

LA DYNAMIQUE DU FAIBLE REVENU DANS QUATRE PAYS

Dennis Batten¹, Miles Corak² et Wen-Hao Chen³

RÉSUMÉ

Nous modélisons la dynamique de la sortie d'une période de faibles revenus au moyen de modèles de risque en temps discret en tenant compte de l'effet de l'hétérogénéité observée et inobservée. Nous mettons l'accent sur la méthodologie qui sous-tend les modèles économétriques de risque et les méthodes d'estimation que nous utilisons. Pour illustrer nos travaux, nous présentons une application des modèles à des données longitudinales recueillies au Royaume-Uni, en Allemagne, aux États-Unis et au Canada. L'analyse porte sur des données provenant des Cross-National Equivalent Files de la BHPS (pour le Royaume-Uni), de la GSOEP (pour l'Allemagne), de la PSID (pour les États-Unis) et de l'EDTR (pour le Canada). Nous utilisons aussi des données administratives canadiennes provenant des dossiers fiscaux.

MOTS-CLÉS : Données d'enquête longitudinales, hétérogénéité inobservée, mélange de distributions, périodes multiples.

1. INTRODUCTION

Aujourd'hui, l'importance de l'étude du marché du travail sous un angle dynamique est généralement reconnue dans le milieu de la recherche universitaire et en politique publique. La collecte de données longitudinales et leur analyse appropriée ont permis de brosser un tableau plus précis des marchés du travail nord-américain et européen qui éclaire les débats stratégiques de l'heure. L'exemple le plus évident est la meilleure compréhension de la nature du faible revenu. L'étude de Bane et Ellwood (1986) réalisée sur des données longitudinales en provenance des États-Unis est souvent citée comme exemple de portrait particulièrement puissant de la population à faible revenu et du processus dynamique qui détermine l'entrée dans la pauvreté et la sortie de celle-ci. La disponibilité progressive de données comparables dans d'autres pays a donné naissance à une profusion de publications sur le sujet, dont les plus récentes sont Bradbury, Jenkins et Micklewright (2001), Jenkins (2000) et Stevens (1999).

Notre étude consiste en une analyse comparative de la dynamique du faible revenu en nous concentrant sur l'Amérique du Nord et sur deux pays européens. Nous utilisons des données longitudinales provenant du Canada, des États-Unis, d'Allemagne et du Royaume-Uni, et nous nous concentrons sur les faits survenus durant les années 1990.

L'article met l'accent sur la méthodologie qui sous-tend les modèles économétriques de risque et les méthodes d'estimation que nous avons utilisés pour exécuter la recherche. À la section 2, nous donnons un bref aperçu des sources de données et attirons l'attention sur certains problèmes qu'il faut résoudre lors de la réalisation d'une analyse comparative de cette nature. Puis, à la section 3, nous décrivons en détail la modélisation des probabilités de transition, y compris la façon de tenir compte de l'hétérogénéité inobservée, et les méthodes d'estimation appliquées aux données d'enquête complexes. Aussi nous donnons un petit exemple qui illustre le fonctionnement des modèles et la façon d'estimer les variances. Enfin, à la section 4, nous résumons brièvement ce que nous avons accompli jusqu'à présent et précisons l'orientation que nous avons l'intention de donner à nos futurs travaux.

¹Statistique Canada, Ottawa, Canada, K1Y 0T6 (dennis.batten@statcan.ca)

²Statistique Canada, Ottawa, Canada, K1Y 0T6 (miles.corak@statcan.ca)

³Statistique Canada, Ottawa, Canada, K1Y 0T6 (wen-hao.chen@statcan.ca)

