein. Ce problème nous oblige à diverger de Semykina et Wooldridge (2006, p. 25) pour la mesure de l'expérience. Semykina et Wooldridge (2006, p. 25) construisent l'expérience à partir de la première donnée disponible additionnée à chaque année du ratio entre le nombre d'heures travaillées et 2 000 heures. Cette méthode n'a pas donné des résultats très différents pour l'EDTR. De toute façon, elle ne permet pas de solutionner le problème précédent puisque aucune variable ne divulgue le niveau initial d'expérience d'une personne n'ayant jamais occupé de travail à temps plein.

Nos instruments respectent tout à fait la spécification de Semykina et Wooldridge (2006, p. 25). Ils comprennent les variables dichotomiques d'années, les années de scolarité, l'âge et son carré, une dichotomique pour le mariage, les autres revenus familiaux de travail que ceux de la répondante et son carré, trois variables pour le nombre d'enfants âgés de 0 à 2 ans, de 3 à 5 ans et de 6 à 17 ans, l'âge du conjoint et son carré, le nombre de semaines que le conjoint a été au chômage et une dichotomique si ces semaines n'étaient pas rapportées.

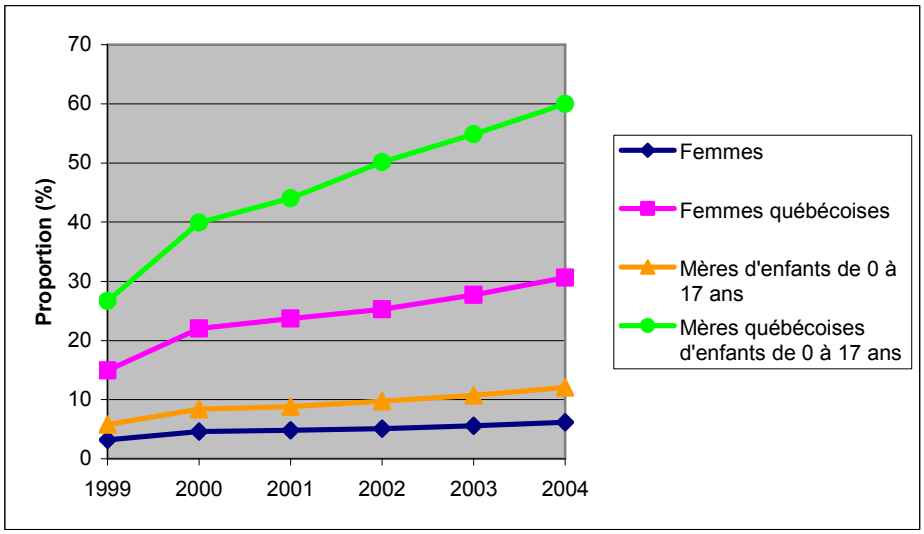
Cette première spécification servira dans la prochaine section à comparer nos résultats à ceux de Semykina et Wooldridge (2006). Par la suite, deux variables de politique s'ajouteront à ce modèle. Une variable dichotomique prend la valeur de 1 si la répondante habite au Québec. Une autre variable dichotomique indique si la femme a déjà été exposée dans le passé à la politique de services de garde du Québec à partir de 1999 et en fonction de l'âge de ses enfants. Malgré l'implantation graduelle de la politique à partir de septembre 1997 (*voir* ann. A), Lefebvre et Merrigan (2007b, p. 15) montrent en contrôlant pour les années 1999 à 2002 que ces effets croissent dans le temps. L'année 1999 sert donc de point de départ à la mesure de la politique. Les mères québécoises ayant plusieurs enfants ont une valeur de 1 dès qu'un de leurs enfants devient éligible à la politique. Les mères à l'extérieur du Québec et les Québécoises sans enfant ont évidemment la valeur de 0. Le tableau 3.3 illustre l'évolution de cette variable dans le temps pour les mères québécoises alors que la figure 3.1 montre la proportion de l'échantillon touchée par la politique selon différentes caractéristiques:

Tableau 3.3 Répartition de la variable dichotomique de politique de services de garde au Québec selon l'âge des enfants et l'année d'exposition.

1997	1998	1999	2000	2001	2002	2003	2004
3	4	5	6	7	8	9	10
2	3	4 (1)	5 (1)	6 (1)	7 (1)	8 (1)	9 (1)
1	2	3 (1)	4 (1)	5 (1)	6 (1)	7 (1)	8 (1)
0	1	2 (1)	3 (1)	4 (1)	5 (1)	6 (1)	7 (1)
	0	1	2 (1)	3 (1)	4 (1)	5 (1)	6 (1)
		0	1 (1)	2 (1)	3 (1)	4 (1)	5 (1)
			0 (1)	1 (1)	2 (1)	3 (1)	4 (1)
				0 (1)	1 (1)	2 (1)	3 (1)
					0 (1)	1 (1)	2 (1)
						0 (1)	1 (1)
							0 (1)

- L'âge de l'enfant se trouve à gauche.
- (1) signifie que la variable dichotomique prend la valeur de 1. Autrement, elle affiche 0.
- Les zones grises indiquent que l'enfant est éligible à la politique de services de garde à ce moment.

Figure 3.1 Proportion des femmes de l'échantillon ayant déjà été exposées à la politique de services de garde à partir de 1999.



Par ailleurs, le probit pour le calcul des termes de correction estimés $\hat{\lambda}_{it2}$ a recours à ces mêmes instruments comme variables indépendantes. Toutefois, au lieu d'estimer la probabilité d'observer un salaire, il le fait plutôt pour la probabilité de participer au marché du travail. En effet, à cause de la définition de la variable d'expérience *yrxfte11*, plusieurs répondantes obtenaient toujours une valeur nulle, mais rapportaient un salaire positif ce qui provoquait des rendements d'expérience non crédibles par la suite. Ainsi, il a fallu éliminer les femmes n'ayant aucune expérience de travail selon la variable *yrxfte11* pour le calcul du rendement de l'expérience. La variable dépendante dichotomique du probit de l'équation de sélection prend donc la valeur 1 si la répondante rapporte un salaire et si son expérience n'est pas nulle. Cette règle diffère de Semykina et Wooldridge (2006) dont la variable dichotomique de l'équation de sélection prend la valeur de 1 uniquement si le salaire est observé ce qui implique dans le PSID que la femme travaille. Aussi, ils ne conservent que les femmes ayant travaillé pour au moins deux années dans la période 1980-1992. Les tableaux D.1 et D.2 résument notre spécification et les variables sélectionnées (*voir* ann. D).

Pour faire partie de l'équation de sélection, seuls les instruments doivent être observés. Parmi les 84 968 observations de départ pour le calcul du ratio inverse de Mill $\left(\hat{\lambda}_{it2}\right)$, 34,66 % ne rapportent pas de salaire et/ou n'ont pas d'expérience de travail au sens de la variable *yrxfte11*. Ce taux est de 35,42 % pour Semykina et Wooldridge (2006) (*voir* tableau 3.4). Par la suite, seules les observations rapportant un salaire et une expérience de travail non nulle sont conservées. Les régressions MCO utilisent ce deuxième échantillon de 55 520 observations. Pour finir, les régressions par effets fixes nécessitent au moins deux observations par individus ce qui diminue les observations considérées à 50 642 (*voir* tableau 3.8). Le programme de Mme Semykina se conforme tout aussi rigoureusement à cette procédure qui repose sur trois échantillons afin de diminuer tout biais potentiel.

En somme, la procédure #1 applique l'équation (18) à notre échantillon et équivaut à un modèle EF-MCDE incluant les termes de correction estimés du biais de sélection $\hat{\lambda}_{it2}$ comme variables explicatives et comme instruments. Si le test statistique sur les coefficients des termes de correction $\hat{\lambda}_{it2}$ se révèle significatif, alors la procédure #2 s'impose, car il y a présence d'un biais de sélection contemporain. La procédure #2 reprend l'équation (47) et revient à une procédure MCDE à laquelle on ajoute non seulement les termes de correction $\hat{\lambda}_{it2}$, mais aussi les moyennes inter temporelles des instruments pour chaque individu ($\bar{z}_i$) comme variables explicatives et comme instruments. Cette procédure nous permet de calculer le rendement de l'expérience en contrôlant pour le biais de sélection, l'hétérogénéité et l'endogénéité. Toutefois, il faut avoir recours à une hypothèse non vérifiable, celle de la linéarité de l'effet fixe en fonction des moyennes inter temporelles des instruments (34). Les paramètres de la première étape des procédures par variables instrumentales révèlent l'impact des instruments sur l'expérience et son carré. L'effet attribuable à la politique de services de garde du Québec par l'intermédiaire de l'expérience peut donc être testé. Ensuite, les tests sur les instruments dévoileront la présence d'instruments faibles ou invalides et un biais possible du test joint sur les termes de correction $\hat{\lambda}_{it2}$. Aussi, les coefficients des moyennes des instruments qui apparaissent dans l'équation structurelle de la procédure #2 offrent une idée générale du rôle joué par l'effet fixe. Finalement, les paramètres de la forme réduite des procédures #3 (48) et #4 (51) dévoileront l'effet direct des instruments sur le salaire dont la contribution de la politique de services de garde du Québec.

3.2 Analyse descriptive

Le tableau 3.4 compare clairement les statistiques descriptives présentées par Semykina et Wooldridge (2006, p. 34) avec celles de notre échantillon. La moyenne du logarithme du salaire horaire semble plus élevée pour l'échantillon canadien, mais Semykina et Wooldridge (2006) utilisent des dollars de 1983 alors que les salaires de cette étude sont en dollars de 2002. Soulignons que Semykina et Wooldrige (2006) ne mentionnent pas avoir exclu les centiles inférieurs et supérieurs de la distribution des salaires et des revenus. Il a fallu appliquer cette condition à notre échantillon afin de rejeter les salaires inférieurs au salaire minimum et les revenus trop négatifs. Cette différence explique peut-être les écarts types moins élevés de notre échantillon. L'écart type empirique de l'expérience de notre échantillon est plus élevé même en incluant les valeurs nulles des femmes ne participant pas au marché du travail. Les années d'éducation des deux échantillons sont comparables quoique la différence soit statistiquement significative à 1 %. Par contre, la moyenne d'âge des femmes ne participant pas au marché du travail est statistiquement plus élevée pour les deux échantillons. Pour un pouvoir d'achat constant, les autres revenus familiaux des femmes canadiennes de notre échantillon sont inférieurs à ceux des Américaines de l'autre groupe, surtout pour les femmes exclues du marché du travail. Par ailleurs, les femmes non-participantes de l'échantillon américain rapportent des revenus familiaux autres supérieurs à la moyenne totale. Le contraire est vrai pour l'échantillon canadien, c'est-à-dire que les femmes non-participantes de notre échantillon disposent de revenus familiaux autres inférieurs à la moyenne totale. Les conjoints canadiens sont aussi beaucoup plus âgés, mais disposent d'une éducation équivalente. Notre échantillon compte davantage de semaines de chômage déclarées par les conjoints. La participation au marché du travail n'influence pas autant la moyenne du nombre d'enfants de 3 à 5 ans dans l'échantillon canadien quoique la définition différente de la participation au marché du travail joue peut-être un rôle ici. La moyenne du nombre d'enfants de 6 à 17 ans de notre échantillon est statistiquement

inférieure à l'échantillon américain spécialement pour les femmes non-participantes au marché du travail. Dans l'ensemble, les deux échantillons se comparent aisément sauf pour quelques exceptions.

Tableau 3.4 Comparaison de statistiques descriptives pour l'échantillon de cette étude et de Semykina et Wooldridge (2006, p. 34).

Variable	Semykina et Wooldridge			Échantillon de cette étude		
	Total	Part.	Non-part.	Total	Part.	Non-part.
Participation	0,73	1	0	0,65	1	0
Log du salaire réel horaire	—	1,94 (0,62)	—	—	2,66 (0,45)	—
Expérience (années)	11,76 (7,76)	12,93 (7,58)	8,51 (7,31)	12,62 (9,92)	14,05 (8,96)	9,93 (11,02)
Éducation (années)	12,94 (2,27)	13,13 (2,24)	12,40 (2,29)	13,09 (3,26)	13,74 (3,05)	11.88 (3,30)
Âge (années)	40,91 (10,28)	40,12 (9,61)	43,12 (11,65)	41,54 (11,62)	39,18 (9,85)	46,00 (13,29)
Mariée	0,86	0,84	0,93	0,85	0,83	0,88
Autre revenu familial ($)	34 461 (40 586)	31 167 (30 996)	44 268 (58 520)	34 880 (26 523)	38 901 (24 870)	27 299 (27 851)
Âge du conjoint (années)	37,07 (18,05)	35,21 (18,14)	42,21 (16,75)	44,15 (12,42)	41,58 (10,56)	49,01 (14,10)
Éducation du conjoint (années)	11,26 (5,22)	11,04 (5,47)	11,88 (4,38)	12,83 (3,70)	13,33 (3,47)	11,88 (3,92)
Semaines de chômage du conjoint	0,98 (4,96)	0,95 (4,79)	1,06 (5,39)	2,55 (8,88)	2,32 (8,09)	2,99 (10,20)
Sem. de chôm. non rapportées	0,08	0,06	0,16	0,025	0,005	0,050
Enfants de 0 à 2 ans	0,14 (0,37)	0,11 (0,33)	0,21 (0,45)	0,16 (0,41)	0,16 (0,41)	0,15 (0,42)
Enfants de 3 à 5 ans	0,18 (0,42)	0,16 (0,40)	0,24 (0,49)	0,17 (0,42)	0,17 (0,42)	0,17 (0,43)
Enfants de 6 à 17 ans	0,82 (1,01)	0,84 (1,01)	0,77 (0,99)	0,71 (0,98)	0,77 (0,98)	0,60 (0,99)
Observations	11 232	8 254	3 978	84 968	55 520	29 448

▪ Les écarts types empiriques sont rapportés entre parenthèses.
▪ Le logarithme du salaire horaire réel est calculé à partir de données en dollars de 1983 pour Semykina et Wooldridge (2006) et en dollars de 2002 pour notre étude.

Les figures 3.2 à 3.14 ci-après montrent l'évolution dans le temps des variables d'intérêt de l'étude pour notre échantillon. Elles comparent aussi l'évolution temporelle de ces variables entre les mères d'au moins un enfant de 0 à 17 ans et les autres femmes. Mentionnons avant tout la présence de bruit statistique qui crée des sauts lors de l'introduction du panel 3 en 1999 et du panel 4 en 2002, un phénomène mentionné aussi dans d'autres études (Lefebre et Merrigan, 2007b, p. 19).

D'abord, le taux de participation des femmes québécoises de notre échantillon a crû d'environ 7 % en 12 ans alors que le reste du Canada a connu une légère hausse et que l'Ontario est resté à un niveau comparable (*voir* figures 3.2 et 3.3). L'écart entre l'Ontario et le Québec diminue de moitié à ce chapitre pour s'établir à 6,4 % en 2004. Étonnamment, le taux de participation des mères de notre échantillon a connu l'ascension la plus fulgurante surtout au Québec où il a augmenté de 10 % en 12 ans (*voir* figures 3.4 et 3.5). Au contraire, les femmes sans enfant du reste du Canada ont vu leur taux de participation rester stable à 60 % alors que celles du Québec les ont légèrement rattrapé de quelques points de pourcentages en 12 ans.

Figure 3.2 Taux de participation de toutes les femmes de l'échantillon au mois d'avril au Québec, en Ontario et dans les autres provinces (RdC) (variable *ml04v2* de l'EDTR).

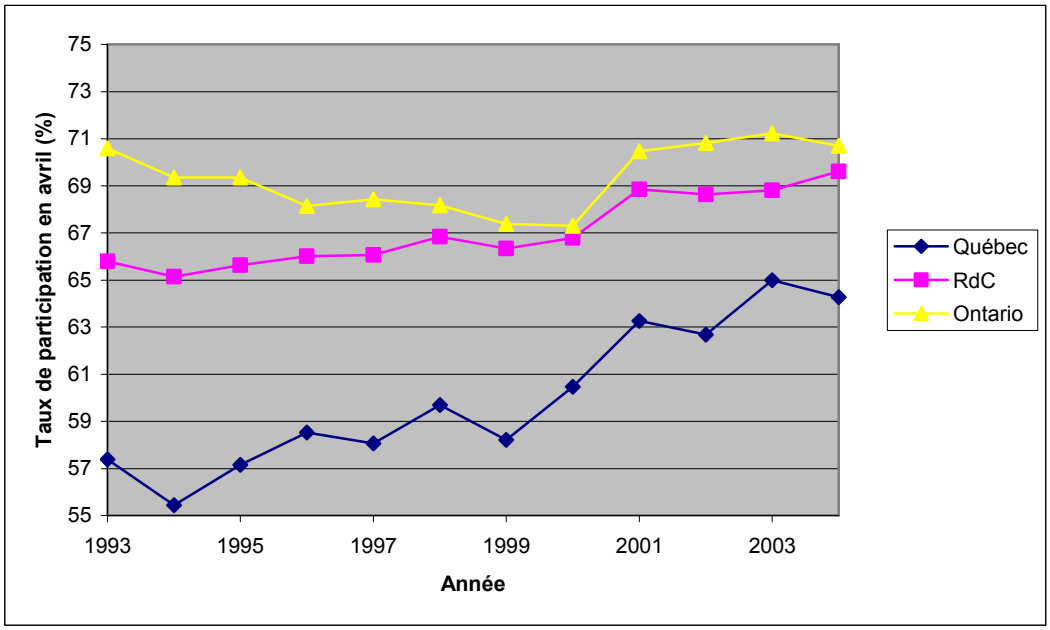


Figure 3.3 Taux de participation de toutes les femmes de l'échantillon au mois d'août au Québec, en Ontario et dans les autres provinces (RdC) (variable *ml08v2* de l'EDTR).

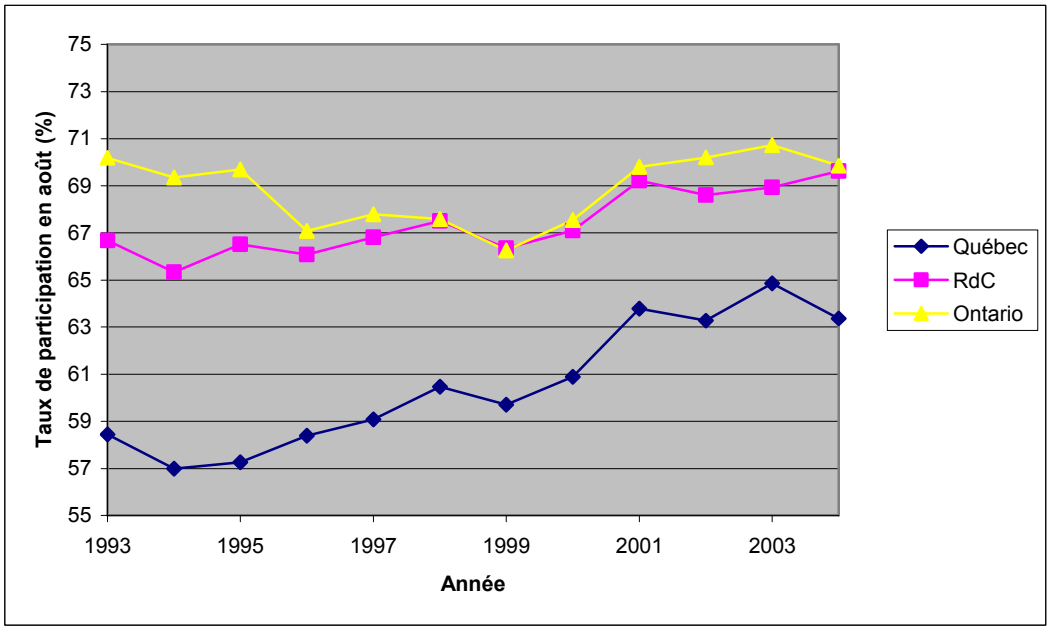


Figure 3.4 Taux de participation des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au mois d'avril au Québec (Qc) et dans les autres provinces (RdC) (variables *ml04v28* de l'EDTR).

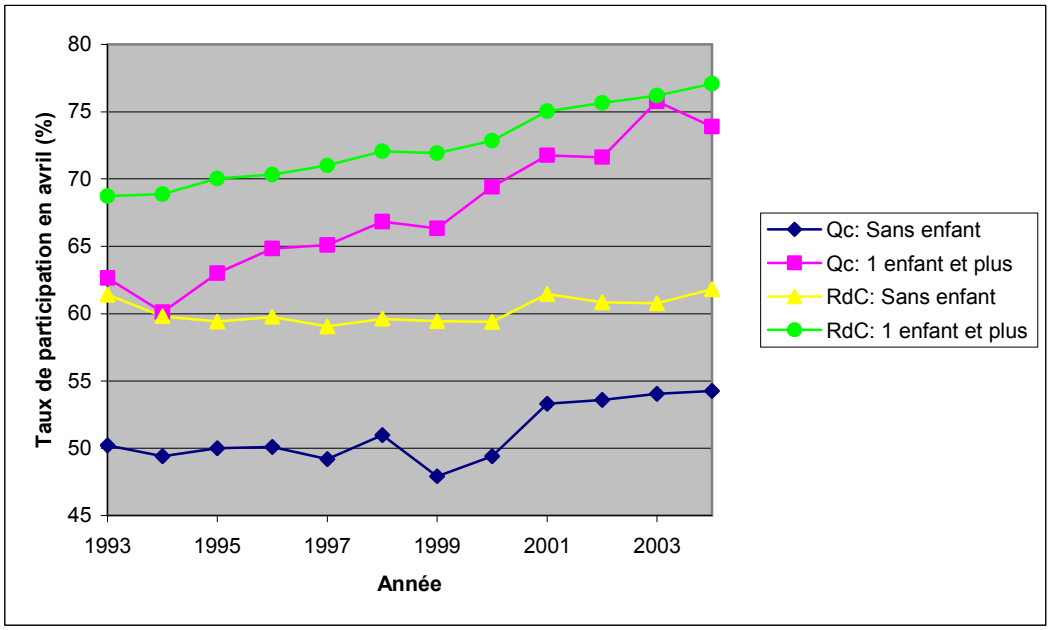
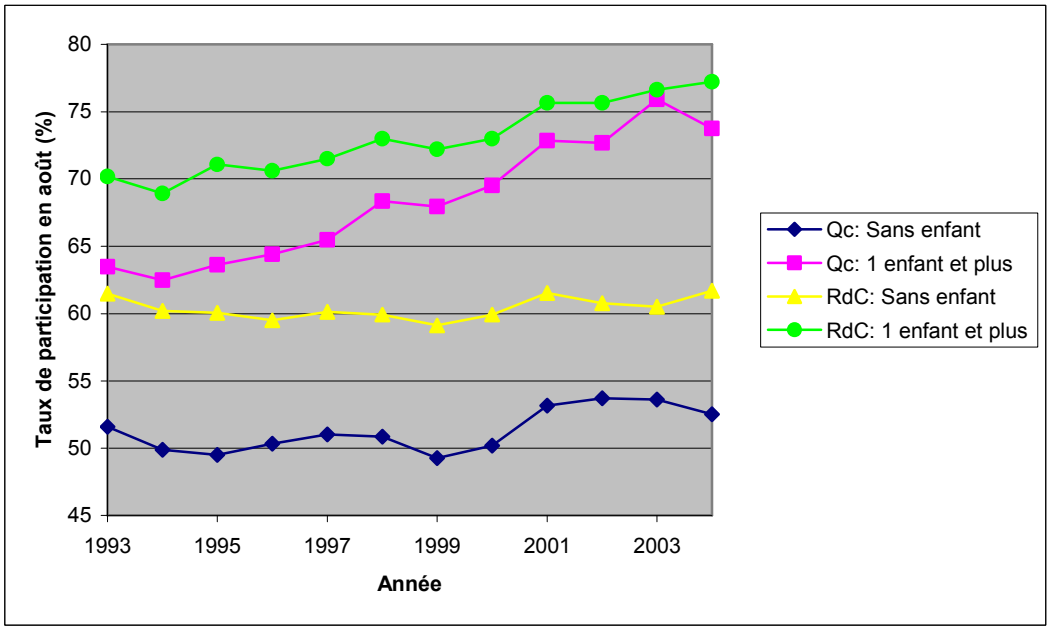


Figure 3.5 Taux de participation des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au mois d'août au Québec (Qc) et dans les autres provinces (RdC) (variables *ml08v28* de l'EDTR).



Ensuite, à l'égard des variables d'offre de travail, le Québec a encore réduit son retard sur le reste du Canada au niveau des heures et des semaines annuelles travaillées (*voir* figures 3.6 à 3.9). Ainsi, l'écart entre l'Ontario et le Québec pour les heures travaillées annuellement est passé de 245 heures en 1993 à 125 heures en 2004. Pour les semaines de travail, la différence entre les deux provinces était de 7 semaines en 1993 et seulement de 3 semaines en 2004. Pourtant, cette dynamique provient surtout des mères de notre échantillon et particulièrement celles du Québec qui ont augmenté leur offre de travail de 278 heures en 12 ans. Celles du reste du Canada l'ont fait de 224 heures si bien que la différence entre les deux groupes était de 93 heures en 1993 et de 39 heures en 2004. Sans égard à la maternité, les Québécoises sans enfant ont quand même réduit l'écart de leurs heures travaillées par rapport à celles du reste du Canada de 241 heures en 1993 à 196 heures en 2004. La même dynamique se reflète sur les semaines travaillées. Ainsi, l'augmentation de l'offre de travail des femmes de notre échantillon provient davantage des mères canadiennes avec un effet plus marqué pour l'ensemble du Québec. Toutes les différences de moyenne précédentes sont statistiquement significatives à 1 %.

Figure 3.6 Heures annuelles moyennes travaillées par toutes les femmes de l'échantillon au Québec, en Ontario et dans les autres provinces (RdC) (variable *alhrwk28* de l'EDTR).

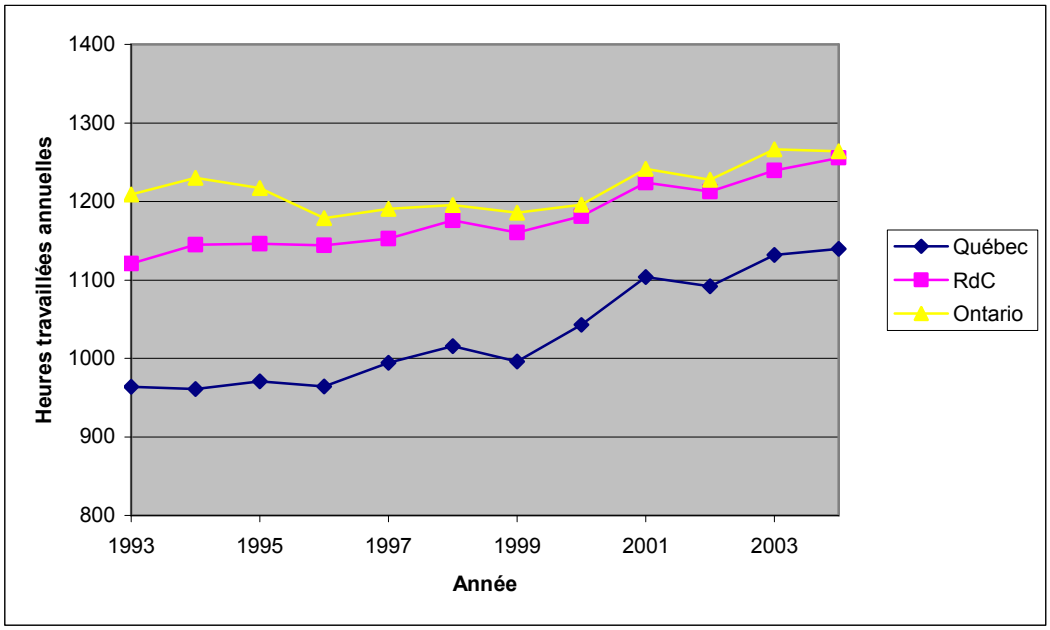


Figure 3.7 Heures annuelles moyennes travaillées par les mères d'au moins 1 enfant de 0 à 17 ans et par les femmes sans enfant de l'échantillon au Québec (Qc) et dans les autres provinces (RdC) (variable *alhrwk28* de l'EDTR).

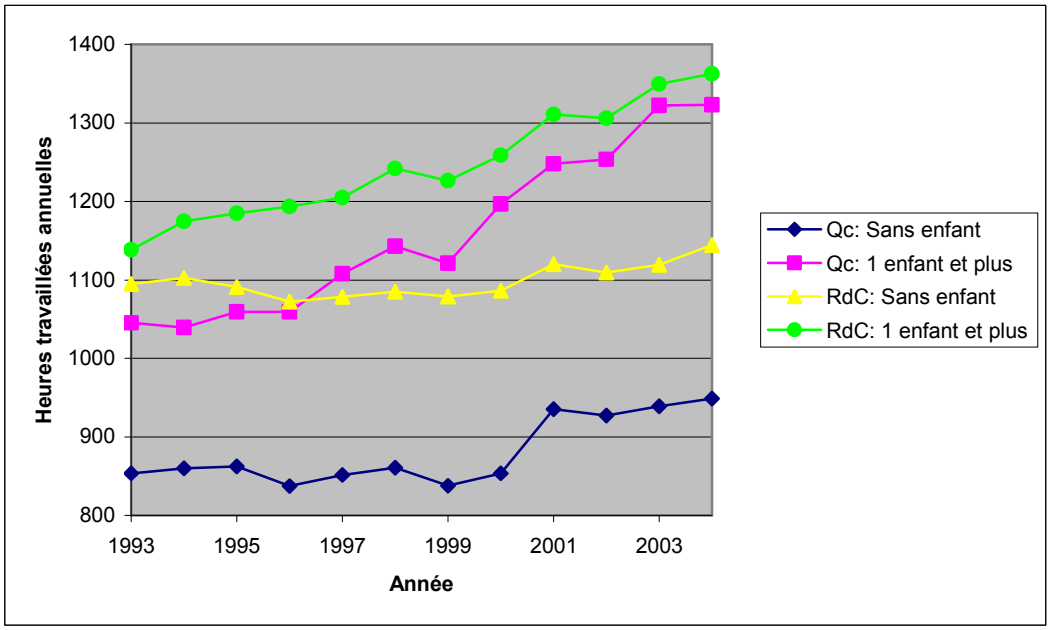


Figure 3.8 Semaines annuelles moyennes travaillées par toutes les femmes au Québec, en Ontario et dans les autres provinces (RdC) (variable *wksem28* de l'EDTR).

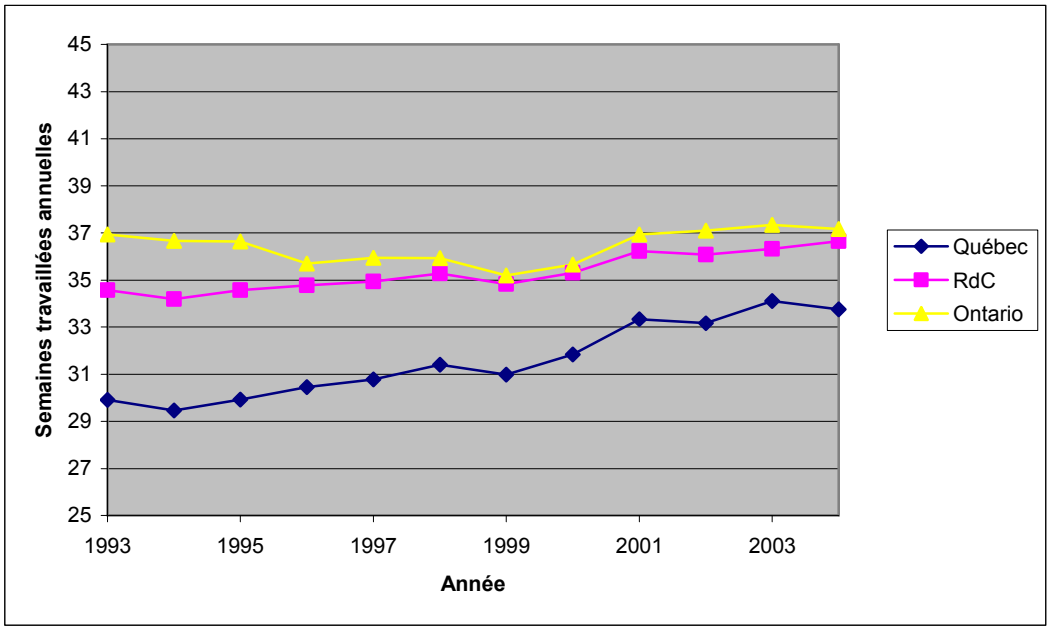


Figure 3.9 Semaines annuelles moyennes travaillées par les mères d'au moins 1 enfant de 0 à 17 ans et par les femmes sans enfant de l'échantillon au Québec (Qc), en Ontario et dans les autres provinces (RdC) (variable *wksem28* de l'EDTR).

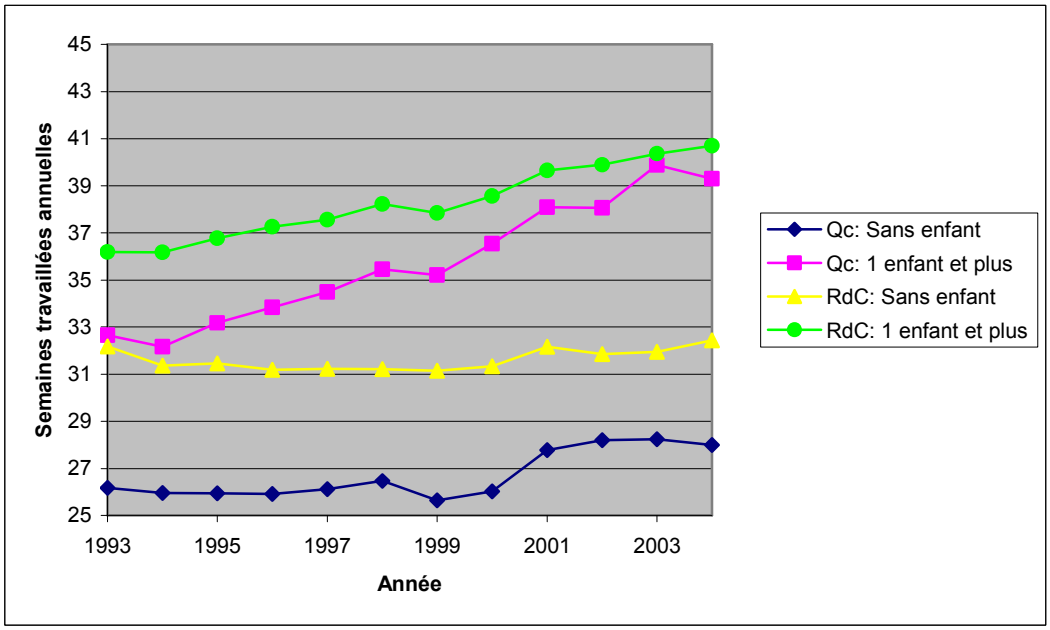


Figure 3.10 Revenu de travail annuel moyen de toutes les femmes au Québec, en Ontario et dans les autres provinces (RdC) (variable *earng42* de l'EDTR corrigée pour l'inflation).

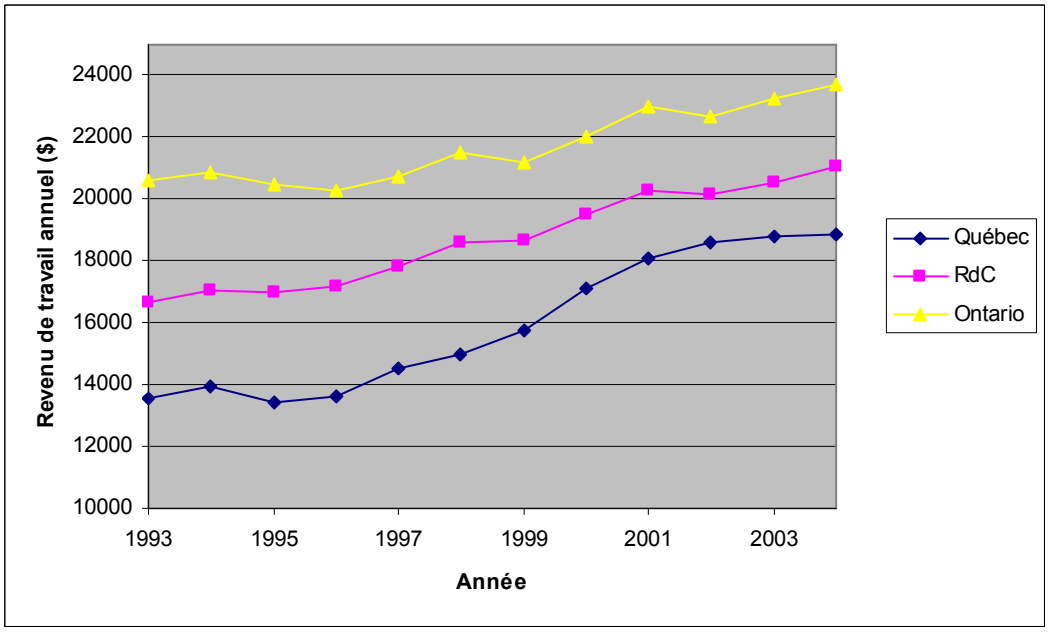
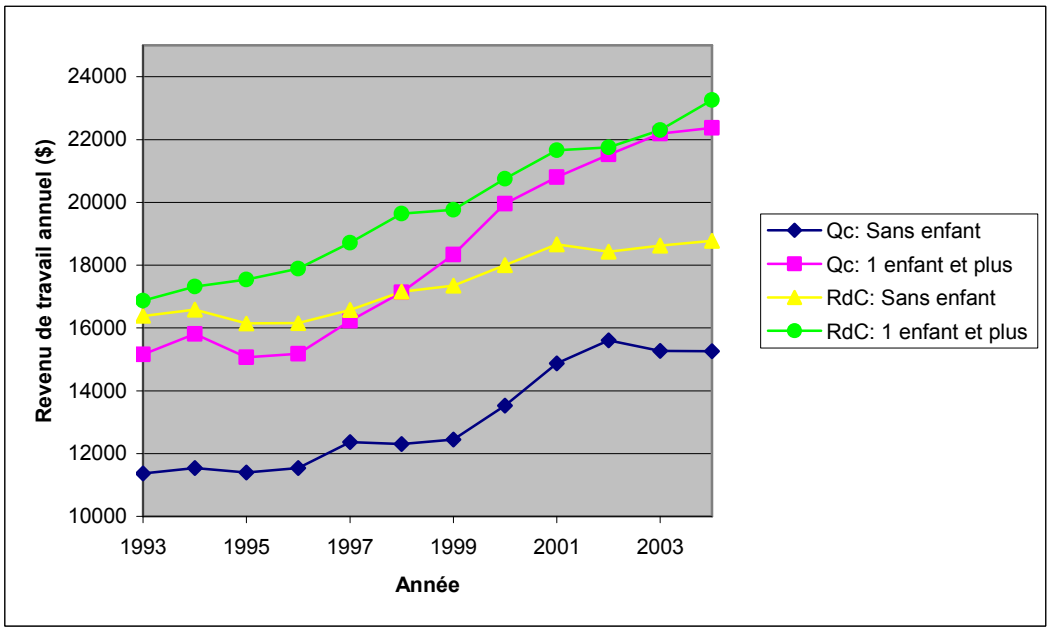


Figure 3.11 Revenu de travail annuel moyen des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au Québec (Qc) et dans les autres provinces (RdC) (variable *earng42* de l'EDTR corrigée pour l'inflation).



Pourtant, le revenu de travail annuel reste inférieur au Québec même par rapport au reste du Canada (*voir* figure 3.10). À l'instar des variables d'offre de travail, le revenu de travail annuel des mères du reste du Canada augmente progressivement au cours de la période considérée. Pour le Québec, l'augmentation débute en 1997 et l'écart avec les autres mères canadiennes est presque éliminé au cours des 8 années suivantes.

Enfin, l'évolution longitudinale des deux variables les plus importantes laisse cependant perplexe (*voir* figures 3.12 à 3.15). D'une part, la moyenne d'expérience des femmes de l'échantillon augmente d'environ 3 ans et demi pour l'ensemble du Canada alors que les autres provinces réduisent légèrement leur retard sur l'Ontario. Néanmoins, les femmes sans enfant ont davantage augmenté leur expérience par rapport aux mères au sens de la variable *yrxfte11* de l'EDTR qui ne comptabilise pas le travail à temps partiel avant le premier emploi à temps plein. Pourtant, les mères canadiennes auraient dû accumuler plus d'expérience selon l'évolution des heures et des semaines annuelles travaillées (*voir* figures 3.9 et 3.11). Une prépondérance des mères dans les postes à temps partiel pourrait peut-être expliquer cette tendance surprenante. Pour le Québec, la moyenne d'expérience des femmes sans enfant a augmenté de 5 années en 12 ans alors que l'expérience des mères québécoises n'a progressé que de 2,2 années. D'autre part, le salaire des provinces progresse d'une manière semblable avec un léger rattrapage des autres provinces sur l'Ontario. Il n'y a pas de différence statistiquement significative entre la progression des salaires moyens des mères et des femmes sans enfant au Québec et dans le reste du Canada. Étonnamment, les mères d'enfants de 0 à 17 ans de notre échantillon gagnent un salaire plus élevé que les autres femmes en moyenne. Ce résultat est statistiquement significatif à 5 % en 2004, mais ce n'est pas le cas pour quelques années. En contrôlant pour diverses variables explicatives, la maternité devrait donner lieu à une

pénalité sur l'expérience et donc sur le salaire. Une analyse économétrique plus rigoureuse permettra de donner une appréciation plus juste de ces remarques.

Figure 3.12 Expérience moyenne en années des femmes de l'échantillon au Québec, en Ontario et dans les autres provinces (RdC) (variable *yrxfte11* de l'EDTR).

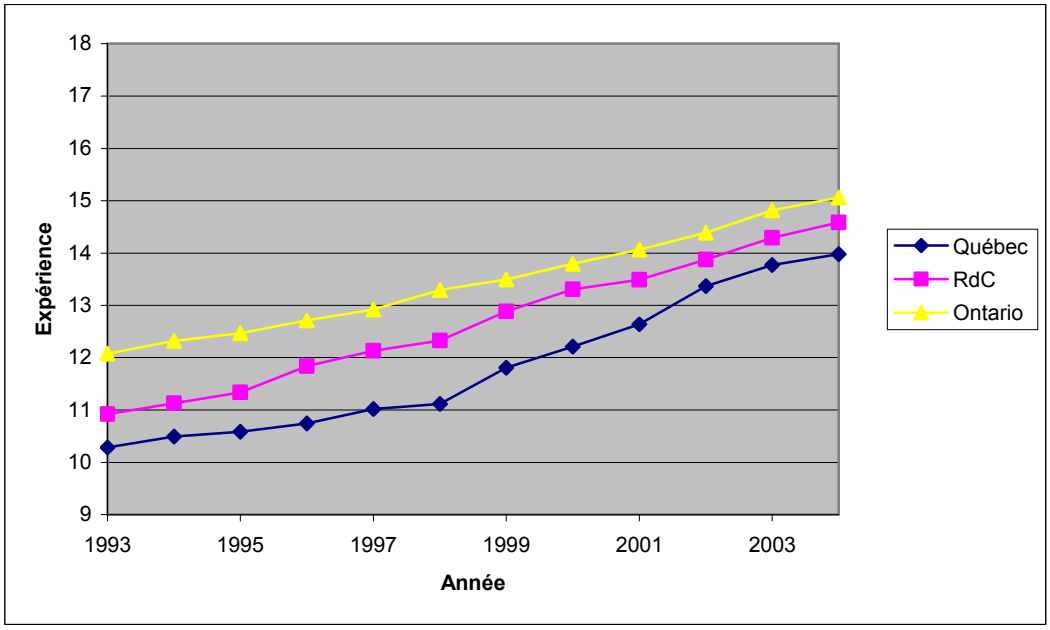


Figure 3.13 Expérience moyenne en années des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au Québec (Qc) et dans les autres provinces (RdC) (variable *yrxfte11* de l'EDTR).