2. SOURCES DES DONNÉES ET QUESTIONS CONNEXES

Les données proviennent des Cross-National Equivalent Files (CNEF) qui regroupent des données longitudinales couvrant plusieurs cycles d'enquête provenant du Canada, des États-Unis, de Grande-Bretagne et d'Allemagne. Les variables ont été définies de la même façon dans les diverses enquêtes afin de faciliter les études internationales (Burkhauser et coll., 2000). Les données extraites des CNEF aux fins de la présente étude proviennent de l'Enquête sur la dynamique du travail et du revenu du Canada (EDTR, 1993 à 1998), du Panel Survey of Income Dynamics des États-Unis (PSID, 1990 à 1996), du British Household Panel Survey du Royaume-Uni (BHPS, 1991 à 1999) et du German Socio-Economic Panel d'Allemagne (GSOEP, 1992 à 1999). L'avantage des CNEF est qu'ils fournissent un ensemble d'estimations sur des variables annuelles du revenu que l'on ne peut obtenir directement à partir des ensembles de données originaux produits d'après les variables originales observées dans les divers pays. L'ensemble contient des données sur le revenu du ménage avant et après transferts gouvernementaux et des estimations au niveau du ménage du revenu annuel du travail, des avoirs, du loyer imputé, des transferts publics et privés et des impôts versés. L'existence de renseignements sur chaque source de revenu des ménages permet de brosser un tableau complet des rôles relatifs du marché du travail, de la structure familiale et de l'État dans la détermination du niveau de revenu.

Une étude comprenant des comparaisons entre plusieurs pays, intégrant des données provenant de diverses enquêtes basées sur des méthodes d'échantillonnage différentes, oblige à prendre des décisions cruciales et difficiles concernant les divers concepts et définitions. Quelles populations comparons-nous? En tenant compte des limites de chaque enquête, de la nature dynamique des populations longitudinales, des taux de non-réponse et d'autres facteurs, nous devons définir l'unité d'analyse et le revenu de l'unité. Devons-nous utiliser les données sur le revenu avant ou après impôt? Pour le CNEF, nous n'utilisons pas les variables après transferts gouvernementaux, étant donné le manque d'uniformité des définitions d'un pays à l'autre. Comment définissons-nous le faible revenu? Il n'existe pas de définition unique du faible revenu pour les quatre pays, les États-Unis étant les seuls à avoir établi un seuil officiel de « pauvreté ». Comment attribuons-nous un « revenu équivalent » à chaque membre de l'unité? Devons-nous considérer le revenu total de l'unité comme étant une fonction de la taille de l'unité ou une fonction du nombre d'adultes et du nombre d'enfants en attribuant des poids différents aux adultes et aux enfants? Il existe, pour la plupart de ces questions, de nombreuses réponses qui sont toutes acceptables. Voici ce que nous avons fait.

Nous avons attribué à chacun des membres d'un même ménage, qu'ils soient ou non unis par les liens du sang ou du mariage, un revenu équivalent. Pour déterminer ce revenu équivalent, nous avons exprimé le revenu total du ménage en dollars constants de 1997, puis nous l'avons divisé par la racine carrée de la taille du ménage, en considérant le revenu de chaque membre du ménage comme étant le revenu du marché avant les transferts gouvernementaux. Le revenu du marché est défini de la même façon que le revenu total sans tenir compte des transferts publics, des prestations de sécurité sociale ni des impôts. Nous pouvons le considérer comme une estimation du revenu disponible éventuel de chaque membre du ménage sous l'hypothèse d'un partage égal. De surcroît, nous utilisons le revenu annuel plutôt que le revenu au moment de l'entrevue. Nous considérons comme ayant un faible revenu durant une année particulière toute personne dont le revenu du ménage est inférieur à la moitié de la médiane nationale cette année-là.

Les échantillons sont conçus de sorte qu'ils soient représentatifs de tous les individus formant la population, y compris les enfants et les personnes qui ne travaillent pas. Cependant, pour l'ajustement de nos modèles de transition, nous n'avons inclus que les individus dans une période de faible revenu ayant débuté durant la période d'observation de l'étude (pas de censure à gauche). Une fois qu'un individu est entré dans une période de faible revenu, nous le suivons jusqu'à la fin de la période ou jusqu'à ce que celle-ci soit censurée à droite à la fin de la période d'enquête. Nous avons inclus dans l'étude les périodes multiples de faible revenu d'un même individu et avons supposé qu'elles étaient indépendantes les unes des autres. Par conséquent, notre unité d'analyse la plus fondamentale est une période de faible revenu et nous cherchons à observer, après qu'un individu commence une période de faible revenu, la fin de cette période. Nous avons éliminé de l'étude tout individu présentant une discontinuité à n'importe quel moment durant la période étudiée (données manquantes pour une année ou plus).