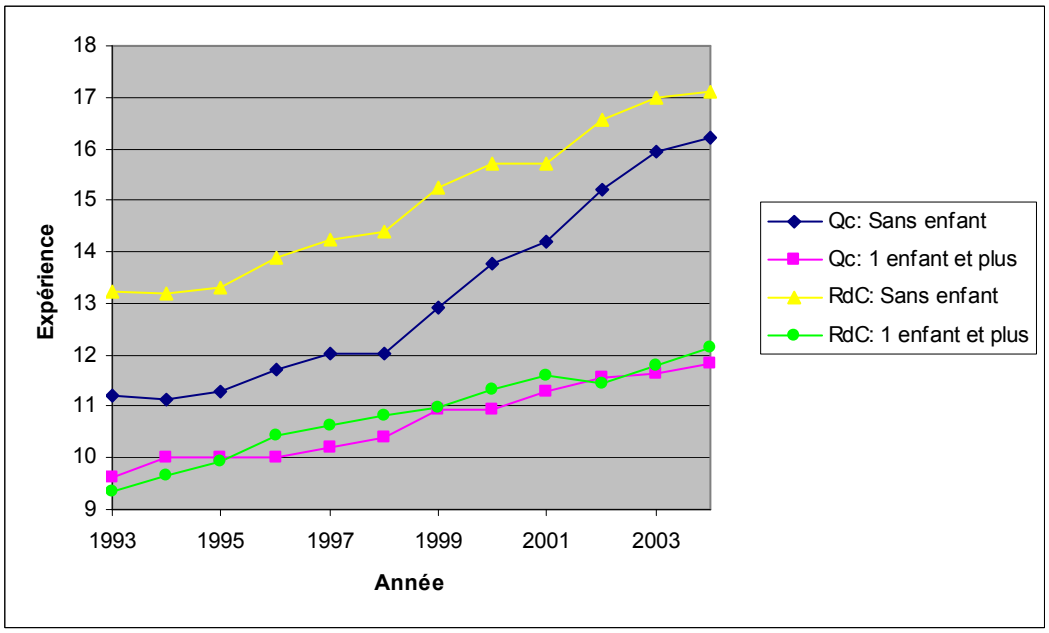


Figure 3.14 Salaire horaire moyen des femmes de l'échantillon au Québec, en Ontario et dans les autres provinces (RdC) (variable *cmphrw28* de l'EDTR corrigée pour l'inflation).

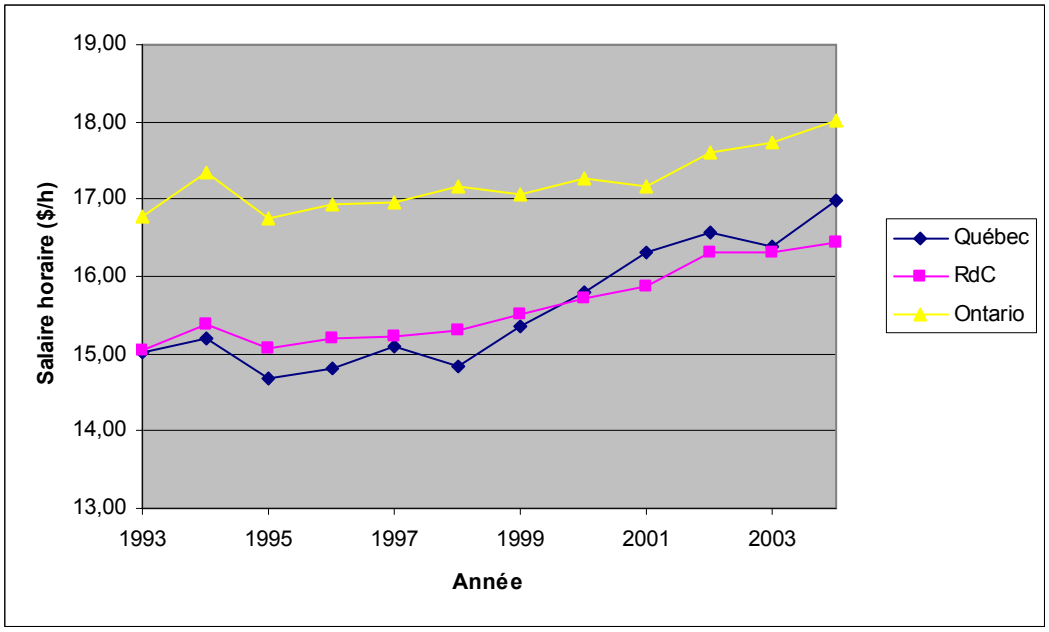
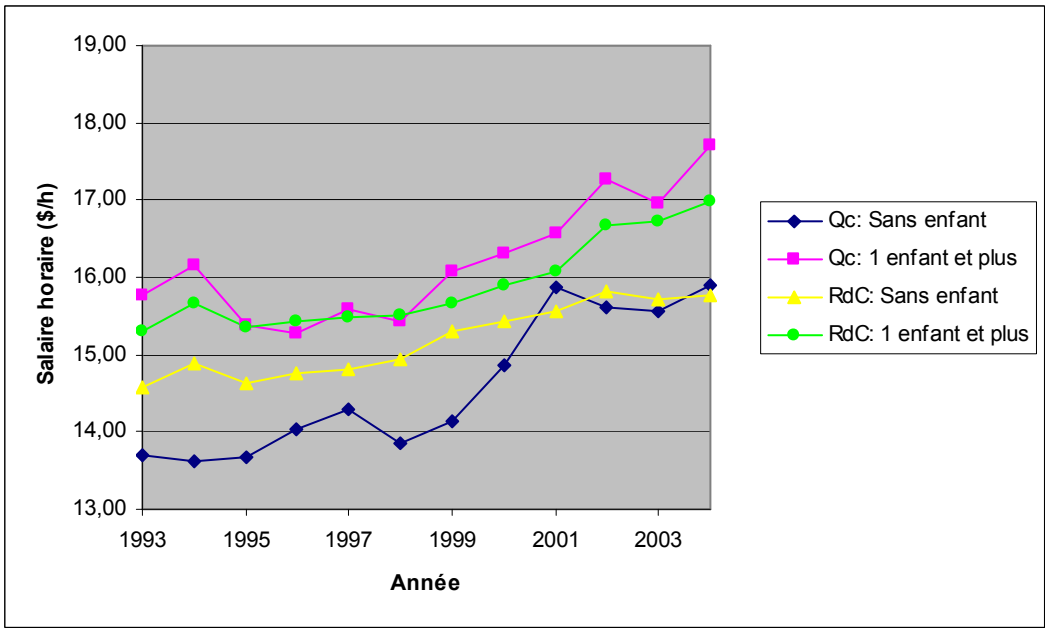


Figure 3.15 Salaire horaire moyen des mères d'au moins 1 enfant de 0 à 17 ans et des femmes sans enfant de l'échantillon au Québec (Qc) et dans les autres provinces (RdC) (variable *cmphrw28* de l'EDTR corrigée pour l'inflation).



3.3 Résultats

Les tableaux 3.5 et 3.6 ci-dessous comparent nos résultats avec ceux fournis par Semykina et Wooldridge (2006, p. 35). Ces derniers obtiennent des rendements allant de 5,5 % à 8,2 % selon la méthode choisie. Tous ces rendements sont significatifs à 1 % à l'exception des procédures #1 et #2 qui ne sont significatives qu'à 5 %. La correction pour l'endogénéité de l'expérience de la méthode MCDE diminue quelque peu le rendement de l'expérience de la méthode MCO de 6,7 % à 6,2 %. Toutefois, le test de Hausman ne peut rejeter l'hypothèse nulle que l'estimateur MCO converge (*voir* tableau 3.9). L'endogénéité ne biaiserait donc pas l'estimateur MCO. L'effet fixe semble jouer un rôle plus important puisque la méthode EF fait passer le rendement de l'expérience de 6,7 % à 8,2 %. Elle affecte aussi le rendement de l'éducation qui passe de 11,4 % à 2,3 % pour la méthode EF. La méthode EF-MCDE ramène le rendement de l'expérience à celui de la méthode MCO. La procédure #1 arrive à des résultats très similaires alors que les termes de correction $\hat{\lambda}_{it2}$ sont significatifs à 5 %. Ainsi, le biais de sélection significatif justifie l'utilisation de la procédure #2 qui arrive à un rendement de l'expérience de 5,5 % significatif à 5 %. Le carré de l'expérience de même que les termes de correction $\hat{\lambda}_{it2}$ ne sont plus significatifs à ce stade.

Le tableau 3.6 présente les résultats calculés à partir de notre échantillon et d'une spécification aussi près que possible de celle de Semykina et Wooldridge (2006). Les rendements de l'expérience vont de 3,2 % à 6,1 %. Le rendement de la méthode MCO est beaucoup plus bas que le résultat de Semykina et Wooldridge (2006) à 3,2 %. Contrairement au cas précédent, la correction pour l'endogénéité de la méthode MCDE a un impact plus important à 5,8 % que la correction pour l'hétérogénéité de la méthode EF à seulement 3,8 %. Tel que précédemment, la méthode EF-MCDE arrive à un rendement proche de la méthode MCO. Toutefois, les rendements de la

méthode EF-MCDE et de la procédure #1 ne sont plus significatifs. Cette difficulté laisse présager un problème au niveau des instruments qui affecte probablement la validité des termes de correction $\hat{\lambda}_{it2}$ non significatifs. Par conséquent, la procédure #2 s'avère la meilleure méthode pour surmonter tous ces problèmes. Le rendement de l'expérience de la procédure #2 de 6,1 % n'est d'ailleurs pas statistiquement différents du résultat final de Semykina et Wooldridge (2006) de 5,5 %. Concernant les autres variables explicatives, la non-linéarité du rendement de l'expérience persiste peu importe la méthode. Le rendement de l'éducation progresse aussi de la même manière à travers les méthodes. Le contrôle pour l'hétérogénéité diminue de plus du tiers le rendement de l'éducation auquel s'ajoute un contrôle pour l'endogénéité. Le rendement de l'éducation de la procédure #2 est presque identique pour les deux échantillons.

Afin d'examiner davantage la perte de significativité de la méthode EF-MCDE et de la procédure #1 du rendement de l'expérience pour notre échantillon, divers tests mesurent la faiblesse et la validité des instruments proposés par Semykina et Wooldridge (2006, p. 25). Les résultats présentés au tableau 3.7 révèlent que seule la méthode MCDE ne dispose pas d'instruments faibles. Les statistiques du R^2 partiel et du R^2 de Shea partiel démontrent clairement la faiblesse des régresseurs exogènes exclus de la variable d'expérience pour la méthode EF-MCDE et pour les procédures #1 et #2. En acceptant un biais relatif de 10 % par rapport à l'estimateur MCO, le test de Stock et Yogo (2005) rejette l'hypothèse nulle que les instruments sont faibles dans tous les cas. Toutefois, cette conclusion ne corrige pas pour l'hétéroscédasticité très probable des erreurs des modèles. De plus, le test des restrictions de suridentification robuste à l'hétéroscédasticité de Wooldridge (1995a) ne peut rejeter l'hypothèse nulle que les instruments sont valides uniquement pour la procédure #2.

Ensuite, une variable dichotomique pour la province de Québec et une autre pour la politique de services de garde s'ajoutent aux instruments de la spécification précédente afin d'évaluer l'impact possible de la politique sur le salaire par l'intermédiaire de l'expérience de travail. Ainsi, les résultats du tableau 3.8 s'approchent considérablement de ceux du tableau 3.6 puisque seulement deux instruments ont été ajoutés. La méthode MCDE affiche donc les mêmes rendements pour l'expérience et pour l'éducation. La méthode EA donne un rendement identique à la méthode MCO à 3,2 %. Le test de Hausman (*voir* équation 32) compare les méthodes EF et EA et donne une statistique de test de $\chi_1^2 = 3{,}39$ et une p-value de 0,0656. Le test rejette donc l'hypothèse nulle à un niveau de significativité de 10 % ce qui constitue une indication que l'estimateur EA est biaisé et que l'estimateur EF est convergent. Le même test peut comparer les estimateur MCDE et MCO ce qui donne une statistique de test de $\chi_1^2 = 259{,}00$ et une p-value proche de 0,0000. L'estimateur MCO est donc probablement biaisé et l'estimateur MCDE, quoique moins efficient, est convergent. Ces deux derniers tests confirment la nécessité d'une correction pour l'endogénéité et l'hétérogénéité. Cependant, les rendements de l'expérience de la méthode EF-MCDE et de la procédure #1 diminuent respectivement à 2,5 % et 1,8 % tout en perdant leur significativité comme pour la spécification précédente. Ce problème donne une statistique de Hausman des estimateurs EF-MCDE et EA-VI de $\chi_1^2 = 0{,}52$ et une p-value de 0,4709. Évidemment, des instruments faibles biaisent probalement l'estimateur EF-MCDE. La procédure #1 pose le même problème et le test du biais de sélection contemporain n'est pas significatif avec une p-value de 0,7872. Dans ce cas, il faut se rabattre sur la procédure #2 qui estime le rendement de l'expérience à 6,4 % et celui de l'éducation à 3,8 %, deux valeurs tout à fait comparables aux résultats du tableau 3.6 et aux résultats de Semykina et Wooldridge (2006). Les termes de correction estimés $\hat{\lambda}_{it2}$ redeviennent alors significatifs à un niveau de 2,51 %. Aucun des tests du biais

de sélection inter temporel n'ont été concluants. Nous omettons donc ces résultats par souci de concision.

Par ailleurs, les coefficients des variables explicatives et instrumentales exprimées en moyennes individuelles de la procédure #2 ne peuvent être analysés simplement. En effet, l'ensemble de ces coefficients est censé représenter l'effet fixe selon l'hypothèse de Semykina et Wooldridge posée par l'équation (34). Les variables exprimées en moyenne sont généralement significatives à l'exception du carré des autres revenus familiaux, de l'âge du conjoint et de son carré, de l'indicateur de semaines de chômage non rapportées et des deux variables de politique.

En outre, les résultats des régressions de première étape sur l'expérience et son carré révèlent les instruments les plus significatifs et les méthodes les plus stables (*voir* tableaux 3.10 à 3.12). De façon générale, presque tous les instruments sont significatifs pour les méthodes MCO et EA-VI. Par contre, beaucoup d'instruments dont ceux sur le conjoint deviennent non significatifs dans les premières étapes de l'estimateur EF-MCDE et dans les procédures #1 et #2. Sur les 17 moyennes individuelles utilisées comme instruments dans la procédure #2, seules 8 d'entre elles sont significatives à 10 %. De plus, le signe des effets des instruments sur l'expérience est identique entre les méthodes MCDE et EA-VI. Par contre, certains coefficients changent de signe de la méthode MCDE à EF-MCDE ou à la procédure #1. Ce phénomène se produit surtout pour les coefficients non significatifs, mais aussi pour l'âge qui est significatif à 1 %. Les signes des coefficients des instruments pour la méthode EF-MCDE et pour la procédure #1 contredisent parfois l'intuition courante. Ainsi, l'âge en déviation par rapport à la moyenne a un effet négatif sur l'expérience significatif à 1 % pour la méthode EF-MCDE et pour la procédure #1. De même, le nombre d'enfants de 0 à 2 ans en déviation par rapport à la moyenne augmente l'expérience de 0,036 année à un niveau de significativité de 1 % pour la

méthode EF-MCDE. Enfin, les coefficients des instruments de la procédure #2 obtiennent le même signe que la méthode EF-MCDE seulement 9 fois sur 17 instruments communs. Pour le nombre d'enfants de 3 à 5 ans et de 6 à 17 ans, les coefficients n'ont pas le même signe, mais sont tous les deux significatifs à 1 %.

D'ailleurs, l'intuition de base de ce travail reposait sur l'hypothèse que la politique de garde du Québec était un bon instrument, c'est-à-dire non corrélé avec les termes d'erreurs et expliquant bien l'expérience. La variable dichotomique du Québec et celle de la politique de services de garde performent mal dans les premières étapes des méthodes MCDE et EF-MCDE puisqu'elles ne sont pas significatives. Pourtant, la dichotomique du Québec augmente l'expérience de 0,288 années à un niveau de significativité de 1 % pour l'estimateur EA-VI. Dans la procédure #1, la variable de la politique de services de garde accroît l'expérience de 0,112 années à 1 % de significativité. De même, la dichotomique du Québec hausse l'expérience de 0,890 années dans la procédure #2 toujours à un niveau de significativité de 1 %. La variable de services de garde est toutefois légèrement négative sans être significative.

Puis, les tests d'instruments du tableau 3.13 livrent des résultats similaires au tableau 3.7. Les instruments paraissent aussi faibles. Le R^2 de Shea partiel ajusté augmente de quelques dix millièmes tandis que la statistique du test de Stock et Yogo (2005) diminue légèrement à l'exception de la procédure #1. Par contre, toutes les statistiques du test des restrictions de suridentification de Wooldridge (1995a) augmentent. La p-value de ce test pour la procédure #2 passe donc de 40,73 % à 51,16 %. Les deux variables de politique ne règlent pas le problème de la faiblesse des instruments, mais améliorent tout de même leur validité.

Finalement, les résultats des formes réduites du tableau 3.14 visent à distribuer entre les instruments l'impact de l'expérience sur le salaire horaire réel. Seuls l'éducation, le statut familial, l'âge et son carré conservent le même signe et une significativité d'au moins 10 % pour les méthodes MCO, EF et pour les procédures #3 et #4. Les autres instruments ne restent pas significatifs pour toutes ces méthodes. L'âge du conjoint et son carré de même que l'indicateur de semaines de chômage non rapportées ne sont jamais significatifs. La procédure #3 ne détecte pas de biais de sélection contemporain alors que les termes de correction $\hat{\lambda}_{it2}$ sont significatifs à 8,54 % pour la procédure #4. Sur les 17 moyennes individuelles introduites dans la procédure #4 pour corriger pour l'hétérogénéité, seules 9 variables demeurent statistiquement significatives à 10 %. Concernant les variables de politique, seule la dichotomique d'exposition aux services de garde augmente le salaire de 4,1 % significatif à 5 % pour la méthode MCO. Autrement, son rendement est négatif, mais non significatif. La dichotomique du Québec n'est jamais significative et obtient un rendement négatif pour les méthodes EF et pour la procédure #3. En somme, ces résultats ne permettent pas de trancher définitivement en faveur de notre hypothèse de départ ce qui requiert davantage d'explications.

Tableau 3.5 Estimés du logarithme du salaire horaire réel de Semykina et Wooldridge (2006, p. 35).

Régresseurs	MCO	MCDE	EF	EF-MCDE	P#1	P#2
Expérience	0,067***	0,062***	0,082***	0,067***	0,068**	0,055**
	(0,006)	(0,011)	(0,010)	(0,023)	(0,030)	(0,026)
Expérience^2	-0,0014***	-0,00166***	-0,00091***	-0,00091***	-0,00066**	-0,00043
	(0,0002)	(0,0004)	(0,00017)	(0,00032)	(0,00033)	(0,00035)
Éducation	0,114***	0,115***	0,023**	0,025**	0,024**	0,035***
	(0,006)	(0,007)	(0,0114)	(0,0115)	(0,0116)	(0,013)
Test de Wald sur les $\hat{\lambda}_{it2}$					χ^2_{13}=26,96 (0,0126)	χ^2_{13}=15,03 (0,305)

▪ Les résultats pour les variables dichotomiques d'année et les termes de correction estimés $\hat{\lambda}_{it2}$ sont omis par souci de concision.
▪ Les écarts-types sont robustes à l'hétéroscédasticité et à la corrélation en série (*cluster*).
▪ Les p-values robustes sont entre parenthèses sous les statistiques des tests de Wald.
▪ Les écarts-types de la procédure #2 sont aussi corrigés pour l'estimation de la première étape.
▪ * = significatif à 10 %; ** = significatif à 5 %; *** = significatif à 1 %.

Tableau 3.6 Estimés du logarithme du salaire horaire réel avec notre échantillon et une spécification semblable à celle Semykina et Wooldridge (2006).

Régresseurs	MCO	MCDE	EF	EF-MCDE	P#1	P#2
Expérience	0,032***	0,058***	0,038***	0,036	0,047	0,061***
	(0,0010)	(0,0019)	(0,0034)	(0,036)	(0,046)	(0,0051)
Expérience^2	-0,00054***	-0,00128***	-0,00054***	-0,00098***	-0,00094***	-0,0013***
	(0,000029)	(0,000059)	(0,000046)	(0,00008)	(0,00012)	(0,00014)
Éducation	0,069***	0,068***	0,041***	0,034***	0,036***	0,036***
	(0,0013)	(0,0013)	(0,0072)	(0,0092)	(0,0101)	(0,0079)
Test de Wald sur les $\hat{\lambda}_{it2}$					χ^2_{12}=6,86 (0,8667)	χ^2_{12}=22,32 (0,0340)

▪ Commentaires identiques au tableau 3.5.

Tableau 3.7 Tests des instruments de nos modèles avec une spécification similaire à celle de Semykina et Wooldridge (2006).

	EF-MCDE		Procédure #1	
	Expérience	Expérience^2	Expérience	Expérience^2
R^2	0,9158	0,7859	0,9181	0,7921
R^2 ajusté	0,9157	0,7857	0,9181	0,7920
R^2 partiel	0,0047	0,3066	0,0059	0,2578
F(14, 13355)	5,02524 (0,0000)	211,15 (0,0000)	6,52228 (0,0000)	302,149 (0,0000)
R^2 de Shea partiel	0,0065	0,4225	0,0041	0,1776
R^2 de Shea partiel ajusté	0,0060	0,4222	0,0034	0,1770
g_{min}	19,83 (5 %) > 16,77 > 10,89 (10 %)		19,83 (5 %) > 11,27 > 10,89 (10 %)	
Wooldridge	χ^2_{12} = 43,8906 (0,0000)		χ^2_{12} = 51,953 (0,0000)	
	MCDE		Procédure #2	
	Expérience	Expérience^2	Expérience	Expérience^2
R^2	0,4909	0,4476	0,4940	0,4507
R^2 ajusté	0,4906	0,4474	0,4935	0,4502
R^2 partiel	0,4806	0,4338	0,0124	0,0126
F(14, 18233)	1441,03 (0,0000)	888,619 (0,0000)	83,7157 (0,0000)	52,4718 (0,0000)
R^2 de Shea partiel	0,2885	0,2604	0,0166	0,0168
R^2 de Shea partiel ajusté	0,2882	0,2601	0,0156	0,0159
g_{min}	1 361,36 > 19,83 (5 %)		49,48 > 19,83 (5 %)	
Wooldridge	χ^2_{12} = 1794,45 (0,0000)		χ^2_{12} = 12,4882 (0,4073)	

▪ La valeur Eigen minimale $(g_{\min})$ est mise en relation avec les valeurs critiques fournies dans les tableaux de Stock et Yogo (2005, p. 100-101). Le biais relatif de l'estimateur VI par rapport à l'estimateur MCO associé à la valeur critique est inscrit entre parenthèses.