Nous avons également inclus dans l'étude des données provenant de la Base de données administratives longitudinales (BDAL) du Canada afin de pouvoir comparer deux sources de données canadiennes. La BDAL est créée d'après divers fichiers de données administratives, dont les fichiers des déclarations de revenus et des prestations fiscales pour enfants. À l'heure actuelle, la BDAL contient des renseignements sur les individus et leur

famille couvrant 19 années (de 1982 à 2000). Des renseignements sur ces individus couvrant d'autres années sont ajoutés à mesure qu'ils deviennent disponibles. Les individus qui figurent dans la BDAL sont sélectionnés selon un plan d'échantillonnage de Bernoulli. Cette méthode permet d'échantillonner facilement les individus et de tenir compte des changements démographiques, comme les naissances, les décès, l'immigration ou l'émigration. (Division des données régionales et administratives, 2002).

3. MODÉLISATION DES PROBABILITÉS DE TRANSITION

Au temps $t=0$, nous observons le début d'une période de faible revenu i . Pour toute autre période t , $t>0$, soit y_{it} , $i=1,\dots,n$ le résultat de la période i où $y_{it}=1$ si la i^{e} période de faible revenu se termine au temps t et $y_{it}=0$, autrement. Supposons que la probabilité conditionnelle de réussite au temps t est modélisée par

$$\text{logit}(p_{it}) = \mathbf{x}_{it}^T \boldsymbol{\beta}, \quad (3.1)$$

où $\text{logit}(a) = \log(a/(1-a))$, \mathbf{x}_{it} est un vecteur de dimension $p \times 1$ de variables indépendantes et, éventuellement, d'interactions entre ces variables indépendantes, et $\boldsymbol{\beta}$ est un vecteur de dimension $p \times 1$ de paramètres. Nous cherchons principalement à faire des inférences au sujet de $\boldsymbol{\beta}$. Supposons pour commencer que nous observons la i^{e} période T_i jusqu'à un succès (fin de la période de faible revenu) ou une censure à droite. Pour une période i donnée, la fonction de vraisemblance sous l'hypothèse d'indépendance d'une année à l'autre est

$$L_i = \prod_t^{T_i} f_{it}, \quad (3.2)$$

où $f_{it} = (1-p_{it})^{1-y_{it}} (p_{it})^{y_{it}}$ dans le cas de la loi de Bernoulli pour une année particulière. Nous pouvons nous servir, pour estimer $\boldsymbol{\beta}$, de logiciels statistiques standard qui utilisent généralement la méthode de Newton-Raphson ou la méthode des scores de Fisher. L'estimateur $\hat{\boldsymbol{\beta}}$ est la solution d'une équation d'estimation de la forme

$$U(\boldsymbol{\beta}) = \sum_i u_i(\boldsymbol{\beta}), \quad (3.3)$$

où $u_i(\boldsymbol{\beta}) = \sum_t (y_{it} - p_{it}) \mathbf{x}_{it}$ sous un modèle logit pour p_{it} comme dans (3.1).

3.1 Hétérogénéité inobservée

L'un des objectifs principaux de l'étude est de trouver des variables indépendantes qui prédisent la probabilité de succès lequel, dans le présent contexte, consiste à observer la fin d'une période de faible revenu. Le vecteur de variables indépendantes peut inclure la durée de la période, ainsi que des caractéristiques socio-économiques et démographiques. Cependant, la probabilité de succès peut être influencée par une variable de classe latente. Une variable inobservable discrète, disons Z , indique la classe latente de la i^{e} période. Nous supposons que la variable prend G valeurs distinctes correspondant chacune à une probabilité distincte de succès. Par conséquent, la probabilité de réussite de la i^{e} période au temps t dépend des caractéristiques observées ainsi qu'inobservées. Par exemple, dans le cas de $G=2$ groupes, nous introduisons la composante inobservée dans le modèle de la même manière que les variables observées, c'est-à-dire

$$h(\mu_{igt}) = \lambda_g \mathbf{x}_{it} + \lambda_g \quad (3.4)$$

où $\mu_{igt} = E(y_{it} / z_i = g, \mathbf{x}_{it}, \boldsymbol{\beta})$ est la moyenne conditionnelle, $\boldsymbol{\beta}_g^T = (\lambda_g^T, \boldsymbol{\alpha}_g^T)$ étant l'effet des variables observées et inobservées pour le g^{e} groupe, et $h(\cdot)$ est la fonction lien. L'un des objectifs est d'identifier l'effet de classe latente, ce qui revient statistiquement à déterminer si $\boldsymbol{\alpha}_g$ diffère d'un groupe latent à l'autre.

3.2 Estimation

Lorsque les données peuvent être considérées comme provenant de deux populations ou plus mélangées en proportions variables, on utilise comme cadre d'analyse un mélange de lois de distribution, qui couvre le problème des classes latentes. Chaque y_i observé est tiré d'une superpopulation P qui est un mélange d'un nombre fini de G populations P_1, \dots, P_G dans certaines proportions π_1, \dots, π_G , avec $\sum_g \pi_g = 1$ et $\pi_g \geq 0$, $g = 1, \dots, G$.

À partir des données et d'une forme connue des distributions, nous souhaitons estimer les paramètres du modèle et du mélange de distribution. Un mélange de lois peut être traité au moyen de l'algorithme d'Expectation-Maximisation (EM) (Dempster, Laird et Rubin, 1977). Il est possible, pour appliquer cette procédure, d'utiliser un progiciel statistique comme SAS et d'écrire un code itératif en se servant d'une fonction définie par SAS à l'étape de la maximisation. Par exemple, avec une fonction lien logit, on utiliserait la fonction *proc logistic de SAS* à l'étape de la maximisation. Essentiellement, on peut pour cela créer G copies des données, puis utiliser les probabilités conditionnelles comme poids sous les conditions de recensement. Le calcul des poids et l'estimation des paramètres sont répétés jusqu'à la convergence. Définissons les variables indicatrices d'appartenance au groupe latent pour la période i comme étant $z_{ig} = 1$ si $i \in P_g$ et $z_{ig} = 0$ si $i \notin P_g$. Nous appliquons l'algorithme EM au mélange de lois de distribution en traitant la variable z_{ig} comme si les données manquaient. La vraisemblance pour l'ensemble des données pour la période i est donnée par

$$L_i = \prod_{g=1}^G (\pi_g f_{ig})^{z_{ig}} \quad (3.5)$$

où $f_{ig} = f_{ig}(y_i^1, \dots, y_i^T | \beta, g)$. En partant d'une valeur initiale pour $\boldsymbol{\varphi} = (\boldsymbol{\beta}, \boldsymbol{\pi})$, disons $\boldsymbol{\varphi}^{(m)}$, l'étape E nécessite le calcul de la pseudo log-vraisemblance complète fondée sur les données incomplètes

$$Q(\boldsymbol{\varphi}, \boldsymbol{\varphi}^{(m)}) = E(l_i | y_i, \boldsymbol{\varphi}^{(m)})$$

où $l_i = \log L_i$. Autrement dit, chaque variable indicatrice z_{ig} est remplacée par son espérance conditionnelle, τ_{ig} , où

$$\tau_{ig}^{(m)} = \pi_g^{(m)} f_{ig}(\boldsymbol{\beta}^{(m)}) / \sum_j \pi_j^{(m)} f_{ij}(\boldsymbol{\beta}_j^{(m)}).$$

Le but de l'étape M est de choisir la valeur de $\boldsymbol{\varphi}$, disons $\boldsymbol{\varphi}^{(m+1)}$, qui maximise $Q(\boldsymbol{\varphi}, \boldsymbol{\varphi}^{(m)})$ sous la contrainte $\sum_g \pi_g = 1$. Pour réaliser l'étape de la maximisation pour chaque itération EM, il suffit de résoudre

$$\partial l_i / \partial \boldsymbol{\varphi} = \partial \log f(\mathbf{y}_i | \mathbf{x}_i, \boldsymbol{\beta}, \boldsymbol{\pi}) / \partial \boldsymbol{\varphi}, \quad (3.6)$$

qui donne lieu aux deux équations d'estimation suivantes

$$\partial l_i / \partial \boldsymbol{\beta} = \sum_g \tau_{ig}^{(m)}(s) f(\mathbf{y}_i | z = g_i, \boldsymbol{\beta}) / \partial \boldsymbol{\beta} \quad (3.7)$$

$$\partial l_i / \partial \boldsymbol{\pi} = \sum_g \tau_{ig}^{(m)}(s) f(\mathbf{y}_i | z = g_i, \boldsymbol{\pi}) / \partial \boldsymbol{\pi}. \quad (3.8)$$