▪ Les p-values sont entre parenthèses sous la statistique F et le test des conditions de suridentification de Wooldridge (1995a).

Tableau 3.8 Estimations pour le logarithme du salaire réel horaire avec la spécification incluant les variables de politique (début).

	MCO	MCDE	EA	EF
	(1)	(2)	(3)	(4)
Expérience	0,032***	0,058***	0,032***	0,038***
	(0,0010)	(0,0019)	(0,00097)	(0,0034)
Expérience^2	-0,00054***	0,00128***	-0,00053***	-0,00054***
	(0,000029)	(0,000059)	(0,000027)	(0,000046)
Éducation	0,069***	0,068***	0,068***	0,041***
	(0,0013)	(0,0013)	(0,0013)	(0,0072)
Observations	55 520	55 520	50 642	50 642
Personnes	18 234	18 234	13 356	13 356
	EA-VI	EF-MCDE	Procédure #1	Procédure #2
	(5)	(6)	(7)	(8)
Expérience	0,050***	0,025	0,018	0,064***
	(0,0014)	(0,0347)	(0,0409)	(0,0051)
Expérience^2	-0,00104***	-0,00098***	-0,00091***	-0,00142***
	(0,000042)	(0,000080)	(0,000109)	(0,000141)
Éducation	0,066***	0,032***	0,031***	0,038***
	(0,0010)	(0,0091)	(0,0096)	(0,0079)
Observations	50 642	50 642	50 642	55 520
Personnes	13 356	13 356	13 356	18 234

▪ Voir la fin du tableau 3.8 à la page 82 pour les commentaires.

Tableau 3.8 Estimations pour le logarithme du salaire horaire réel avec la spécification incluant les variables de politique (suite).

	Procédure #1	Procédure #2
Éducation (moy.)		0,023***
		(0,0079)
Âge (moy.)		-0,013**
		(0,0063)
Âge^2 (moy.)		0,000146**
		(0,000073)
Mariée (moy.)		0,052***
		(0,0092)
Autre revenu familial (10 000 $) (moy.)		0,018***
		(0,0052)
Autre revenu familial^2 (moy.)		0,00010
		(0,00053)
Enfants de 0 à 2 ans (moy.)		0,034***
		(0,0116)
Enfants de 3 à 5 ans		0,029**
		(0,0117)
Enfants de 6 à 17 ans (moy.)		-0,014***
		(0,0048)
Âge du conjoint		-0,00046
		(0,0037)
Âge du conjoint^2		0,000016
		(0,000043)
Éducation du conjoint		0,043***
		(0,0047)
Éducation du conjoint^2		-0,0011***
		(0,00016)
Semaines de chômage du conjoint (moy.)		-0,0014***
		(0,00050)

▪ Voir la fin du tableau 3.8 à la page suivante pour les commentaires.

Tableau 3.8 Estimations pour le logarithme du salaire horaire réel avec la spécification incluant les variables de politique (suite et fin).

	Procédure #1	Procédure #2
Indicateur de semaines de chômage non rapportées (moy.)		0,015 (0,0616)
Québec (moy.)		0,0033 (0,0091)
Politique de garde (moy.)		0,036 (0,0220)
Test de Wald sur les $\hat{\lambda}_{it2}$	χ^2_{12}=7,97 (0,7872)	χ^2_{12}= 23,32 (0,0251)

▪ Le salaire est en dollars constants de 2002 pour chaque province avec le contenu de panier de 2005 (série 326-0021 de CANSIM).
▪ Les résultats pour les variables dichotomiques d'année, les termes de correction estimés $\hat{\lambda}_{it2}$ et la constante sont omis par souci de concision.
▪ Les écarts-types sont robustes à l'hétéroscédasticité et à la corrélation en série (*cluster*) sauf pour le modèle à effets aléatoires.
▪ Les p-values robustes sont entre parenthèses sous les statistiques des tests de Wald.
▪ * = significatif à 10 %; ** = significatif à 5 %; *** = significatif à 1 %.
▪ « (moy.) » désigne la moyenne de la variable pour un individu donné pour les t périodes pour lesquelles les informations sont disponibles conformément à la procédure #2 tirée de Semykina et Wooldridge (2006, p. 20).

Tableau 3.9 Tests de Hausman (*voir* tableaux 3.5, 3.6 et 3.8).

Estimateurs comparés	Semykina et Wooldridge (2006)	Sans les variables de politique	Avec les variables de politique
MCDE et MCO	$\chi^2_1 = 0{,}29$ (0,5876)	$\chi^2_1 = 259{,}00$ (0,0000)	$\chi^2_1 = 259{,}00$ (0,0000)
EF et EA	—	—	$\chi^2_1 = 3{,}39$ (0,0656)
EF-MCDE et EA-VI	—	—	$\chi^2_1 = 0{,}52$ (0,4709)

Tableau 3.10 Premières étapes des modèles MCDE et EF-MCDE (*voir* tableau 3.8) (début).

Variables instrumentales	MCDE		EF-MCDE	
	Expérience	Expérience^2	Expérience	Expérience^2
Observations	55 520		50 642	
Personnes	18 234		13 356	
R^2 ajusté	0,4907	0,4475	0,9157	0,7857
Éducation	0,030	0,818	-0,167***	-3,510***
	(0,0213)	(0,8080)	(0,0200)	(0,8667)
Âge	1,2798***	6,113**	-0,831***	-57,680***
	(0,0702)	(3,1131)	(0,2408)	(14,9222)
Âge^2	-0,0076***	0,199***	-0,000013	0,558***
	(0,00094)	(0,0431)	(0,00033)	(0,0217)
Mariée	-0,521***	-19,810***	-0,024	-2,774**
	(0,1305)	(4,6157)	(0,0242)	(1,1137)
Autre revenu familial (10 000 $)	0,1101*	3,719	-0,00083	0,449
	(0,0660)	(2,5949)	(0,0063)	(0,4196)
Autre revenu familial^2	-0,0181***	-0,657***	-0,00034	-0,043
	(0,0064)	(0,2465)	(0,00064)	(0,0400)
Enfants de 0 à 2 ans	-0,025	-7,593***	0,036***	0,021
	(0,0604)	(1,5574)	(0,0126)	(0,5162)
Enfants de 3 à 5 ans	-0,032	-8,105***	-0,041***	-0,564
	(0,0692)	(1,9892)	(0,0149)	(0,6991)
Enfants de 6 à 17 ans	-0,745***	-22,753***	-0,057***	1,016
	(0,0610)	(2,0199)	(0,0124)	(0,7539)

▪ Voir la fin du tableau 3.10 à la page suivante pour les commentaires.

Tableau 3.10 Premières étapes des modèles MCDE et EF-MCDE (*voir* tableau 3.8) (suite et fin).

Variables instrumentales	MCDE		EF-MCDE	
	Expérience	Expérience^2	Expérience	Expérience^2
Observations	55 520		50 642	
Personnes	18 234		13 356	
R^2 ajusté	0,4907	0,4475	0,9157	0,7857
Âge du conjoint	0,186***	6,812***	0,024	2,074
	(0,0611)	(2,5449)	(0,0281)	(1,5871)
Âge du conjoint^2	-0,0024***	-0,089***	-0,00025	-0,027
	(0,00077)	(0,0332)	(0,00030)	(0,0190)
Éducation du conjoint	0,534***	13,610***	-0,00034	5,956**
	(0,0932)	(3,5274)	(0,0594)	(2,6810)
Éducation du conjoint^2	-0,018***	-0,482***	0,00081	-0,189**
	(0,0030)	(0,1126)	(0,0017)	(0,0788)
Semaines de chômage du conjoint	-0,019***	-0,406*	0,000033	0,020
	(0,0057)	(0,2142)	(0,00046)	(0,0265)
Indicateur de semaines de chômage non rapportées	2,750**	100,531*	-0,131	-10,97
	(1,2092)	(58,1357)	(0,0993)	(8,6110)
Québec	0,250	10,620*	0,132	1,456
	(0,1603)	(6,1258)	(0,0974)	(4,1919)
Politique de garde	-0,230	-16,505**	0,054	-1,423
	(0,2418)	(7,9011)	(0,0402)	(1,4841)

▪ Des variables dichotomiques d'année et une constante sont incluses dans toutes les régressions.
▪ Les écarts-types sont robustes à l'hétéroscédasticité et à la corrélation en série (*cluster*).
▪ * = significatif à 10 %; ** = significatif à 5 %; *** = significatif à 1 %.

Tableau 3.11 Premières étapes du modèle EA-VI et de la procédure #1 (*voir* tableau 3.8) (début).

Variables instrumentales	EA-VI		Procédure #1	
	Expérience	Expérience^2	Expérience	Expérience^2
Observations	50 642		50 642	
Personnes	13 356		13 356	
R^2 ajusté	—	—	0,9180	0,7920
Éducation	0,051***	1,775***	-0,162***	-3,656***
	(0,0108)	(0,4113)	(0,0198)	(0,8646)
Âge	0,958***	-8,025***	-0,890***	-62,494***
	(0,0305)	(1,1605)	(0,1616)	(19,7077)
Âge^2	-0,0036***	0,374***	0,00055	0,624***
	(0,00036)	(0,0137)	(0,00035)	(0,0214)
Mariée	-0,316***	-13,521***	-0,0195	-2,101*
	(0,0554)	(2,1052)	(0,0234)	(1,0836)
Autre revenu familial (10 000 $)	0,012	0,902	0,00222	0,223
	(0,0173)	(0,6562)	(0,0064)	(0,4199)
Autre revenu familial^2	-0,0035**	-0,175***	-0,000794	-0,025
	(0,0018)	(0,0668)	(0,00065)	(0,0405)
Enfants de 0 à 2 ans	-0,026	-3,766***	0,026*	1,8634***
	(0,0298)	(1,1281)	(0,0141)	(0,6267)
Enfants de 3 à 5 ans	-0,161***	-6,389***	-0,0301*	2,501***
	(0,0302)	(1,1497)	(0,0155)	(0,7475)
Enfants de 6 à 17 ans	-0,290***	-7,249***	-0,0157	4,783***
	(0,0230)	(0,8733)	(0,0128)	(0,7566)

▪ Voir la fin du tableau 3.11 à la page suivante pour les commentaires.

Tableau 3.11 Premières étapes du modèle EA-VI et de la procédure #1 (*voir* tableau 3.8) (suite et fin).

Variables instrumentales	EA-VI		Procédure #1	
	Expérience	Expérience^2	Expérience	Expérience^2
Observations	50 642		50 642	
Personnes	13 356		13 356	
R^2 ajusté	—	—	0,9180	0,7920
Âge du conjoint	0,053*	2,208**	-0,0262	-0,719
	(0,0286)	(1,0857)	(0,0274)	(1,5559)
Âge du conjoint^2	-0,00084***	-0,036***	0,00044	0,0102
	(0,00031)	(0,0119)	(0,00030)	(0,0187)
Éducation du conjoint	0,582***	16,309***	-0,0020	5,293*
	(0,0392)	(1,4900)	(0,0609)	(3,0423)
Éducation du conjoint^2	-0,019***	-0,564***	0,00073	-0,185**
	(0,0013)	(0,0484)	(0,0017)	(0,0931)
Semaines de chômage du conjoint	-0,0022*	-0,034	0,000063	0,0152
	(0,0013)	(0,0509)	(0,00045)	(0,0257)
Indicateur de semaines de chômage non rapportées	0,477**	15,120*	0,066	0,936
	(0,2257)	(8,5795)	(0,0992)	(8,6326)
Québec	0,288***	11,121***	0,133	4,235
	(0,0761)	(2,8941)	(0,1003)	(4,1701)
Politique de garde	-0,0068	-7,186**	0,112***	1,648
	(0,0946)	(3,5954)	(0,0389)	(1,4709)

▪ Les résultats pour les variables dichotomiques d'année, les termes de correction estimés $\hat{\lambda}_{it2}$ et la constante sont omis par souci de concision.

▪ Les écarts-types sont robustes à l'hétéroscédasticité et à la corrélation en série (*cluster*) sauf pour le modèle à effets aléatoires avec instruments.

▪ * = significatif à 10 %; ** = significatif à 5 %; *** = significatif à 1 %.

Tableau 3.12 Premières étapes de la procédure #2 (*voir* tableau 3.8) (début).

Variables instrumentales	Procédure #2	
	Expérience	Expérience^2
Observations	55 520	
Individus	18 234	
R^2 ajusté	0,4937	0,4504
Éducation	-0,369***	-7,649***
	(0,0733)	(2,1798)
Âge	-0,183	-46,705***
	(0,1669)	(6,7319)
Âge^2	0,0104***	0,842***
	(0,0021)	(0,0869)
Mariée	0,300***	6,859***
	(0,0826)	(2,4028)
Autre revenu familial (10 000 $)	-0,155***	-5,547***
	(0,0416)	(1,6330)
Autre revenu familial^2	0,0194***	0,638***
	(0,0040)	(0,1500)
Enfants de 0 à 2 ans	0,923***	30,024***
	(0,1056)	(4,4744)
Enfants de 3 à 5 ans	0,998***	33,156***
	(0,1125)	(4,6325)
Enfants de 6 à 17 ans	0,512***	20,263***
	(0,0860)	(3,2902)
Âge du conjoint	0,022	0,484
	(0,1200)	(4,2485)
Âge du conjoint^2	-0,00078	-0,0283
	(0,0014)	(0,0524)
Éducation du conjoint	-0,073	-4,168
	(0,3034)	(7,1613)
Éducation du conjoint^2	0,0013	0,048
	(0,0087)	(0,2113)

▪ Voir la fin du tableau 3.12 à la page 89 suivante pour les commentaires.

Tableau 3.12 Premières étapes de la procédure #2 (*voir* tableau 3.7) (suite).

Variables instrumentales	Procédure #2	
	Expérience	Expérience^2
Semaines de chômage du conjoint	-0,0045	-0,146
	(0,0030)	(0,1032)
Indicateur de semaines de chômage non rapportées	0,434	23,936
	(0,9350)	(44,730)
Québec	0,890***	32,498***
	(0,2759)	(9,7607)
Politique de garde	-0,098	-5,750
	(0,1520)	(3,9138)
Éducation (moy.)	0,234***	1,987
	(0,0745)	(2,1866)
Âge (moy.)	1,0076***	34,634***
	(0,1529)	(5,9443)
Âge^2 (moy.)	-0,012***	-0,395***
	(0,0019)	(0,0773)
Mariée (moy.)	-0,614***	-17,862***
	(0,1641)	(5,5080)
Autre revenu familial (10 000 $) (moy.)	-0,161	-7,524*
	(0,1012)	(3,9848)
Autre revenu familial^2 (moy.)	0,0049	0,403
	(0,0099)	(0,3809)
Enfants de 0 à 2 ans (moy.)	-0,310***	-11,417***
	(0,1193)	(3,1154)
Enfants de 3 à 5 ans (moy.)	0,055	-1,967
	(0,1424)	(4,0176)
Enfants de 6 à 17 ans (moy.)	-0,948***	-29,787***
	(0,0975)	(3,2642)

▪ Voir la fin du tableau 3.12 à la page suivante pour les commentaires.

Tableau 3.12 Premières étapes de la procédure #2 (*voir* tableau 3.7) (suite et fin).

Variables instrumentales	Procédure #2	
	Expérience	Expérience^2
Âge du conjoint (moy.)	0,194	7,248
	(0,1386)	(5,0213)
Âge du conjoint^2 (moy.)	-0,0018	-0,068
	(0,0016)	(0,0625)
Éducation du conjoint (moy.)	0,374	8,793
	(0,3190)	(8,0960)
Éducation du conjoint^2 (moy.)	-0,012	-0,244
	(0,0093)	(0,2461)
Semaines de chômage du conjoint (moy.)	-0,0206**	-0,272
	(0,0101)	(0,3803)
Indicateur de semaines de chômage non rapportées (moy.)	3,770**	130,075
	(1,6975)	(80,4742)
Québec (moy.)	-0,191	-3,885
	(0,3046)	(10,6948)
Politique de garde (moy.)	-0,141	-11,740
	(0,3228)	(10,0267)

▪ Les résultats pour les variables dichotomiques d'année et les termes de correction estimés $\hat{\lambda}_{it2}$ sont omis par souci de concision.

▪ Les écarts-types sont robustes à l'hétéroscédasticité et à la corrélation en série (*cluster*). Par contre, ils ne tiennent pas compte de la correction du biais de sélection.

▪ * = significatif à 10 %; ** = significatif à 5 %; *** = significatif à 1 %.

Tableau 3.13 Tests des instruments de nos modèles avec la spécification contenant les variables de politique.

	EF-MCDE		Procédure #1	
	Expérience	Expérience^2	Expérience	Expérience^2
R^2	0,9158	0,7859	0,9181	0,7921
R^2 ajusté	0,9157	0,7857	0,9180	0,7920
R^2 partiel	0,0049	0,3066	0,0064	0,2598
F(14, 13355)	4,69804 (0,0000)	185,691 (0,0000)	6,16496 (0,0000)	264,434 (0,0000)
R^2 de Shea partiel	0,0068	0,4226	0,0049	0,1977
R^2 de Shea partiel ajusté	0,0063	0,4223	0,0041	0,1970
g_{min}	20,12 (5 %) > 15,29 > 10,96 (10 %)		20,12 (5 %) > 11,69 > 10,96 (10 %)	
Wooldridge	χ^2_{14} = 47,8102 (0,0000)		χ^2_{14} = 56,405 (0,0000)	
	MCDE		Procédure #2	
	Expérience	Expérience^2	Expérience	Expérience^2
R^2	0,4910	0,4478	0,4943	0,4510
R^2 ajusté	0,4907	0,4475	0,4937	0,4504
R^2 partiel	0,4807	0,4339	0,0122	0,0124
F(14, 18233)	1261,02 (0,0000)	780,673 (0,0000)	70,4517 (0,0000)	42,8164 (0,0000)
R^2 de Shea partiel	0,2887	0,2606	0,0166	0,0169
R^2 de Shea partiel ajusté	0,2883	0,2603	0,0156	0,0159
g_{min}	1 192,54 > 20,12 (5 %)		42,66 > 20,12 (5 %)	
Wooldridge	χ^2_{14} = 1 814,89 (0,0000)		χ^2_{14} = 13,1909 (0,5116)	

▪ La valeur Eigen minimale $(g_{\min})$ est mise en relation avec les valeurs critiques fournies dans les tableaux de Stock et Yogo (2005, p. 100-101). Le biais relatif de l'estimateur VI par rapport à l'estimateur MCO associé à la valeur critique est inscrit entre parenthèses.

▪ Les p-values sont entre parenthèses sous la statistique F et le test des conditions de suridentification de Wooldridge (1995a).

Tableau 3.14 Estimations pour le logarithme du salaire horaire réel en forme réduite avec la spécification contenant les variables de politique (début).

Variable explicative	MCO	EF	Procédure #3	Procédure #4
Observations	55 520	50 642	50 642	55 520
Personnes	18 234	13 356	13 356	18 234
R^2	0,9805	0,1666	0,1684	0,9806
Éducation	0,059***	0,029***	0,030***	0,022***
	(0,0014)	(0,0071)	0,0072	(0,0082)
Âge	0,053***	0,046***	0,047***	0,067***
	(0,0035)	(0,0119)	(0,0138)	(0,0113)
$Âge^2$	-0,00053***	-0,00056***	-0,00059***	-0,00060***
	(0,000044)	(0,000097)	(0,00011)	(0,00013)
Mariée	0,048***	0,017*	0,016*	0,026**
	(0,0080)	(0,0094)	(0,0095)	(0,0105)
Autre revenu familial (10 000 $)	0,014***	0,0016	0,0036	0,0089**
	(0,0034)	(0,0022)	(0,0030)	(0,0044)
Autre revenu familial^2	0,00025	-0,000029	-0,00025	-0,00055
	(0,00034)	(0,00024)	(0,00031)	(0,00045)
Enfants de 0 à 2 ans	0,043***	-0,0042	-0,0088	0,0074
	(0,0056)	(0,0040)	(0,0060)	(0,0085)
Enfants de 3 à 5 ans	0,041***	-0,0056	-0,010	0,0074
	(0,0056)	(0,0046)	(0,0067)	(0,0094)
Enfants de 6 à 17 ans	-0,021***	-0,0045	-0,0067	-0,0018
	(0,0035)	(0,0038)	(0,0046)	(0,0064)
Âge du conjoint	0,0014	-0,0092	-0,011	-0,013
	(0,0033)	(0,0083)	(0,0084)	(0,0096)
Âge du conjoint^2	$-4{,}2\cdot10^{-6}$	0,000057	0,000076	0,000052
	(0,000038)	0,000088	(0,000090)	(0,00011)
Éducation du conjoint	0,056***	0,024	0,0266	0,0091
	(0,0044)	(0,0248)	(0,0247)	(0,0255)
Éducation du conjoint^2	-0,0015***	-0,00015	-0,00022	0,00027
	(0,00015)	(0,00076)	(0,00076)	(0,00078)
Semaines de chômage du conjoint	-0,0014***	-0,00020	-0,00022	-0,00015
	(0,00029)	0,00015	(0,00015)	(0,00019)
Indicateur de semaines de chômage non rapportées	0,059	0,011	0,012	0,0076
	(0,0432)	(0,0249)	(0,0253)	(0,0359)
Québec	0,0099	-0,038	-0,0410	0,021
	(0,0083)	(0,0407)	(0,0407)	(0,0489)
Politique de garde	0,041**	-0,010	-0,0079	-0,0051
	(0,0173)	(0,0119)	(0,0118)	(0,0158)

▪ Voir à la fin du tableau 3.14 à la page 93 pour les commentaires.

Tableau 3.14 Estimations pour le logarithme du salaire horaire réel en forme réduite avec la spécification contenant les variables de politique (suite et fin).

Variable explicative	MCO	EF	Procédure #3	Procédure #4
Éducation (moy.)				0,039*** (0,0082)
Âge (moy.)				-0,0080 (0,0107)
Âge^2 (moy.)				-0,000016 (0,00012)
Mariée (moy.)				0,021 (0,0135)
Autre revenu familial (10 000 $) (moy.)				0,0092* (0,0052)
Autre revenu familial^2 (moy.)				0,00054 (0,00052)
Enfants de 0 à 2 ans (moy.)				0,034*** (0,0110)
Enfants de 3 à 5 ans (moy.)				0,041*** (0,0115)
Enfants de 6 à 17 ans (moy.)				-0,029*** (0,0062)
Âge du conjoint (moy.)				0,015 (0,0103)
Âge du conjoint^2 (moy.)				-0,000068 (0,00011)
Éducation du conjoint (moy.)				0,048* (0,0261)
Éducation du conjoint^2 (moy.)				-0,0018** (0,00080)
Semaines de chômage du conjoint (moy.)				-0,0023*** (0,00052)
Indicateur de semaines de chômage non rapportées (moy.)				0,057 0,0608
Québec (moy.)				-0,016 (0,0496)
Politique de garde (moy.)				0,051** 0,0261
Test de Wald sur les $\hat{\lambda}_{it2}$			χ^2_{12} = 10,9139 (0,5363)	χ^2_{12} = 19,1321 (0,0854)

▪ Voir à la fin du tableau 3.14 à la page suivante pour les commentaires.

▪ Le salaire est en dollars constants de 2002 pour chaque province avec le contenu de panier de 2005 (série 326-0021 de CANSIM).
▪ Les résultats pour les variables dichotomiques d'années et les termes de correction estimés $\hat{\lambda}_{it2}$ sont omis par souci de concision.
▪ Les écarts-types sont robustes à l'hétéroscédasticité et à la corrélation en série (*cluster*).
▪ Les p-values robustes sont entre parenthèses sous les statistiques des tests de Wald.
▪ * = significatif à 10 %; ** = significatif à 5 %; *** = significatif à 1 %.
▪ « (moy.) » désigne la moyenne de la variable pour un individu donné pour les t périodes pour lesquelles les informations sont disponibles conformément à la procédure #4 tirée de Wooldridge (1995b, p. 130).

3.4 Discussion

D'abord, nous avons essayé de respecter autant que possible la spécification de Semykina et Wooldridge (2006), mais le principal obstacle fut la conception de la variable d'expérience de l'EDTR *yrxfte11*. Aucune alternative ne s'est avérée concluante afin d'inclure dans nos calculs l'expérience du travail à temps partiel avant le premier emploi à temps plein. Cette différence de mesure de l'expérience explique peut-être l'écart de 3,5 % entre notre rendement de l'expérience en MCO et celui de Semykina et Wooldridge (2006). Les rendements des tableaux 3.5 et 3.6 et les tests de Hausman du tableau 3.9 montrent aussi que l'endogénéité joue un rôle beaucoup plus important dans notre échantillon alors que le contraire s'applique pour les résultats de Semykina et Wooldridge (2006).

Ces divergences restent cependant mineures comparativement à la perte de significativité des rendements de l'expérience pour la méthode EF-MCDE et pour la procédure #1. Cette anomalie a suscité une investigation plus profonde qui a révélé des instruments faibles et non valides à l'exception de la procédure #2. Ce constat jette un doute sur le résultat du test du biais de sélection contemporain. En effet, les termes de correction $\hat{\lambda}_{it2}$ sont significatifs dans la procédure #2 de notre échantillon, mais pas pour la procédure #1. Le contraire s'applique pour Semykina et Wooldridge (2006). D'ailleurs, les auteurs ne testent pas leurs instruments et ne fournissent pas leurs premières étapes. Pourtant, la significativité de leurs résultats diminue pour les procédures #1 et #2 comparativement aux autres méthodes. Le carré de l'expérience n'est plus statistiquement significatif dans leur procédure #2.

Le faible pouvoir explicatif des régresseurs exogènes exclus s'explique par une interaction particulière entre l'expérience et les variables dichotomiques d'année en déviation par rapport à la moyenne individuelle. L'expérience est une variable

singulière à modéliser, car elle varie fréquemment d'une année telle que rapportée par la variable *yrxfte11* de l'EDTR. Ainsi, si l'expérience augmente constamment d'une année pour toutes les personnes, alors l'expérience est parfaitement corrélée avec l'année en déviation par rapport à la moyenne individuelle. La composante orthogonale du vecteur d'expérience à une matrice de variables dichotomiques d'année est alors nulle. Autrement dit, la matrice des variables dichotomiques d'année explique parfaitement le vecteur d'expérience en déviation par rapport à la moyenne individuelle lorsque l'expérience augmente toujours d'une année. Pour des données empiriques, les personnes n'occupent pas toujours un emploi et l'expérience stagne parfois si bien que cette relation demeure très forte, mais approximative.

Cette interaction explique les R^2 supérieurs à 90 % pour les premières étapes de l'expérience de la méthode EF-MCDE et de la procédure #1 (*voir* tableaux 3.7 et 3.13). Toutefois, le R^2 partiel et le R^2 de Shea chutent drastiquement pour ces deux méthodes, car ces deux statistiques ne considèrent que les composantes orthogonales aux régresseurs exogènes inclus dont les variables dichotomiques d'année qui détiennent un très fort pouvoir explicatif. La différence entre le R^2 et le R^2 partiel ou de Shea n'est pas aussi marquée pour la méthode MCDE, car les composantes orthogonales détiennent un meilleur pouvoir explicatif lorsque les moyennes individuelles ne sont pas soustraites. Le R^2 partiel et le R^2 ne diminuent pas autant pour l'expérience au carré justement parce que les variables dichotomiques d'année n'expliquent pas aussi bien cette variable. Pour la procédure #2, le R^2 se compare à celui de la méthode MCDE, mais cette fois-ci, la baisse marquée du R^2 partiel et du R^2 de Shea se produit autant pour l'expérience que pour l'expérience au carré. À la différence de la méthode MCDE, les régresseurs exogènes inclus comprennent les moyennes des instruments. Le vecteur des composantes orthogonales de l'expérience et de l'expérience au carré par rapport à la matrice de régresseurs exogènes inclus est

probablement faible. Bref, toutes ces complications nuisent grandement à la robustesse et à l'efficience de nos résultats.

En définitive, Semykina et Wooldridge (2006) optent pour la procédure #2 parmi toutes les autres à cause de leur test du biais de sélection contemporain positif. De même, la procédure #2 nous apparaît comme la meilleure. D'une part, elle corrige pour l'endogénéité et pour l'hétérogénéité, deux problèmes décelés par les tests de Hausman (*voir* tableau 3.9). D'autre part, la procédure #1 ne détecte pas de biais de sélection contemporain, mais la faiblesse et l'invalidité de ses instruments biaisent peut-être cette conclusion. Ainsi, le R^2 de Shea partiel ajusté de la procédure #2 est 2,5 fois plus élevé que celui de la méthode EF-MCDE (*voir* tableau 3.13) et le test des restrictions de suridentification de Wooldridge (1995a) conclut que ses instruments sont valides. Quoique le R^2 de Shea partiel ajusté ne soit que de 1,56 %, la procédure #2 constitue un compromis entre des méthodes qui présentent toutes des faiblesses dans un contexte d'endogénéité, d'hétérogénéité et de biais de sélection. À tout le moins, le test de Stock et Yogo (2005) non robuste à l'hétéroscédasticité établit à 5 % au maximum le biais de l'estimateur de la procédure #2 par rapport à l'estimateur MCO. Après tout, la procédure #2 n'arrive pas à des rendements de l'expérience et de l'éducation statistiquement différents des résultats de Semykina et Wooldridge (2006). Le prix de cette correction est l'hypothèse non testable de la linéarité de l'effet fixe en fonction des moyennes inter temporelles des instruments (34).

La procédure #2 permet-elle de conclure que la politique de services de garde du Québec a augmenté le salaire des mères du Québec? Dans la première étape de la procédure #2 de l'expérience, la variable dichotomique du Québec a un effet positif significatif et important, mais la variable de services de garde n'a pas d'impact significatif une fois ce facteur clé contrôlé (*voir* tableau 3.12). Les résultats des

modèles en forme réduite ne raffinent guère ces remarques puisque peu de variables sont significatives pour les autres méthodes que MCO. La variable de services de garde a bel et bien un rendement de 4,1 % sur le salaire significatif à 5 %, mais les autres rendements sont négatifs et/ou non significatifs. Bref, il s'avère difficile d'isoler la politique une fois l'effet du Québec pris en compte. Après avoir contourné les problèmes d'endogénéité, d'hétérogénéité, de biais de sélection, de faiblesse et d'invalidité des instruments, l'impact de la politique de services de garde du Québec se limite à une retombée territoriale. La qualité des données et notre méthodologie ne nous permettent pas de préciser l'analyse au niveau des mères québécoises ayant été exposées à la politique.

CONCLUSION

Ce travail cherchait à établir un lien empirique entre la nouvelle politique de services de garde dont s'est graduellement doté le Québec dès septembre 1997 et une possible prime salariale des mères québécoises causée par une augmentation de leur offre de travail et donc de leur expérience.

Le premier chapitre a permis de situer notre démarche au sein des études existantes. Ainsi, la littérature économique reconnaît généralement la présence d'une pénalité salariale maternelle. De leur côté, Lefebvre et Merrigan (2005a, 2005b, 2007a, 2007b) et Baker *et al.* (2005) ont découvert que la politique québécoise a significativement accru l'offre de travail des mères du Québec par rapport au reste du Canada. Finalement, le lien entre l'expérience et les salaires appuie notre hypothèse sur une possible prime salariale obtenue par les mères québécoises par rapport au reste du Canada. Le deuxième chapitre a tenté de clarifier la méthode de Semykina et Wooldridge (2006) concernant les données de panel en présence d'endogénéité et d'un biais de sélection. Il investigue aussi les différents tests permettant de détecter des instruments faibles ou invalides. Enfin, le chapitre trois a présenté l'application de cette méthode et les résultats obtenus.

De cette analyse ressort des problèmes liés à la conception de la variable d'expérience *yrxfte11* et à la faiblesse et à l'invalidité des instruments qui biaisent nos estimateurs. La procédure #2 nous apparaît comme le meilleur choix puisqu'elle corrige pour l'endogénéité, l'hétérogénéité, le biais de sélection et ses instruments quoique possiblement faibles réussissent le test des restrictions de suridentification de

Wooldridge (1995a). Un tel compromis fait toutefois appel à l'hypothèse non vérifiable de la linéarité de l'effet fixe en fonction des moyennes inter temporelles des instruments. La première étape de l'expérience de la procédure #2 montre un impact positif de la variable dichotomique du Québec tandis que la variable de la politique de services de garde n'est pas significative. Les Québécoises ont indéniablement profité d'une augmentation de leur expérience, mais raffiner cette analyse uniquement pour les mères ayant été exposées à la nouvelle politique de services de garde dépasse les capacités de notre méthodologie pour l'instant.

Ce travail souligne l'importance de divulguer les résultats des premières étapes et de tester pour la faiblesse et la validité des instruments. Toutes corrections pour l'endogénéité devraient aborder ces deux facettes fondamentales. Il aurait été intéressant d'effectuer ces tests avec les données et la spécification de Semykina et Wooldridge (2006) afin de constater si ce problème se pose aussi pour le PSID. Sans cette analyse, le fort pouvoir explicatif entre l'expérience et les variables dichtomiques d'année en déviation par rapport aux moyennes individuelles serait passé inaperçu. Inclure l'expérience à temps partiel avant le premier emploi à temps plein de même qu'une mesure plus précise de l'expérience auraient peut-être permis d'introduire suffisamment d'inégalité dans la variable endogène pour résoudre cette interaction néfaste.

Dans l'avenir, il faudra aussi tester ou relaxer l'hypothèse sur la forme de l'effet fixe de la procédure #2 ou à tout le moins évaluer la sensibilité des estimations à des postulats alternatifs. Ces recherches permettraient idéalement de réduire les écarts types, car l'évaluation de politiques nécessite des estimateurs précis et des instruments performants.

ANNEXE A

PRINCIPALES DISPOSITIONS CONCERNANT LES SERVICES DE GARDE ET L'ÉDUCATION PRÉSCOLAIRE AU CANADA

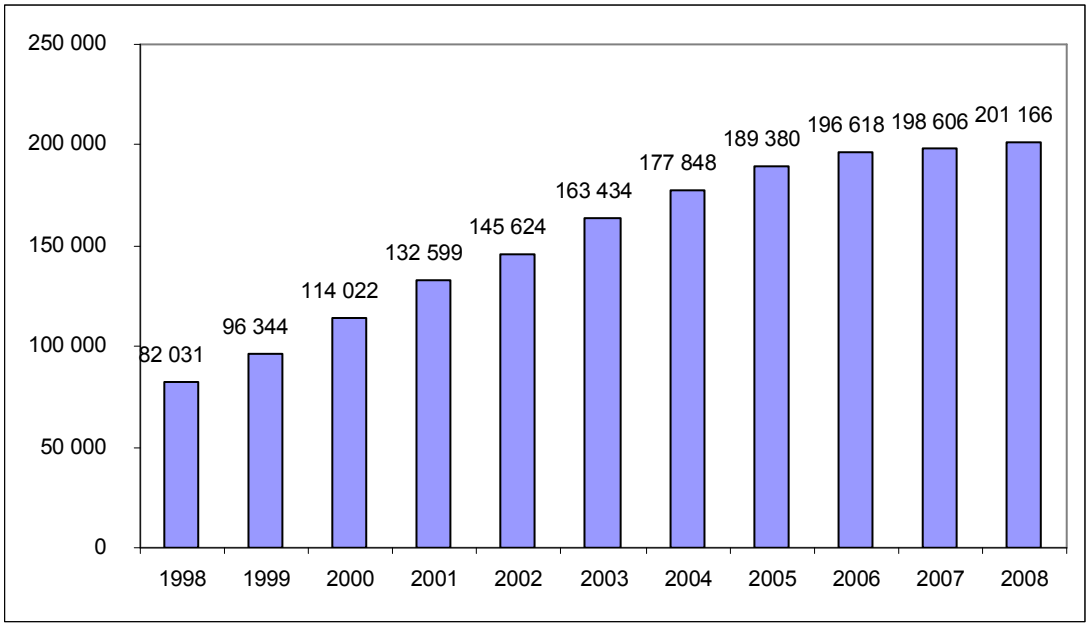
Tableau A.1 Introduction des services de garde au Québec.
(Tiré de Lefebvre et Merrigan, 2005a, p. 29.)

Québec
Services de garde
▪ Le 1ier septembre 1997, les milieux de garde reconnus par le ministère (garderies à but non lucratif et par la suite garderies privées et garderies en milieu familial) ont commencé à offrir des places à 5 $/jour aux enfants qui avaient atteint l'âge de 4 ans au 30 septembre. ▪ Le 1ier septembre 1998, admissibilité des 3 ans (âge atteint au 30 sept.). ▪ Le 1ier septembre 1999, admissibilité des 2 ans (âge atteint au 30 sept.). ▪ Le 1ier septembre 2000, admissibilité des 1 an et des moins de 1 an, (tous les enfants de moins de 5 ans et les enfants de 5 ans non admissibles à la maternelle).
Maternelle
▪ Pour les enfants de 5 ans (âge atteint au 30 septembre), la maternelle à plein temps (plutôt qu'à mi-temps) est offerte par toutes les commissions scolaires depuis septembre 1997. La fréquentation est optionnelle, mais si l'enfant fréquente la maternelle il doit être présent à temps plein. ▪ Dans quelques centres urbains, les commissions scolaires offrent une pré maternelle 4 ans à demi temps qui visent les enfants handicapés ou dont le(s) parent(s) est (sont) bénéficiaire(s) de l'aide sociale.
Garde avant/après l'école (5 ans et plus)
▪ À partir de septembre 1998, le ministère de l'Éducation subventionne les services de garde avant et après l'école (en milieu scolaire pour les enfants de 5 à 12 ans). Les commissions scolaires doivent offrir ces services en contrepartie d'une contribution de 5 $/jour lorsque la demande et le nombre d'élèves le justifient.

Tableau A.2 Services de garde à l'extérieur du Québec.
(Tiré de Lefebvre et Merrigan, 2005a, p. 29.)

Autres provinces
Services de garde
▪ Toutes les provinces ont un programme de subventions aux services de garde : celles-ci tiennent compte du revenu familial (la subvention est versée aux familles à revenu très modeste). ▪ Plusieurs provinces (CB, AL, MAN, ON, NB, IPE, TNL) ont un programme supplémentaire (du type *Enhanced Child Care Subsidy* ou *Employment Support*) afin de réduire le coût des services de garde pour les familles qui ont des revenus de travail faibles et rendre le « travail » plus « payant ».
Maternelle
▪ En Ontario la plupart des commissions scolaires offrent la pré maternelle (*junior kindergarten*), à demi temps (2h30 par jour), aux enfants de 4 ans et la plupart des enfants fréquentent cette pré maternelle publique qui est optionnelle. ▪ Toutes les autres provinces que le Québec offrent une maternelle publique à demi temps (2h30 par jour), sauf au Nouveau-Brunswick et en Nouvelle-Écosse où la maternelle est plutôt à temps plein (en Ontario, les commissions scolaires francophones offrent la maternelle à temps plein). La fréquentation de la maternelle est obligatoire en Colombie-Britannique, au Nouveau-Brunswick et en Nouvelle-Écosse. L'âge requis est de 5 ans (le mois d'anniversaire varie d'une province à l'autre).
Garde avant/après l'école (5 ans et plus)
▪ Quelques commissions scolaires offrent des services de garde à l'école, mais il n'y a pas d'initiatives généralisées de la part des ministères de l'Éducation provinciaux pour amener les écoles à offrir des services de garde aux élèves.

Figure A.1 Évolution du nombre de places en services de garde à contribution réduite au Québec.
(Tirée de Québec, 2004, p. 75; Québec, 2008.)



ANNEXE B

VARIANCE DE L'ESTIMATEUR DE LA RÉGRESSION INSTRUMENTALE EN DEUX ÉTAPES CORRIGEANT POUR LE BIAIS DE SÉLECTION

Selon Semykina et Wooldridge (2006, p. 30), la variance de $\left(\hat{\theta}-\theta\right)$ et donc de l'estimateur $\hat{\theta}$ tend asymptotiquement vers

$$\sqrt{N}\left(\hat{\theta}-\theta\right)\xrightarrow{a} Normal\left(0,\left(C'D^{-1}C\right)^{-1}C'D^{-1}BD^{-1}C\left(C'D^{-1}C\right)^{-1}\right)$$

où :

$$\hat{C}=N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}s_{it2}\hat{h}_{it}'\hat{w}_{it}\ ;$$

$$\hat{D}=N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T}s_{it2}\hat{h}_{it}'\hat{h}_{it}\ ;$$

$$\hat{B}=N^{-1}\sum_{i=1}^{N}\hat{p}_i\hat{p}_i'\ ;$$

$$p_i=\sum_{t=1}^{T}\left(s_{it2}\hat{h}_{it}'\hat{e}_{it1}-\hat{F}\hat{r}_i\right);$$

$$\hat{e}_{it1}=y_{it1}-\hat{w}_{it1}\hat{\theta}\,.$$

Les deux valeurs $\hat{F}$ et $\hat{r}_i$ nécessitent toutefois quelques explications. Définissons d'abord $q_{it}\equiv\left(1,z_{it},\bar{z}_i\right)$ un vecteur ligne de $1\times\left(1+2L\right)$ puis $\pi_t=\left(\eta_{t2},\delta_{t2}',\xi_{t2}'\right)$, c'est-

à-dire les coefficients du probit pour le temps t, et finalement le vecteur colonne $\pi = \left(\pi_1', \pi_2', ..., \pi_T'\right)'$.

D'une part,

$$\hat{F} = N^{-1} \sum_{i=1}^{N} \sum_{t=1}^{T} s_{it2} \hat{h}_{it}' \left(\hat{\theta} \nabla_{\pi} \hat{w}_{it}'\right).$$

Le vecteur des paramètres $\hat{\theta}'$ provient des estimations de la régression instrumentale en deux étapes. Ensuite, $\nabla_{\pi} \hat{w}_{it}'$ est une matrice $1 + K + L + T \times \left(1 + 2L\right) \cdot \left(T - 1\right) + 1$ remplie de 0 sauf à une exception près. Pour chaque couple (i,t), à la ligne $1 + K + L + t$ et à la colonne $\left(1 + 2L\right) \cdot \left(t - 1\right) + 1$, apparaît le résultat de l'expression $-q_{it} \lambda_{it2} \cdot \left(q_{it} \pi_t + \lambda_{it2}\right)$ ce qui donne visuellement :

$$\nabla_{\pi} w_{it}' = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & -q_{it} \lambda_{it2} \cdot \left(q_{it} \pi_t + \lambda_{it2}\right) & 0 \\ 0 & 0 & 0 & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}$$

D'autre part, $\hat{r}_i$ est un long vecteur colonne et se calcule ainsi :

$$\hat{r}_i = \left(\hat{r}_{i1}', \hat{r}_{i2}', ..., \hat{r}_{iT}'\right)'$$

où

$$\hat{r}_{it} = \hat{A}_t^{-1} \left\{\Phi\left(q_{it} \hat{\pi}_t\right)\left[1 - \Phi\left(q_{it} \hat{\pi}_t\right)\right]\right\}^{-1} \phi\left(q_{it} \hat{\pi}_t\right) q_{it}' \left[s_{it2} - \Phi\left(q_{it} \hat{\pi}_t\right)\right]$$

et

$$\hat{A}_t = N^{-1}\sum_{i=1}^{N}\{\Phi(q_{it}\hat{\pi}_t)[1-\Phi(q_{it}\hat{\pi}_t)]\}^{-1}[\phi(q_{it}\hat{\pi}_t)]^2 q'_{it}q_{it} \ .$$

En fait, le terme $\hat{A}_t$ représente la matrice de variance covariance du probit de la période t.