Les équations (3.7) et (3.8) sont toutes deux de la forme de (3.3). Si les données proviennent d'une enquête à plan de sondage complexe, un estimateur de (3.3) cohérent par rapport au plan est donné par

$$\hat{U}(\boldsymbol{\beta}) = \sum_i w_i(s) u_i(\boldsymbol{\beta}), \quad (3.9)$$

où $w_i(s)$ est le poids de sondage pour la i^{e} période dans l'échantillon s .

3.3 Caractérisation de la dynamique du faible revenu

Aux fins du présent article, nous illustrons la méthode en modélisant la probabilité qu'une période de faible revenu se termine à l'aide d'une seule variable explicative, nommément la durée de la période de faible revenu avant qu'elle se termine. En outre, nous choisissons une période de référence commune (1993 à 1998) sur les ensembles de données par souci de simplicité et pour assurer la comparabilité entre les pays. Supposons le modèle qui suit pour la sortie d'une période de « faible » revenu :

$$\text{logit}(p_{it(g)}) = \alpha_g + \lambda_{g2}d_{2i}(t) + \lambda_{g3}d_{3i}(t) + \lambda_{g4}d_{4i}(t)$$

sous un mélange de $G=2$ distributions, où $d_{ji}(t)=1$ si $t=j$ années et 0 , autrement et λ_{gt} représente l'effet de régression associé à la sortie de la i^{e} unité d'une période de faible revenu après t années sous le g^{e} groupe. Les estimations des paramètres sont présentées au tableau (3.1) pour les cinq ensembles de données étudiés. Le tableau (3.2) donne la probabilité qu'une période de faible revenu se termine après une durée particulière (1, ..., 4+).

Les caractéristiques des deux groupes représentés aux tableaux (3.1) et (3.2) semblent être semblables pour chaque ensemble de données. Pour les périodes de faible revenu appartenant au premier groupe, la probabilité que la période se termine paraît très faible au début, mais elle augmente à mesure que s'accroît la longueur de la période. Ce groupe semble caractéristique d'environ 29 % des périodes de faible revenu en Grande-Bretagne, 36 % et 37 % des périodes de faible revenu des ensembles de données de la BDAL et de l'EDTR, respectivement, au Canada, et 37 % des périodes de faible revenu en Allemagne. Le second groupe, qui est caractéristique du reste des périodes de faible revenu dans chaque pays, indique une situation inverse. Les périodes de faible revenu appartenant à ce groupe semblent avoir une plus forte probabilité de se terminer à leur début que celles appartenant au premier groupe. Par contre, plus elles se prolongent, moins il est probable qu'elles se terminent.

Tableau 3.1 : Estimations paramétriques de la probabilité qu'une période de faible revenu se termine (1993 à 1998)

Paramètres du modèle	Canada				Allemagne		Grande-Bretagne		* États-Unis	
	BDAL		EDTR		1	2	1	2	1	2
<u>G</u>	1	2	1	2	1	2	1	2	1	2
\forall	-2,81	-0,02	-3,15	0,24	-3,18	-0,02	-3,78	-0,09	0,46	-0,86
δ_2	0,81	-0,78	0,30	-0,06	0,75	-0,51	0,31	-0,32	-1,98	-0,45
δ_3	1,56	-2,46	1,75	1,42	1,27	-2,48	1,40	-1,70		
δ_{4+}	1,77	-4,95	1,85	-4,76	1,92	-5,52	3,03	-4,16		
B	0,36	0,64	0,37	0,63	0,37	0,63	0,29	0,71	0,37	0,63

* Nota : une erreur de lecture du fichier CNEF pour les données de 1997 des États-Unis a empêché de poursuivre l'analyse pour ce pays au-delà de la deuxième année.