ANNEXE C

VARIANCE DE L'ESTIMATEUR DE LA RÉGRESSION CORRIGEANT POUR LE BIAIS DE SÉLECTION

Selon Wooldridge (1995b, p. 131-132), la variance de $\left(\hat{\theta}-\theta\right)$ et donc de l'estimateur $\hat{\theta}$ tend asymptotiquement vers

$$\sqrt{N}\left(\hat{\theta}-\theta\right)\xrightarrow{a} Normal\left(0, C^{-1}BC^{-1}\right)$$

où pour i = 1,…, N et t = 1,…,T :

$$\hat{C} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{w}'_{it}\hat{w}_{it} \;;$$

$$\hat{B} = N^{-1}\sum_{i=1}^{N}\hat{p}_i\hat{p}'_i \;;$$

$$\hat{p}_i = \hat{q}_i - \hat{F}\hat{r}_i \;;$$

$$\hat{q}_i = \sum_{t=1}^{T} s_{it}\hat{w}'_{it}\hat{e}_{it} \;.$$

Les deux valeurs $\hat{F}$ et $\hat{r}_i$ se calculent presque de la même façon que dans l'annexe B :

$$\hat{F} = N^{-1}\sum_{i=1}^{N}\sum_{t=1}^{T} s_{it2}\hat{w}'_{it}\left(\hat{\theta}'\nabla_{\pi}\hat{w}'_{it}\right);$$

$$\hat{r}_i = \left(\hat{r}'_{i1}, \hat{r}'_{i2}, ..., \hat{r}'_{iT}\right)'$$

où

$$\hat{r}_{it} = \hat{A}_t^{-1}\left\{\Phi\left(q_{it}\hat{\pi}_t\right)\left[1-\Phi\left(q_{it}\hat{\pi}_t\right)\right]\right\}^{-1}\phi\left(q_{it}\hat{\pi}_t\right)q'_{it}\left[s_{it2}-\Phi\left(q_{it}\hat{\pi}_t\right)\right]$$

et

$$\hat{A}_t = N^{-1}\sum_{i=1}^{N}\left\{\Phi\left(q_{it}\hat{\pi}_t\right)\left[1-\Phi\left(q_{it}\hat{\pi}_t\right)\right]\right\}^{-1}\left[\phi\left(q_{it}\hat{\pi}_t\right)\right]^2 q'_{it}q_{it} \ .$$

ANNEXE D

SPÉCIFICATION ET VARIABLES UTILISÉES

Tableau D.1 Spécification de l'échantillon (début).

Conditions	Ligne de programmation Stata
Femme	`sex99==2`
Âgée de 18 à 65 ans	`age26>=18 & age26<=65`
Soutien économique principal ou conjointe au sein de la famille de recensement	`rmjce26<=2`
La répondante n'est pas une immigrante.	`immst15==2`
La répondante est en couple.	`marst26<=2`
La principale source de revenus ne provient pas d'un emploi autonome.	`majri42~=3`
La répondante n'est pas une professionnelle de l'agriculture.	`nocg1s6~=46 & nocg1e6~=46`
La répondante n'est pas une éducatrice.	`noccds6~="E217" & noccde6~="E217"`
Province de résidence connue	`pvreg25<=59`
La répondante ne divulgue pas de valeurs aberrantes sur son âge, son salaire, ses heures de travail ou son éducation.	`age_diff>=6 & sal_nowork==0 & freework==0 & scol_weird_indicator==0`
Autres variables disponibles incluant l'information sur le conjoint et les enfants si la répondante est une mère.	`yrschl18>=0 & yrschl18<=99.5 & earng42~=99999996 7 ou 8 & earng46~=99999996 7 ou 8 & yrxfte11~=96 7 ou 8 & cmphrw28~=999.96 7 ou 8 & et fusion des données réussies`
Le salaire horaire (*cmphrw28*) corrigé pour l'inflation (*salaireipc*) est compris entre les centiles inférieur et supérieur de la distribution. Le revenu de travail familial autre que celui de la répondante utilise le deuxième centile. Le salaire horaire peut aussi être manquant ou nul à cette étape.	`keep if (salaireipc>=w1 & salaireipc<=w99 & rfam_autreipc>=f2 & rfam_autreipc<=f98) | (salaireipc==. & rfam_autreipc>=f2 & rfam_autreipc<=f98)`

Tableau D.1 Spécification de l'échantillon (suite).

Conditions	Ligne de programmation Stata
Pour considérer qu'une personne participe au marché du travail	
La répondante doit déclarer un salaire et une expérience de travail positive au sens de la variable *yrxfte11* (ou *exp*).	`g participation = (salaireipc~=. & exp>0 & exp<.)`
Condition supplémentaire pour les modèles à effets fixes	

Deux observations par personne au minimum	`quietly by id: gen t=_n`
	`egen maxt=max(t), by(id)`
	`drop if maxt==1`

▪ Les observations qui affichent « Code de traitement provisoire », « Ne sait pas » et « Refus » sont automatiquement exclues de l'échantillon.

Tableau D.2 Variables de l'EDTR utilisées pour les modèles.

	Nom	Variable de l'EDTR pertinente
Variable dépendante	Logarithme du salaire horaire réel	cmphrw28 série 326-0021 de CANSIM
Variables indépendantes	Expérience et son carré Éducation Variables dichotomiques de temps	yrxfte11 yrschl18 year
Instruments	Éducation Statut familial (mariée et conjointe de fait) Âge et son carré Revenus familiaux réels autres que ceux de la répondante et son carré Nombre d'enfants de 0 à 2 ans Nombre d'enfants de 3 à 5 ans Nombre d'enfants de 6 à 17 ans Âge du conjoint et son carré Éducation du conjoint et son carré Semaines de chômage Dichotomique si les semaines de chômage ne sont pas rapportées Variable dichotomique pour le Québec Dichotomique d'exposition à la politique Variables dichotomiques de temps	yrschl18 marst26 age26 (earng46-earng42) et série 326-0021 de CANSIM age26 et relag44 de l'enfant idem idem age26 du conjoint yrschl18 du conjoint wksuem28 du conjoint wksuem28 du conjoint pvreg25 age26 des enfants, year et pvreg25 year

ANNEXE E

PROGRAMME MATA

```
clear
clear mata

* Set working directory
local working_directory K:\
cd "`working_directory'"


* Set memory requirements
set mem 1m
set matsize 2000
set mem 300m
set matsize 4000
set more off

* Specify database
local database Semykina_merge_07.dta

* SET THE LOCALS
*Time identifier
local timevariable year

* Cross-section unit identifier
local ident id

* Dependant variable in the primary equation
local depvar lsalaireipc

* Explanatory variables in the primary equation instrumented
local indep_ins exp exp2
* Explanatory variables in the primary equation not instrumented
local indep_noins educ

* Explanatory variables in the selection equation
local probits age age2 mariee rfam_autreipc rfam_autreipc2 enfants_0_2 enfants_3_5
enfants_6_17 age_h age_h2 educ_h educ_h2 sem_chom_h no_report_chom qc
* Explanatory variables in the selection equation without mean
local probitsnomean


* Instruments used at the second step with their mean
local instrumeans age age2 mariee rfam_autreipc rfam_autreipc2 enfants_0_2 enfants_3_5
enfants_6_17 age_h age_h2 educ_h educ_h2 sem_chom_h no_report_chom qc
* Instruments without mean
local instrunomean
```

```
* Note: in general, there may be fewer instruments than variables in the selection
equation, but it must be the case that the selection equation constains all the
variables from the instruments.

* Policies
local policies garde_0_4

* Variables with missing values
local missing participation



**************************************
**************************************
**************************************
* If everything above is set properly,*
* the program should run smoothly.    *
**************************************
**************************************
**************************************
cd "`working_directory'"

use "`database'", clear

sort `ident' `timevariable'

qui: sum `timevariable'
local minT `r(min)'
local maxT `r(max)'

qui: tab `timevariable'
scalar T = r(r)
mata
T = st_numscalar("T")
end

local t = 0
while `t' <= `maxT'-`minT' {
       local var = `minT'+`t'
       local time2 `time2' `var'
       local t = `t'+1
}

local t = 1
while `t' <= `maxT'-`minT' {
       local var = `minT'+`t'
       local time3 `time3' `var'
       local t = `t'+1
}

foreach var of local time2 {
       local time `time' y_`var'
}

foreach var of local time3 {
       local time4 `time4' y_`var'
}

mata: TT=T

qui: tab `timevariable', gen(y)

local j = 1
foreach var of local time {
```

```
	rename y`j' `var'
	local j = `j'+1
}

local instrumeans `instrumeans' `policies'

local regprobit y `indep_noins' `probits' `probitsnomean'

foreach var of local indep_noins {
local moy_indep_noins `moy_indep_noins' m`var'
}

foreach var of local indep_noins {
	egen m`var' = mean(`var'), by(`ident')
}

foreach var of local probits {
	local moy_probit `moy_probit' m`var'
}
local moy_probit `moy_indep_noins' `moy_probit'

foreach var of local instrumeans {
	local moy_instru `moy_instru' m`var'
}
local moy_instru `moy_indep_noins' `moy_instru'

local regprobit `regprobit' `moy_probit'

foreach var of local regprobit {
	foreach t of local time2 {
		local regprobit1 `regprobit1' `var'_`t'
	}
}
foreach var of local regprobit {
	foreach t of local time2 {
		local regprobit`t' `regprobit`t'' `var'_`t'
	}
}
foreach t of local time2 {
	local regprobit2 `regprobit2' `regprobit`t''
}

*********************************************
********* PROCEDURE 4.1 SEMYKINA P. 17 *******
* GENERATE PROBIT VARIABLES
gen cons=1
sort `ident' `timevariable'

foreach var of local probits {
	egen m`var' = mean(`var'), by(`ident')
}

foreach var of local policies {
	egen m`var' = mean(`var'), by(`ident')
}

local generation `indep_noins' `probits' `probitsnomean' `moy_probit'
foreach var of local generation {
	foreach t of local time2 {
		gen `var'_`t' = `var'*y_`t'
	}
}


* FIND LAMBDA (SEMYKINA P. 17 ET P. 20)
```

```
* LOOK FOR PI FOR A GIVEN T
tsset `ident' `timevariable'

probit `missing' `regprobit1', nocons
mat bprobit = e(b)
mat cov_search = e(V)

////////////// GENERATE COVARIANCE PROBIT /////////////////
probit `missing' `regprobit2', nocons
mat bprobit2 = e(b)
mat cov_search2 = e(V)
mata
bprobit2 = st_matrix ("bprobit2")
cov_search2 = st_matrix("cov_search2")
////////////////////////////////////////////////////////////
end
////////////////// GENERATE LAMBDA ///////////////////////
predict prob_hat
predict xbhat, xb
gen lambda = normalden(xbhat)/prob_hat
drop prob_hat xbhat

foreach var of local time3 {
       gen lambday`var' = lambda*y_`var'
}

local lambda lambda
foreach var of local time3 {
       local lambda `lambda' lambday`var'
}


mata
//////////////////// GENERATE At //////////////////////
// GENERATE PROBIT MATRIX TO COMPUTE MATRIX PI
bprobit = st_matrix ("bprobit")
P=colshape(bprobit,T)

// GENERATE MATRICES FOR qit
end
local varlist cons `indep_noins' `probits' `probitsnomean' `moy_probit'


mata
st_view(QQ=., ., tokens("`varlist'"))

// CORRECTION OF QQ FOR NON OBSERVED t
st_view(G1=., ., tokens("`ident'"))
st_view(G2=., ., tokens("`time'"))
shape = panelsetup(G1, 1, 1)

QQ0=J(T*rows(shape),cols(QQ), .)
for (g=1; g<=cols(QQ); ++g) {
       QQd=QQ[.,g]
       res_QQg = J(T, rows(shape), .)
              for (i=1; i<=rows(shape); ++i) {
                     panelsubview(QQd_=., QQd, i, shape)
                     panelsubview(images_=., G2, i, shape)
                     QQg = images_'*QQd_
                     res_QQg[.,i]=QQg
              }
QQg=vec(res_QQg)
QQ0[.,g]=QQg
}
////////////////////////////////////////////////
```

```
end

foreach var of local missing {
       drop if `var'==0
}

* CLEAN DATABASE
local bigdrop `indep_noins' `probits'
foreach var of local bigdrop {
       local mbigdrop `mbigdrop' m`var'
}
local bigdrop `bigdrop' `mbigdrop' `probitsnomean'
foreach t of local time2 {
       foreach var of local bigdrop {
              local ybigdrop `ybigdrop' `var'_`t'
       }
}
drop `ybigdrop'

save "dummy_07.dta", replace
**************************************
*************** FE-2SLS ***************
**************************************

drop if `depvar'==.


* KEEP ONLY 2 OR MORE OBSERVATIONS PER CROSS-SECTION UNIT
quietly by `ident': gen t=_n
egen maxt=max(t), by(`ident')
drop if maxt==1
drop maxt t

sort `ident' `timevariable'

local fe2sls cons `depvar' `indep_ins' `indep_noins' `instrumeans' `instrunomean'
`time4' `lambda'
foreach var of local fe2sls {
       egen mm`var' = mean(`var'), by(`ident')
       egen mmt`var' = mean(`var')
       gen d`var' = `var' - mm`var'
}
drop mm* mmt*

local ddepvar d`depvar'
local indepvar `indep_ins' `indep_noins'
foreach var of local indepvar {
       local dindepvar `dindepvar' d`var'
}
local instru `indep_noins' `instrumeans' `instrunomean'
foreach var of local instru {
       local dinstru `dinstru' d`var'
}
foreach var of local indep_noins {
       local dindep_noins `dindep_noins' d`var'
}
foreach var of local indep_ins {
       local dindep_ins `dindep_ins' d`var'
}
foreach var of local instrumeans {
       local dinstrumeans `dinstrumeans' d`var'
}
foreach var of local instrunomean {
       local dinstrunomean `dinstrunomean' d`var'
}
```

```
foreach var of local policies {
	local dpolicies `dpolicies' d`var'
}
foreach var of local time4 {
	local dtime4 `dtime4' d`var'
}
foreach var of local lambda {
	local dlambda `dlambda' d`var'
}

local dinstru `dinstru' `dtime4' `dlambda'
local dindepvar `dindepvar' `dtime4' `dlambda'

mata
st_view(y=., ., tokens("`ddepvar'"))
st_view(X=., ., tokens("`dindepvar'"))
st_view(Z=., ., tokens("`dinstru'"))


K= cols(X)-T-1
L= cols(Z)-T

b = invsym(X'X)*X'y
e=y-X*b

b_iv = invsym((X'Z)*invsym(Z'Z)*(Z'X))*(X'Z)*invsym(Z'Z)*(Z'y)
e_iv=y-X*b_iv

st_view(V1=., ., "`ident'")
st_view(V2=., ., tokens("`time'"))
info = panelsetup(V1, 1, 2)
N = rows(info)
res_u = J(T, rows(info), .)
res_uz = J(T, rows(info), .)


// GENERATE MATRIX U FOR COVARIANCE
for (i=1; i<=rows(info); ++i) {
	panelsubview(e_=., e, i, info)
	panelsubview(images_=., V2, i, info)
	u = images_'*e_
	res_u[.,i]=u
}
u = vec(res_u)

for (i=1; i<=rows(info); ++i) {
	panelsubview(e_iv_=., e_iv, i, info)
	panelsubview(images_=., V2, i, info)
	uz = images_'*e_iv_
	res_uz[.,i]=uz
}
uz = vec(res_uz)


// GENERATE MATRIX Y TO VERIFY
res_y2 = J(T, rows(info), .)
for (i=1; i<=rows(info); ++i) {
	panelsubview(y2_=., y, i, info)
	panelsubview(images_=., V2, i, info)
	y2 = images_'*y2_
	res_y2[.,i]=y2
}
y2 = vec(res_y2)
```

```
// GENERATE MATRIX X FOR COVARIANCE
X0=J(T*rows(info),cols(X), .)

for (g=1; g<=cols(X); ++g) {
       Xd=X[.,g]
       res_Xg= J(T, rows(info), .)
              for (i=1; i<=rows(info); ++i) {
                     panelsubview(Xd_=., Xd, i, info)
                     panelsubview(images_=., V2, i, info)
                     Xg = images_'*Xd_
                     res_Xg[.,i]=Xg
                     }
Xg=vec(res_Xg)
X0[.,g]=Xg
}

// GENERATE MATRIX Z for COVARIANCE
Z0=J(T*rows(info),cols(Z), .)

for (g=1; g<=cols(Z); ++g) {
       Zd=Z[.,g];
       res_Zg= J(T, rows(info), .)
              for (i=1; i<=rows(info); ++i) {
                     panelsubview(Zd_=., Zd, i, info)
                     panelsubview(images_=., V2, i, info)
                     Zg = images_'*Zd_
                     res_Zg[.,i]=Zg
                     }
Zg=vec(res_Zg)
Z0[.,g]=Zg
}



// COVARIANCE FE TO VERIFY
E=J(cols(X),cols(X),0)
for (i=1; i<=rows(info); ++i) {
       Xi=J(T,cols(X),.)
       ui=J(T,1,.)
       for(l=T-1; l>=0; --l) {
                     Xi[T-l,.]=X0[T*i-l,.]
                     ui[T-l,.]=u[T*i-l,.]
       }
       Ei = Xi'*ui*ui'*Xi
       E=E+Ei
}

N1 = rows(y)
ka = cols(X)
M = rows(info)
// Correction cluster http://www.stata.com/support/faqs/stat/robust.html
corr = ((N1-1)*M)/((N1-ka)*(M-1))
E=E*corr

// CALCULATE COV WOOLDRIDGE 2002 P. 275 (10.59)
cov = invsym(X'*X)*E*invsym(X'*X)
se = sqrt(diagonal(cov))

// CALCULATE COVARIANCE WOOLDRIDGE 2002 P. 275 (10.59)
// CALCULATE SUM FIRST
E=J(cols(Z0),cols(Z0),0)
for (i=1; i<=rows(info); ++i) {
       Zi=J(T,cols(Z0),.)
       ui=J(T,1,.)
```

```
        for(l=T-1; l>=0; --l) {
                        Zi[T-l,.]=Z0[T*i-l,.]
                        ui[T-l,.]=uz[T*i-l,.]
        }
        Ei = Zi'*ui*ui'*Zi
        E=E+Ei
}


// http://www.stata.com/support/faqs/stat/robust.html
// CORRECTION STATA [XT] p. 150
N1 = rows(y)
ka = cols(X)
M = rows(info)
L2=cols(Z)
corr = ((N1-1)*M)/((M-1)*(N1-ka))

E=E*corr

// CALCULATE COV
// cov robust cluster like stata
covz =
invsym(X'Z*invsym(Z'Z)*Z'X)*X'Z*invsym(Z'Z)*E*invsym(Z'Z)*Z'X*invsym(X'Z*invsym(Z'Z)*Z
'X)

// cov wooldrige p. 275 (10.59) doesn't fit stata
covz_wooldridge = invsym((X0'Z0)*invsym(E)*(Z0'X0))


// cov no robust no cluster
covz2 = (e_iv'*e_iv)/(N1-ka)*invsym((X0'Z0)*invsym(Z0'Z0)*(Z0'X0))

sez = sqrt(diagonal(covz))
sez2 = sqrt(diagonal(covz2))
end


tsset `ident' `timevariable'
* FIXED EFFECTS
xtreg `depvar' `indepvar' `time4', fe robust cluster(`ident')
* RANDOM EFFECTS
xtreg `depvar' `indepvar' `time4', re robust cluster(`ident')


******* Autocorrelation test *******
preserve
sort `ident' `timevariable'

* KEEP ONLY 3 OR MORE OBSERVATIONS PER CROSS-SECTION UNIT
quietly by `ident': gen t=_n
egen maxt=max(t), by(`ident')
keep if maxt>=3
qui: sum maxt if t==1
scalar Tmean = r(mean)

predict double resid, e
by `ident': g resid_lag=resid[_n-1]

* For each T
forvalues x = 3/6 {
display "T=`x'"
reg resid resid_lag if maxt==`x', robust cluster(`ident') noconstant
test resid_lag = -1/(`x'-1)
}
```

```
* For all T
reg resid resid_lag, robust cluster(`ident') noconstant
test resid_lag = -1/(3-1)
test resid_lag = -1/(4-1)
test resid_lag = -1/(5-1)
test resid_lag = -1/(6-1)
scalar list Tmean
test resid_lag = -1/(Tmean-1)
drop resid resid_lag maxt t
restore
**********************************

* Heteroskedasticity test
reg `ddepvar' `dindep_ins' `dindep_noins' `dtime4'
estat hettest

************** COMPARE FE WITH CLUSTER AND ROBUST *****************
******************** SE SHOULD BE IDENTICAL ***********************
xtreg `depvar' `indepvar' `time4' `lambda', fe robust cluster(`ident')
reg `ddepvar' `dindep_ins' `dindep_noins' `dtime4' `dlambda', robust cluster(`ident')
mata
(b, se, b:/se, 2*ttail(N1-ka-M+1, abs(b:/se)))
end

* FE-2SLS
xtivreg `depvar' `indep_noins' `time4' (`indep_ins' = `instrumeans' `instrunomean'
`time4'), re first
xtivreg `depvar' `indep_noins' `time4' (`indep_ins' = `instrumeans' `instrunomean'
`time4'), fe first

* FE-2SLS
ivregress 2sls `ddepvar' `dindep_noins' `dtime4' (`dindep_ins' = `dinstrumeans'
`dinstrunomean' `dtime4'), first
estat firststage, all forcenonrobust
estat overid

******* Autocorrelation test *******
preserve
sort `ident' `timevariable'

* KEEP ONLY 3 OR MORE OBSERVATIONS PER CROSS-SECTION UNIT
quietly by `ident': gen t=_n
egen maxt=max(t), by(`ident')
keep if maxt>=3
qui: sum maxt if t==1
scalar Tmean = r(mean)

predict double resid, residuals
by `ident': g resid_lag=resid[_n-1]

* For each T
forvalues x = 3/6 {
display "T=`x'"
reg resid resid_lag if maxt==`x', robust cluster(`ident') noconstant
test resid_lag = -1/(`x'-1)
}

* For all T
reg resid resid_lag, robust cluster(`ident') noconstant
test resid_lag = -1/(3-1)
test resid_lag = -1/(4-1)
test resid_lag = -1/(5-1)
test resid_lag = -1/(6-1)
scalar list Tmean
test resid_lag = -1/(Tmean-1)
```

```
drop resid resid_lag maxt t
restore
**********************************


* FE-2SLS robust
ivregress 2sls `ddepvar' `dindep_noins' `dtime4' (`dindep_ins' = `dinstrumeans'
`dinstrunomean' `dtime4'), robust first
estat firststage, all forcenonrobust
estat overid

* FE-2SLS robust cluster
ivregress 2sls `ddepvar' `dindep_noins' `dtime4' (`dindep_ins' = `dinstrumeans'
`dinstrunomean' `dtime4'), robust cluster(`ident') first
estat firststage, all forcenonrobust
estat overid

* J-TEST
ivreg `depvar' `indep_noins' `time4' (`indep_ins' = `instrumeans' `instrunomean'
`time4'), robust cluster(`ident') nocons
predict e, resid

reg e `dindep_noins' `dinstrumeans' `dinstrunomean' `time4', robust cluster(`ident')
nocons
test `dindep_noins' `dinstrumeans' `dinstrunomean' `time4'


local Zm 0
local instruZm `dinstrumeans'
foreach var of local instruZm {
       local Zm = `Zm'+1
}
scalar Zm = "`Zm'"
local Xk 0
local indepXk `dindep_ins'
foreach var of local indepXk {
       local Xk = `Xk'+1
}
scalar Xk = "`Xk'"
mata
Fstat = st_numscalar("r(F)")
Zm = st_strscalar("Zm")
Zm = `Zm'
Xk = st_strscalar("Xk")
Xk = `Xk'
Zm
Xk
degchi = Zm-Xk
J=Zm*Fstat
Probchi2 = chi2tail(degchi,J)
J
Probchi2


end
******* COMPARE FE-2SLS WITHOUT CLUSTER AND NON-ROBUST *************
*********** SE A LITTLE BIT LOWER AS EXPECTED **********************
* xtivreg no robust no cluster
xtivreg `depvar' `indep_noins' `time4' `lambda' (`indep_ins' = `instrumeans'
`instrunomean' `time4' `lambda'), re first
xtivreg `depvar' `indep_noins' `time4' `lambda' (`indep_ins' = `instrumeans'
`instrunomean' `time4' `lambda'), fe first
test `time4'
test `lambda'
```

```
* Procedure 1 no robust no cluster
ivregress 2sls `ddepvar' `dindep_noins' `dtime4' `dlambda' (`dindep_ins' =
`dinstrumeans' `dinstrunomean' `dtime4' `dlambda'), first
estat firststage, all
estat overid
test `dtime4'
test `dlambda'

******* Autocorrelation test *******
preserve
sort `ident' `timevariable'

* KEEP ONLY 3 OR MORE OBSERVATIONS PER CROSS-SECTION UNIT
quietly by `ident': gen t=_n
egen maxt=max(t), by(`ident')
keep if maxt>=3
qui: sum maxt if t==1
scalar Tmean = r(mean)

predict double resid, residuals
by `ident': g resid_lag=resid[_n-1]

* For each T
forvalues x = 3/6 {
display "T=`x'"
reg resid resid_lag if maxt==`x', robust cluster(`ident') noconstant
test resid_lag = -1/(`x'-1)
}

* For all T
reg resid resid_lag, robust cluster(`ident') noconstant
test resid_lag = -1/(3-1)
test resid_lag = -1/(4-1)
test resid_lag = -1/(5-1)
test resid_lag = -1/(6-1)
scalar list Tmean
test resid_lag = -1/(Tmean-1)
drop resid resid_lag maxt t
restore
**********************************



* Procedure 1 robust no cluster
ivregress 2sls `ddepvar' `dindep_noins' `dtime4' `dlambda' (`dindep_ins' =
`dinstrumeans' `dinstrunomean' `dtime4' `dlambda'), robust first
estat firststage, all forcenonrobust
estat overid
test `dtime4'
test `dlambda'

* Procedure 1 robust cluster
ivregress 2sls `ddepvar' `dindep_noins' `dtime4' `dlambda' (`dindep_ins' =
`dinstrumeans' `dinstrunomean' `dtime4' `dlambda'), robust cluster(`ident') first
estat firststage, all forcenonrobust
estat overid
test `dtime4'
test `dlambda'


mata
// XTIVREG NO ROBUST NO CLUSTER
// SHOULD BE IDENTICAL, THIS IS JUST A CHECK
(b_iv, sez2, b_iv:/sez2, 2*ttail(N1-ka, abs(b_iv:/sez2)))
end
```

```
*********************************************************************
mata
// XTIVREG ROBUST CLUSTER
// SHOULD BE IDENTICAL, THIS IS JUST A CHECK
(b_iv, sez, b_iv:/sez, 2*ttail(N1-ka, abs(b_iv:/sez)))

///////////////// WALD TEST SELECTION BIAS ///////////////////////
// [R] R-Z p. 416

b=b_iv[|rows(b_iv)-T+1, 1\rows(b_iv),1|]
V=covz[|rows(covz)-T+1, cols(covz)-T+1\rows(covz),cols(covz)|]
R=I(T)
r=J(T,1,0)

W=(R*b-r)'*invsym(R*V*R')*(R*b-r)
Probchi2 = chi2tail(T,W)

W
Probchi2
/////////////////////////////////////////////////////////////
end


use "dummy_07.dta", clear


* FE-2SLS
* Test biais de sélection intertemporel (Semykina p. 12)

* KEEP ONLY 2 OR MORE OBSERVATIONS PER CROSS-SECTION UNIT
quietly by `ident': gen t=_n
egen maxt=max(t), by(`ident')
drop if maxt==1
drop maxt t

sort `ident' `timevariable'

by `ident': g sit=(`depvar'~=.)
by `ident': g lag1=sit[_n-1]
by `ident': g lag2=sit[_n+1]
by `ident': g lag3=sum(sit[_n-1])

gsort `ident' -`timevariable'
by `ident': g lag4=sum(sit[_n-1])
sort `ident' `timevariable'

local fe2sls cons `depvar' `indep_ins' `indep_noins' `instrumeans' `instrunomean'
`time4' `lambda' lag1 lag2 lag3 lag4
foreach var of local fe2sls {
       egen mm`var' = mean(`var'), by(`ident')
       egen mmt`var' = mean(`var')
       gen d`var' = `var' - mm`var'
}

xtivreg `depvar' `indep_noins' `time4' (`indep_ins' = `instrumeans' `instrunomean'
`time4'), fe
xtivreg `depvar' `indep_noins' `time4' lag1 (`indep_ins' = `instrumeans'
`instrunomean' `time4' lag1), fe
xtivreg `depvar' `indep_noins' `time4' lag2 (`indep_ins' = `instrumeans'
`instrunomean' `time4' lag2), fe
xtivreg `depvar' `indep_noins' `time4' lag3 (`indep_ins' = `instrumeans'
`instrunomean' `time4' lag3), fe
```

```
xtivreg `depvar' `indep_noins' `time4' lag4 (`indep_ins' = `instrumeans'
`instrunomean' `time4' lag4), fe

* CORRECTION TIME DUMMIES
ivreg `ddepvar' `dindep_noins' `dtime4' dlag1 (`dindep_ins' = `dinstrumeans'
`dinstrunomean' `dtime4' dlag1), robust cluster(`ident')
ivreg `ddepvar' `dindep_noins' `dtime4' dlag2 (`dindep_ins' = `dinstrumeans'
`dinstrunomean' `dtime4' dlag2), robust cluster(`ident')
ivreg `ddepvar' `dindep_noins' `dtime4' dlag3 (`dindep_ins' = `dinstrumeans'
`dinstrunomean' `dtime4' dlag3), robust cluster(`ident')
ivreg `ddepvar' `dindep_noins' `dtime4' dlag4 (`dindep_ins' = `dinstrumeans'
`dinstrunomean' `dtime4' dlag4), robust cluster(`ident')


use "dummy_07.dta", clear

drop if `depvar'==.


mata
////////////////////////////////////////////////////////////////
////////////////////////////////////////////////////////////////
////////////// POOLED S2LS SEMYKINA P. 28 (30) ////////////////////
//////////////REVIEW MATRICES FOR POOLED 2SLS ////////////////////
// GENERATE MATRICES FOR COMPUTATION OF O
end
local indepvar cons `indep_ins' `indep_noins' `lambda'
local instru `indep_noins' `instrumeans' `instrunomean' `lambda'
local instru_ `moy_instru' `time4'

mata
st_view(X=., ., tokens("`indepvar'"))
st_view(Z=., ., tokens("`instru'"))
st_view(Z_=., ., tokens("`instru_'"))
st_view(y=., ., tokens("`depvar'"))

// GENERATE MATRIX W
W=J(rows(X), cols(X)+cols(Z_), .)

for (i=1;i<=cols(X)-T;++i) {
       W[.,i]=X[.,i]
}
for (i=1;i<=cols(Z_); ++i) {
       W[.,cols(X)-T+i]=Z_[.,i]
}
for (i=1;i<=T;++i) {
       W[.,cols(X)-T+cols(Z_)+i]=X[.,cols(X)-T+i]
}

// GENERATE MATRIX H
H=J(rows(X), cols(Z)+cols(Z_)+1, .)

H[.,1]=X[.,1]

for (i=1;i<=cols(Z)-T;++i) {
       H[.,i+1]=Z[.,i]
}

for (i=1;i<=cols(Z_); ++i) {
       H[.,1+cols(Z)-T+i]=Z_[.,i]
}
for (i=1;i<=T;++i) {
       H[.,1+cols(Z)-T+cols(Z_)+i]=X[.,cols(X)-T+i]
}
```

```
////////////////// POOLED 2SLS ESTIMATOR ////////////////////
O = invsym((W'H)*invsym(H'H)*(H'W))*(W'H)*invsym(H'H)*(H'y)
end
use "dummy_07.dta", clear

qui: ivreg `depvar' `indep_noins' `moy_instru' `time4' `lambda' (`indep_ins' =
`instrumeans' `moy_instru' `time4' `lambda'), robust cluster(`ident')
scalar clust = e(N_clust)

* Obtain residuals even if lwage not available
predict e, resid
replace e=0 if `depvar'==.

mata
st_view(e=., ., "e")


///// CORRECTION STATA 9 BOOK [R] P. 54 FOR CLUSTER AND DEGREES OF FREEDOM
/////////////////
// http://www.stata.com/support/faqs/stat/robust.html
st_view(VV1=., ., "`ident'")
info2 = panelsetup(VV1, 1, 1)
st_view(VV2=., ., tokens("`time'"))


N = rows(y)
ka = cols(X)-TT-1
M = st_numscalar("e(N_clust)")
L2=cols(Z)-TT
corr = ((N-1)*M)/((M-1)*(N-ka-2*TT-L2))

end
* reajust identity matrix
recode `depvar' (.=0)
g selection=(`depvar'~=0)

mata
// reajust identity matrix
st_view(selection=., ., ("selection"))
st_view(V1=., ., "`ident'")
st_view(V2=., ., tokens("`time'"))
info = panelsetup(V1, 1, 1)

end
local varlist cons `indep_noins' `probits' `probitsnomean' `moy_probit'

mata

////////// ADD ZEROS WITHIN MATRICES FOR COV ////////////
st_view(X=., ., tokens("`indepvar'"))
st_view(Z=., ., tokens("`instru'"))
st_view(Z_=., ., tokens("`instru_'"))
st_view(y=., ., tokens("`depvar'"))
st_view(Q=., ., tokens("`varlist'"))
st_view(lambda_v=., ., "lambda")

// GENERATE MATRIX W
WW=J(rows(X), cols(X)+cols(Z_), .)
for (i=1;i<=cols(X)-T;++i) {
        WW[.,i]=X[.,i]
}
for (i=1;i<=cols(Z_); ++i) {
        WW[.,cols(X)-T+i]=Z_[.,i]
}
for (i=1;i<=T;++i) {
```

```
	WW[.,cols(X)-T+cols(Z_)+i]=X[.,cols(X)-T+i]
}

// GENERATE MATRIX H
HH=J(rows(X), cols(Z)+cols(Z_)+1, .)
HH[.,1]=X[.,1]
for (i=1;i<=cols(Z)-T;++i) {
	HH[.,i+1]=Z[.,i]
}
for (i=1;i<=cols(Z_); ++i) {
	HH[.,1+cols(Z)-T+i]=Z_[.,i]
}
for (i=1;i<=T;++i) {
	HH[.,1+cols(Z)-T+cols(Z_)+i]=X[.,cols(X)-T+i]
}

// CORRECTION OF MATRIX W FOR COV
W0=J(T*rows(info),cols(W), .)
for (g=1; g<=cols(W); ++g) {
	Wd=WW[.,g];
	res_Wg= J(T, rows(info), .)
		for (i=1; i<=rows(info); ++i) {
			panelsubview(Wd_=., Wd, i, info)
			panelsubview(images_=., V2, i, info)
			Wg = images_'*Wd_
			res_Wg[.,i]=Wg
			}
Wg=vec(res_Wg)
W0[.,g]=Wg
}

// CORRECTION OF MATRIX H FOR COV
H0=J(T*rows(info),cols(H), .)
for (g=1; g<=cols(H); ++g) {
	Hd=HH[.,g];
	res_Hg= J(T, rows(info), .)
		for (i=1; i<=rows(info); ++i) {
			panelsubview(Hd_=., Hd, i, info)
			panelsubview(images_=., V2, i, info)
			Hg = images_'*Hd_
			res_Hg[.,i]=Hg
			}
Hg=vec(res_Hg)
H0[.,g]=Hg
}

// CORRECTION OF MATRIX e FOR COV
res_u = J(T, rows(info), .)

for (i=1; i<=rows(info); ++i) {
	panelsubview(e_=., e, i, info)
	panelsubview(images_=., V2, i, info)
	u = images_'*e_
	res_u[.,i]=u
}
e0 = vec(res_u)

// CORRECTION OF MATRIX Q FOR COV
Q0=J(T*rows(info),cols(Q), .)

for (g=1; g<=cols(Q); ++g) {
	Qd=Q[.,g];
	res_Qg= J(T, rows(info), .)
		for (i=1; i<=rows(info); ++i) {
			panelsubview(Qd_=., Qd, i, info)
```

```
                        panelsubview(images_=., V2, i, info)
                        Qg = images_'*Qd_
                        res_Qg[.,i]=Qg
                        }
Qg=vec(res_Qg)
Q0[.,g]=Qg
}

// CORRECTION OF LAMBDA_V FOR COV
res_lambda = J(T, rows(info), .)

for (i=1; i<=rows(info); ++i) {
       panelsubview(lambda_=., lambda_v, i, info)
       panelsubview(images_=., VV2, i, info)
       lambda = images_'*lambda_
       res_lambda[.,i]=lambda
}
lambda_v0 = vec(res_lambda)

// CORRECTION OF SELECTION INDICATOR FOR COV
res_selection = J(T, rows(info), .)
for (i=1; i<=rows(info); ++i) {
       panelsubview(selection_=., selection, i, info)
       panelsubview(images_=., V2, i, info)
       selec = images_'*selection_
       res_selection[.,i]=selec
}
selection0 = vec(res_selection)

//// VERIFICATION OF COVARIANCE ///////
// COMPUTE SUM FIRST
E=J(cols(H0),cols(H0),0)
for (i=1; i<=rows(info); ++i) {
       Hi=J(T,cols(H0),.)
       ei=J(T,1,.)
       for(l=T-1; l>=0; --l) {
                        sit2=selection0[T*i-l,1]
                        Hi[T-l,.]=sit2*H0[T*i-l,.]
                        ei[T-l,.]=sit2*e0[T*i-l,.]
       }
       Ei = Hi'*ei*ei'*Hi
       E=E+Ei
}
E=E*corr

// COMPUTE COVARIANCE
cov_hat =
invsym(W'H*invsym(H'H)*H'W)*W'H*invsym(H'H)*E*invsym(H'H)*H'W*invsym(W'H*invsym(H'H)*H
'W)
se_hat = sqrt(diagonal(cov_hat))
end
use "dummy_07.dta", clear

************ MATRICES SHOULD BE EQUIVALENT ************
**************** THIS IS JUST A CHECK *****************
* ivreg `depvar' `indep_noins' `moy_instru' `time4' `lambda' (`indep_ins' =
`indep_noins' `instrumeans' `moy_instru' `time4' `lambda'), robust cluster(`ident')
* mat list e(V)
mata
// cov_hat
//////////////////////////////////////////////////////

//////////////////////////////////////////////////////
/// GENERATE MATRICES C AND D SEMYKINA P. 30 (37) and (38)
/// MISTAKE SEMYKINA P. 30 (37)
```

```
/// SHOULD READ C=[sum(sit2*hit'*wit)]/N
/// NOT C=[sum(sit2*wit'*hit)]/N

C=(H'W)
D=(H'H)

/// GENERATE GAMMA ///
gamma=J(T,1,0)
gam1=O[|1,1\rows(O)-T,1|]
gam2=O[|rows(O)-T+1,1\rows(O),1|]
for (t=2; t<=T; ++t) {
        gamma[t,1]=gam2[1,1]
}
gamma=gam2+gamma
gamma=gam1\gamma
////////////////////

// GENERATE F SEMYKINA P. 30 AT THE BOTTOM ///
F=J(cols(H0),cols(Q)*T,0)
for (i=1;i<=rows(info);++i) {
        for(t=1;t<=T;++t) {
                        sit2=selection0[(t+(i-1)*T),1]
                        qit=Q0[(t+(i-1)*T),.]
                        lambda_it2=lambda_v0[(t+(i-1)*T),.]
                        Pt=P[.,t]
                        hit=H0[(t+(i-1)*T),.]

                        Jacob_it=J(cols(W), cols(Q)*T, 0)
                        bertha=-1*qit*lambda_it2*((qit*Pt)+lambda_it2)
                                for (c=1; c<=cols(bertha);++c) {
                                        Jacob_it[cols(W)-T+t, ((cols(Q)*(t-
1))+c)]=bertha[., c]
                                }
Fi=sit2*hit'*gamma'*Jacob_it
F=F+Fi
        }
}
////////////////////////////////////////////////

///////////////// VALUES OF At ////////////////////
t=1
inva=cov_search2[|(t-1)*cols(Q)+1, (t-1)*cols(Q)+1 \ t*cols(Q),t*cols(Q)|]

for (t=2;t<=T; ++t) {
inva2=cov_search2[|(t-1)*cols(Q)+1, (t-1)*cols(Q)+1 \ t*cols(Q),t*cols(Q)|]
inva=inva\inva2
}

cov_probit = cov_search2
At=inva

////////////////////////////////////////////
// At should be identical to cov_probit but
// without zeros.
// cov_probit
// At
////////////////////////////////////////////

////////////////////////////////////////////////
////////////// DEMONSTRATION ri  FOR i=1 ////////////
/// THIS IS JUST A CHECK BETWEEN THE COV MATRIX ///
///// OF THE PROBIT AND SEMYKINA'S CALCULATION ////
////////////////////////////////////////////////
i=1
        ri=J(cols(Q0), T,0)
```

```
        for (t=1;t<=T;++t) {
                invat=inva[|(t-1)*cols(QQ0)+1,1\t*cols(QQ0),.|]
                sit2=selection0[(t+(i-1)*T),1]
                qit=Q0[(t+(i-1)*T),.]
                Pt=P[.,t]
                rit=invat*invsym(normal(qit*Pt)*(1-
normal(qit*Pt)))*normalden(qit*Pt)*qit'*(sit2-normal(qit*Pt))
                ri[.,t]=rit
        }
        ri1=vec(ri)

i=1
                ri=J(cols(Q0), T,0)
                for (t=1;t<=T;++t) {
                        At=J(cols(QQ0), cols(QQ0), 0)
                                for(g=1;g<=rows(shape);++g) {
                                        qit=QQ0[(t+(g-1)*T),.]
                                        Pt=P[.,t]
                                        b1=invsym(normal(qit*Pt)*(1-normal(qit*Pt)))
                                        b2=(normalden(qit*Pt)*normalden(qit*Pt))*qit'*qit
                                        bertha_it=b1*b2
                                        At=(At+bertha_it)
                                }
                        sit2=selection0[(t+(i-1)*T),1]
                        qit=Q0[(t+(i-1)*T),.]
                        Pt=P[.,t]
                        rit=invsym(At)*invsym(normal(qit*Pt)*(1-
normal(qit*Pt)))*normalden(qit*Pt)*qit'*(sit2-normal(qit*Pt))
                        ri[.,t]=rit

                }
                ri2=vec(ri)

// The difference should be very small.
minmax(ri1-ri2)
////////////////////////////////////////

////////////////////////////////////////////////////////
/// GENERATE B SEMYKINA P. 30 WITHOUT FRi IN EACH Pi ///
/// AS WOOLDRIDGE 1995 P. 132 (A.4) ////////////////////
// Dri IS OUTSIDE THE SUM FOR t=1,2,...,T
B=J(cols(H0),cols(H0),0)

for (i=1;i<=rows(info);++i) {
        pi=J(cols(H0),1,0)
        ki=J(cols(H0),1,0)
        ri=J(cols(Q0), T,0)

        for (s=1;s<=T;++s) {
                sit2=selection0[(s+(i-1)*T),1]
                eit=e0[(s+(i-1)*T),.]
                hit=H0[(s+(i-1)*T),.]
                kit=(sit2*hit'*eit)
                ki=ki+kit
        }
        for (t=1;t<=T;++t) {
                invat=inva[|(t-1)*cols(QQ0)+1,1\t*cols(QQ0),.|]
                sit2=selection0[(t+(i-1)*T),1]
                qit=Q0[(t+(i-1)*T),.]
                Pt=P[.,t]
                rit=invat*invsym(normal(qit*Pt)*(1-
normal(qit*Pt)))*normalden(qit*Pt)*qit'*(sit2-normal(qit*Pt))
                ri[.,t]=rit
        }
ri=vec(ri)
```

```
pi=ki-(F*ri)
B=B+(pi*pi')
}
////////////////////////////////////////////////////

// SE_APPENDIX SEMYKINA P. 30 (36)
cov_=invsym(C'*invsym(D)*C)*C'*invsym(D)*B*invsym(D)*C*invsym(C'*invsym(D)*C)
cov_=cov_*corr
se_appendix=sqrt(diagonal(cov_))
end

* POOLED OLS
* Heteroskedasticity test
reg `depvar' `indep_ins' `indep_noins' `time4'
estat hettest

reg `depvar' `indep_ins' `indep_noins' `time4', robust cluster(`ident')

* POOLED 2SLS no robust no cluster
ivregress 2sls `depvar' `indep_noins' `time4' (`indep_ins' = `indep_noins'
`instrumeans' `instrunomean' `time4'), first
estat firststage, all
estat overid
test `time4'

* POOLED 2SLS robust no cluster
ivregress 2sls `depvar' `indep_noins' `time4' (`indep_ins' = `indep_noins'
`instrumeans' `instrunomean' `time4'), robust first
estat firststage, all forcenonrobust
estat overid
test `time4'

* POOLED 2SLS robust cluster
ivregress 2sls `depvar' `indep_noins' `time4' (`indep_ins' = `indep_noins'
`instrumeans' `instrunomean' `time4'), robust cluster(`ident') first
estat firststage, all forcenonrobust
estat overid
test `time4'


* IVREG SHOULD BE IDENTICAL TO se_hat
* Procedure 2 no robust no cluster
ivregress 2sls `depvar' `indep_noins' `moy_instru' `time4' `lambda' (`indep_ins' =
`indep_noins' `instrumeans' `instrunomean' `moy_instru' `time4' `lambda'), first
estat firststage, all
estat overid
test `time4'
test `lambda'

* Procedure 2 robust no cluster
ivregress 2sls `depvar' `indep_noins' `moy_instru' `time4' `lambda' (`indep_ins' =
`indep_noins' `instrumeans' `instrunomean' `moy_instru' `time4' `lambda'), robust
first
estat firststage, all forcenonrobust
estat overid
test `time4'
test `lambda'

* Procedure 2 robust cluster
ivregress 2sls `depvar' `indep_noins' `moy_instru' `time4' `lambda' (`indep_ins' =
`indep_noins' `instrumeans' `instrunomean' `moy_instru' `time4' `lambda'), robust
cluster(`ident') first
estat firststage, all forcenonrobust
estat overid
test `time4'
```

```
test `lambda'


preserve
****************** Instrument tests ******************************
* CONCENTRATION PARAMETER, Stock, Wright, and Yogo
qui: ivregress 2sls `depvar' `indep_noins' `moy_instru' `time4' `lambda' (`indep_ins'
= `indep_noins' `instrumeans' `instrunomean' `moy_instru' `time4' `lambda'), robust
cluster(`ident') first
predict double resid, residuals
keep if lsalaireipc~=.
mata
st_view(V1=., ., "`ident'")
st_view(V2=., ., tokens("`time'"))
info = panelsetup(V1, 1, 1)
end
drop cons
g cons=1

        * Specify variables
        local endo `indep_ins'
        local X1i cons `indep_noins' `moy_instru' `time4' `lambda'
        local X2i `instrumeans' `instrunomean'
        local Zi `X1i' `X2i'


quietly {
local nendo = 0
foreach var of local endo {
        local nendo = `nendo'+1
        local y`nendo' `var'
}

* Correction for cluster
mata
st_view(fit=., ., tokens("`endo'"))
R2all=J(cols(fit),1,.)
Adj_R2all=J(cols(fit),1,.)
Fall=J(cols(fit),1,.)
Prob_Fall=J(cols(fit),1,.)
Gall=J(cols(fit),1,.)

for (w=1; w<=`nendo'; ++w) {
        st_view(yall=., ., tokens("`endo'"))
        st_view(X1i=., ., tokens("`X1i'"))
        st_view(X2i=., ., tokens("`X2i'"))
        st_view(Zi=., ., tokens("`Zi'"))
        yi = yall[.,w]

        Pi = invsym(Zi'Zi)*Zi'yi
        Pi1 = Pi[|1,1\cols(X1i),1|]
        Pi2 = Pi[|cols(X1i)+1,1\.,1|]

        // R2
        y_mean = mean(yi)
        y_=J(rows(yi),cols(yi),y_mean)
        yhat = Zi*Pi
        ui = yi - yhat
        dyhat = yhat-y_
        dyhat2 = dyhat:^2
        dy = yi-y_
        dy2=dy:^2
        SCE = colsum(dyhat2)
        SCT = colsum(dy2)
        R2 = SCE/SCT
```

```
R2all[w,.]=R2

// Adj_R2
Adj_R2 = 1-((rows(yi)-1)/(rows(yi)-rows(Pi))*(1-R2))
Adj_R2all[w,.]=Adj_R2

// F-stat
// GENERATE MATRIX Zi for COVARIANCE
Zi0=J(T*rows(info),cols(Zi), .)
res_u = J(T, rows(info), .)

for (g=1; g<=cols(Zi); ++g) {
        Zid=Zi[.,g];
        res_Zig= J(T, rows(info), .)
                for (i=1; i<=rows(info); ++i) {
                        panelsubview(Zid_=., Zid, i, info)
                        panelsubview(images_=., V2, i, info)
                        Zig = images_'*Zid_
                        res_Zig[.,i]=Zig
                        }
Zig=vec(res_Zig)
Zi0[.,g]=Zig
}
// GENERATE MATRIX ui for COVARIANCE
for (i=1; i<=rows(info); ++i) {
        panelsubview(ui_=., ui, i, info)
        panelsubview(images_=., V2, i, info)
        u = images_'*ui_
        res_u[.,i]=u
}
ui0 = vec(res_u)

// COVARIANCE FE TO VERIFY
E=J(cols(Zi),cols(Zi),0)
for (i=1; i<=rows(info); ++i) {
        Zii=J(T,cols(Zi),.)
        uii=J(T,1,.)
        for(l=T-1; l>=0; --l) {
                        Zii[T-l,.]=Zi0[T*i-l,.]
                        uii[T-l,.]=ui0[T*i-l,.]
        }
        Ei = Zii'*uii*uii'*Zii
        E=E+Ei
}

N1 = rows(yi)
ka = cols(Zi)
M = rows(info)
// Correction cluster http://www.stata.com/support/faqs/stat/robust.html
corr = ((N1-1)*M)/((N1-ka)*(M-1))
E=E*corr

// CALCULATE COV WOOLDRIDGE 2002 P. 275 (10.59)
cov = invsym(Zi'*Zi)*E*invsym(Zi'*Zi)
se = sqrt(diagonal(cov))
(Pi, se)


R=J(rows(Pi2),rows(Pi),0)
for (i=1;i<=rows(Pi2);++i) {
        R[i,i+rows(Pi1)]=1
}
q=J(rows(Pi2),1,0)

F = ((R*Pi-q)'*invsym(R*cov*R')*(R*Pi-q))/rows(q)
```

```
        Fall[w,.]=F
        Prob_F =  1-F(rows(Pi2),M-1,F)
        Prob_Fall[w,.]=Prob_F
}
end

* Partial R2
* The partial R2 is the R2 between y and X2 after
* eliminating the effects of X1.
local mx1x2
foreach v of local X2i {
        reg `v' `X1i', noconstant
        predict double mx`v', residuals
        local mx1x2 `mx1x2' mx`v'
}
local PR2
foreach v of local endo {
        reg `v' `X1i', noconstant
        predict double mx1`v', residuals
        reg mx1`v' `mx1x2', noconstant
        local PR2 `PR2' P_R2_`v'
        matrix P_R2_`v' = e(r2)
}
drop mx*
local PR2_
foreach var of local PR2 {
        local PR2_ `PR2_' `var' \

}
local characters = length("`PR2_'")-2
local PR2_mat = substr("`PR2_'", 1, `characters')
matrix PR2_all = (`PR2_mat')
matrix list PR2_all

mata
// Stock and Yogo minimum eigenvalue statistic
XpX = quadcross(X1i, X1i)
XpY = quadcross(X1i, yall)
XpZ = quadcross(X1i, X2i)
YpZ = quadcross(yall, X2i)
YpZbar = quadcross(yall, (X1i, X2i))
ZpZ = quadcross(X2i, X2i)
ZbarpZbar = quadcross((X1i, X2i), (X1i, X2i))
XpXi = cholinv(XpX)
YpMxZ = YpZ - XpY'*XpXi*XpZ
ZpMxZ = ZpZ - XpZ'*XpXi*XpZ
YpY = quadcross(yall, yall)
YpMzbarY = YpY - YpZbar*cholinv(ZbarpZbar)*YpZbar'
Sigvv = YpMzbarY / (rows(yall) - cols(X1i) - cols(X2i))
Sigvvisq = matpowersym(0.5*(Sigvv + Sigvv'), -0.5)
G = Sigvvisq * (YpMxZ*cholinv(ZpMxZ)*YpMxZ') * Sigvvisq / cols(X2i)
if (!issymmetric(G)) {
        _makesymmetric(G)
}
G = min(symeigenvalues(G))

PR2_all = st_matrix("PR2_all")
Test_letters1 = ("R-sq.", "Adj. R-sq.", "Partial R-sq.", "Robust F", "Prob > F")
Test_results1 = (R2all , Adj_R2all , PR2_all , Fall , Prob_Fall)
s1 = "Minimum eigenvalue statistic ="
end

* Shea's partial R2
local e=0
local SR2
```

```
if `nendo'~= 1 {
	foreach yi of local endo {
		local e = `e'+1
		display "`yi'"
		local y0 = subinstr("`endo'", "`yi'", "",1)
		display "`y0'"
		reg `yi' `y0' `X1i', noconstant
		predict double yi_tilde, residuals
		reg `yi' `X1i' `X2i', noconstant
		predict double yi_hat
		local c = 0
		local y0_hat
			foreach col of local y0 {
				local c = `c'+1
				reg `col' `X1i' `X2i', noconstant
				predict double y0_hat`c'
				local y0_hat `y0_hat' y0_hat`c'
			}
		reg yi_hat `y0_hat' `X1i', noconstant
		predict double yi_tilde_hat, residuals
		reg yi_tilde yi_tilde_hat
		matrix SR2_`e' = e(r2)
		local SR2 `SR2' SR2_`e'
		drop yi_tilde yi_hat yi_tilde_hat y0_hat*
	}
	local SR2_
	foreach var of local SR2 {
		local SR2_ `SR2_' `var' \

	}
	local characters = length("`SR2_'")-2
	local SR2_mat = substr("`SR2_'", 1, `characters')
	matrix SR2_all = (`SR2_mat')
	mata
	SR2_all = st_matrix("SR2_all")
	SR2_alli = J(cols(yall), 2, .)
	SR2_alli[.,1]=SR2_all

	for (i=1; i<=cols(yall); ++i) {
		SR2 = SR2_all[i,1]
		ASR2i = 1-(((1-SR2)*(rows(yall)-1))/(rows(yall)-cols(Zi)+1))
		SR2_alli[i,2]=ASR2i
	}
	SR2_all = SR2_alli
	s2 = "Shea's partial R-squared"
	Test_letters2 = ("   Shea's   ", "      Shea's      " \ "Partial R-sq.", "Adj.
Partial R-sq.")
	end
}
else {
	mata
	s2 = "No Shea's statistic"
	Test_letters2 = 0
	SR2_all = 0
	end
}


* Wooldridge's (1995) score test
* of overidentifying restrictions
* It needs the residuals.

local Y_hat
local f = 0
foreach var of local endo {
```

```
        local f = `f'+1
        reg `var' `Zi', noconstant
        predict double Y_hat`f'
        local Y_hat `Y_hat' Y_hat`f'
}
display "`Y_hat'"

local j = 0
local k_hat
local over_instru = wordcount("`X2i'") - wordcount("`endo'")
display"`over_instru'"
foreach var of local X2i {
        local j = `j'+1
        reg `var' `X1i' `Y_hat', noconstant
        predict double q_hat`j', residuals
        gen k_hat`j' = q_hat`j'*resid
        local k_hat `k_hat' k_hat`j'
        drop q_hat`j'
        if `j' == `over_instru' {
                continue, break
        }
}

drop resid
drop Y_hat*
display "`k_hat'"
g coni=1
reg coni `k_hat', noconstant
drop coni k_hat*
mata
RSS = st_numscalar("e(rss)")
W = rows(yall)-RSS
Probchi2 = chi2tail(`over_instru',W)
s3 = "Wooldridge's (1995) score test of overidentifying restrictions:"
Test_letters3 = (" Score chi2", "W", "p-value")
Test_results3 = (`over_instru' , W , Probchi2)
end
}

mata
Test_letters1
Test_results1

display(s1)
G

display(s2)
Test_letters2
SR2_all


display(s3)
Test_letters3
Test_results3
end
*****************************************************************
restore


mata
///////////////// WALD TEST SELECTION BIAS ///////////////////////
// [R] R-Z p. 416

b=O[|rows(O)-T+1, 1\rows(O),1|]
V=cov_[|rows(cov_)-T+1, cols(cov_)-T+1\rows(cov_),cols(cov_)|]
```

```
R=I(T)
r=J(T,1,0)

W=(R*b-r)'*invsym(R*V*R')*(R*b-r)
Probchi2 = chi2tail(T,W)

W
Probchi2
////////////////////////////////////////////////////////////

// SHOULD BE IDENTICAL, THIS IS JUST A CHECK
(O, se_hat, O:/se_hat, 2*ttail(N-M-ka+1, abs(O:/se_hat)))

// POOLED S2LS SEMYKINA P. 28 (30)
(O, se_appendix, O:/se_appendix, 2*ttail(N-M-ka+1, abs(O:/se_appendix)))
mata clear
end
*******************************************************************************
```

RÉFÉRENCES

Baker, M., J. Gruber et K. Milligan. 2005. « Universal Childcare, Maternal Labor Supply, and Family Well-Being ». *NBER Working Paper*, no 11832 (décembre), 61 p. <http://www.nber.org/papers/w11832> (19 mai 2007).

Baltagi, B. H. et Y. Chang. 2000. « Simultaneous equations with incomplete panels ». *Econometric Theory*, vol. 16, no 2 (avril), p. 269-279.

Blau, D. et J. Currie. 2004. « Preschool, Day Care, and Afterschool Care : Who's Minding the Kids? ». *NBER Working Paper*, no 10670 (août), 124 p.

Blau, Francine D. et Lawrence M. Kahn. 2006. « Changes in the Labor Supply Behavior of Married Women : 1980-2000 », *Institute for the Study of Labor (IZA)*, Discussion Paper No. 2180, 61 p. <http://ftp.iza.org/dp2180.pdf> (8 juillet 2007).

Cleveland, Gordon, Morley Gunderson et Douglas Hyatt. 1996. « Child Care Costs and the Employment Decision of Women : Canadian Evidence ». *Revue canadienne d'Economique*, vol. 29, no 1 (août), p. 132-151.

Cragg, John G. et Stephen G. Donald. 1993. « Testing Identifiability and Specification in Instrumental Variable Models ». *Econometric Theory*, vol. 9, no 2 (juin), p. 222-240.

Dustmann, Christian et Maria Engracia Rochina-Barrachina. 2000. « Selection Correction in Panel Data Models: An Application to Labour Supply and Wages », *Institute for the Study of Labor (IZA)*, Discussion Paper No. 162 (juin), 56 p.

Duthil, Gérard. 2004. *Économie du travail et des ressources humaines*. Paris : L'Harmattan, 364 p.

French, Eric, Bhashkar Mazumder et Christopher Taber. 2006. « The Changing Pattern of Wage Growth for Low Skilled Workers ». *Federal Reserve Bank of Chicago*, WP 2005-24 (mai), 19 p.

Gladden, Tricia et Christopher Taber. 2006. « The Relationship Between Wage Growth and Wage Levels », *Journal of Applied Econometrics*, 2ième tour, juillet 2006, 27p. <http://www.faculty.econ.northwestern.edu/faculty/taber/tabglad.pdf> (8 juillet 2007).

Goldsmith, Arthur H. et Jonathan R. Veum. 2002. « Wages and the Composition of Experience », *Southern Economic Journal*, vol. 69, no 2 (octobre), p. 429-443.

Greene, William H. 2008. *Econometric Analysis*, 6e éd. Upper Saddle River (New Jersey): Prentice Hall, 1178 p.

Harkness, Susan et Jane Waldfogel. 1999. « The Family Gap in Pay: Evidence from Seven Industrialised Countries », *Centre for Analysis of Social Exclusion, LSE*, no 30 (décembre), 38 p.

Hahn, Jinyong et Jerry Hausman. 2003. « Weak Instruments: Diagnosis and Cures in Empirical Econometrics », *The American Economic Review*, vol. 93, no 2 (mai), p. 118-125.

Heckman, James J. 1974. « Effects of Child-Care Programs on Women's Work Effort », *Journal of Political Economy*, vol. 82, no 2 (mars-avril), p. S136-S163.

Heckman, James J. 1976. « The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models », *Annals of Economic and Social Measurement*, 5, p. 475-492.

Heckman, James J. 1979. « Sample Selection Biais as a Specification Error », *Econometrica*, vol. 47, no 1 (janvier), p. 153-161.

Hotchkiss, Julie L. et M. Melinda Pitts. 2007. « The Role of Labor Market Intermittency in Explaining Gender Wage Differentials », *Federal Reserve Bank of Atlanta Working Paper Series*, 2007-1, 22 p. <http://www.frbatlanta.org/filelegacydocs/wp0701.pdf> (8 juillet 2007).

Kaufman, Bruce E. 1991. *The Economics of Labor Markets*, éd. rev. de *The Economics of Labor Markets and Labor Relations*, c1989. Montréal: Dryden Press, 778 p.

Kunze, Astrid. 2002. « The Timing of Careers and Human Capital Depreciation ». *Institute for the Study of Labor Discussion Paper* (*IZA DP)*, no 509 (juin), 37 p.

Lee, L.-F. 1978. « Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables », *International Economic Review*, vol. 19, no 2 (juin), p. 415-433.

Lefebvre, Pierre et Philip Merrigan. 2005a. « Low-fee ($5/day/child) Regulated Childcare Policy and the Labor Supply of Mothers with Young Children: A Natural Experiment from Canada ». *Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi (CIRPÉE)*, Cahier de recherche 05-08, 50 p.

Lefebvre, Pierre et Philip Merrigan. 2005b. « The Québec's Experiment of $5 per day per Child Childcare Policy and Mother's Labor Supply: Evidence Based on the Five Cycles of the NLSCY ». *Centre Interuniversitaire sur le Risque, les Politiques Économiques et l'Emploi (CIRPÉE)*, 40 p.

Lefebvre, Pierre et Philip Merrigan. 2007a. *Childcare Policy and the Labor Supply of Mothers with Young Children : A Natural Experiment from Canada*. Montréal (Qué.): Université du Québec à Montréal, version révisée de 2005a à paraître dans le *Journal of Labor Economics*, 23 p.

Lefebvre, Pierre, Philip Merrigan et Matthieu Verstraete. 2007b. *Dynamic Labour Supply Effects of Childcare Subsidies : Evidence from a Canadian Natural Experiment*. Montréal (Qué.): Université du Québec à Montréal, à paraître dans le *Journal of Labor Economics*, 41p.

Michalopoulos, C. et P. Robins. 2000. « Employment and Child-Care Choices in Canada and the United States ». *Revue canadienne d'Economique*, vol. 33, no 2 (mai), p. 435-470.

Nielsen, Helena Skyt, Marianne Simonsen et Mette Verner. 2002. « Does the Gap in Family-Friendly Policies Drive the Family Gap? ». *Working Papers* 02-19, Department of Economics of Aarhus School of Business, 37 p.

Oaxaca, Ronald L. 1973. « Male-Female Wage Differentials in Urban Labor Markets ». *International Economic Review*, vol. 14, no 3 (octobre), p. 693-709.

Powell, Lisa M. 1997. « The Impact of Child Care Costs on the Labour Supply of Married Mothers : Evidence from Canada ». *Revue canadienne d'Economique*, vol. 30, no 3 (août), p. 577-594.

Québec, ministère de l'Emploi, de la Solidarité sociale et de la Famille. 2004. *Rapport annuel de gestion 2003-2004*. F-488 (2004-11). Québec : Direction des communications, 122 p. < http://www.messf.gouv.qc.ca> (15 août 2007).

Québec, ministère de la Famille et des Aînés. 2008. *Places en services de garde*. <http://www.mfa.gouv.qc.ca/services-de-garde/portrait-des-services-de-garde/places-en-services-de-garde/> (14 juin 2008).

Semykina, Anastasia et Jeffrey M. Wooldridge. 2006. *Estimating Panel Data Models in the Presence of Endogeneity and Selection: Theory and Application*. Mimeo: Michigan State University Department of Economics, 35 p.

Shea, John. 1997. « Instrument Relevance in Multivariate Linear Models : A Simple Measure ». *The Review of Economics and Statistics*, vol. 79, no 2 (mai), p. 348-352.

StataCorp. 2007. « ivregress postestimation – Postestimation tools for ivregress », *Stata Base Reference Manual*, 10e éd., vol. 2. College Station (Texas): StataCorp LP, p. 45-56.

Stock, James H., Jonathan H. Wright et Motohiro Yogo. 2002. « A survey of weak instruments and weak identification in generalized method of moments ». *Journal of Business and Economic Statistics*, vol. 20, no 4 (octobre), p. 518-529.

Stock, James H. et Motohiro Yogo. 2005. « Testing for weak instruments in linear IV regression ». *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, éd. par D. W. K. Andrews et J. H. Stock, chap. 5. New York: Cambridge University Press, p. 80-108.

Waldfogel, Jane. 1997. « The Effect of Children on Women's Wages ». *American Sociological Review*, vol. 62, no 2 (avril), p. 209-217.

Waldfogel, Jane. 1998. « Understanding the "Family Gap" in Pay for Women with Children ». *The Journal of Economic Perspectives*, vol. 12, no 1 (hiver), p. 137-156.

Wooldridge, Jeffrey M. 1995a. « Score diagnostics for linear models estimated by two stages least squares ». *Advances in Econometrics and Quantitative Economics: Essays in Honor of Professor C. R. Rao*, éd. par G. S. Maddala, P. C. B. Phillips et T. N. Srinivasan, Oxford: Blakwell, p. 66-87.

Wooldridge, Jeffrey M. 1995b. « Selection corrections for panel data models under conditional mean independence assumptions ». *Journal of Econometrics*, vol. 68, no 1 (juillet), p. 115-132.

Wooldridge, Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge (Mass.): MIT Press, 752 p.

Zhang, Xuelin. 2002. « Progression de la rémunération des travailleurs peu qualifiés au Canada d'après l'EDTR (1993-1998) ». *Statistique Canada – Direction des études analytiques : document de recherche*, 11F0019MIF, no 194, p. 25.