Tableau 3.2 : Probabilité qu'une période de faible revenu se termine après t années, 1993 à 1998

<i>Durée</i>	<i>Canada</i>				<i>Allemagne</i>		<i>Grande-Bretagne</i>		<i>États-Unis</i>	
	BDAL		EDTR							
<i>G</i>	<i>1</i>	<i>2</i>	<i>1</i>	<i>2</i>	<i>1</i>	<i>2</i>	<i>1</i>	<i>2</i>	<i>1</i>	<i>2</i>
<i>1</i>	0,057	0,495	0,041	0,560	0,040	0,495	0,022	0,476	0,613	0,297
<i>2</i>	0,119	0,310	0,055	0,545	0,081	0,371	0,030	0,399	0,179	0,212
<i>3</i>	0,223	0,077	0,198	0,840	0,129	0,076	0,085	0,143		
<i>4+</i>	0,261	0,007	0,214	0,012	0,221	0,004	0,321	0,014		
<i>B</i>	0,36	0,64	0,37	0,63	0,37	0,63	0,29	0,71	0,37	0,63

Les ensembles de données observés semblent être caractérisés par des profils semblables lorsqu'on utilise uniquement la variable de durée pour expliquer la probabilité qu'une période de faible revenu se termine. Cependant, quand on considère des variables explicatives supplémentaires (sexe, âge, structure familiale, etc.), les modèles donnent des résultats distincts pour les quatre pays.

3.4 Estimation de la variance

Pour faire des comparaisons appropriées entre les pays et à l'intérieur de ceux-ci, ainsi que pour déterminer la signification de nos estimations, nous devons produire une bonne estimation de la variance. Nous allons donc donner des exemples de comparaisons et proposer un estimateur de la variance.

Pour tester la signification des paramètres sous-jacents, nous devons réaliser un test d'hypothèse qui indiquera si la valeur des paramètres diffère de zéro. Par exemple, pour tester si la valeur d'un paramètre est significative au Canada, nous réalisons le test $H_0 : \Phi_{\text{CAN}}^{(1)} = 0$. Pour déterminer s'il existe des différences significatives entre pays ou, dans le cas du Canada, entre les populations représentées par les deux échantillons étudiés, nous devons vérifier une autre hypothèse. Par exemple, pour voir si les effets sont différents au Canada et aux États-Unis, nous testons $H_0 : \Phi_{\text{CAN}}^{(1)} = \Phi_{\text{USA}}^{(1)}$, et pour évaluer toute différence entre les deux ensembles de données disponibles pour le Canada, nous testons $H_0 : \Phi_{\text{LAD}}^{(1)} = \Phi_{\text{SLID}}^{(1)}$, où $\Phi^{(1)} \subseteq \Phi$. Pour exécuter ces tests statistiques, nous devons estimer la variance du vecteur de coefficients estimés. Elle est donnée par

$$\text{Var}(\hat{\Phi}) = E_M \text{Var}_S(\hat{\Phi}) + \text{Var}_M E_S(\hat{\Phi})$$

où E_M et Var_M représentent l'espérance et la variance fondées sur le modèle, respectivement. E_S et Var_S représentent l'espérance et la variance sous le plan de sondage, respectivement. Une estimation de la variance est donnée par

$$\text{est}(\text{Var}(\hat{\Phi})) = v_S(\hat{\Phi}) + \text{est}(\text{Var}_M E_S(\hat{\Phi})).$$

Quand n/N est faible, $\text{est}(\text{Var}_M E_S(\hat{\Phi}))$ est faible. Par conséquent, une bonne approximation pour $\text{Var}(\hat{\Phi})$ est

$$\text{est}(\text{Var}(\hat{\Phi})) \approx v_S(\hat{\Phi}).$$

Pour calculer $v_S(\hat{\phi})$, nous supposons qu'on peut approximer le plan de sondage par une sélection avec remise d'unités primaires d'échantillonnage (UPE) sur chaque strate, puis nous utilisons une méthode d'estimation de la variance, comme celle du Jackknife ou du Bootstrap.

3.5 Extensions futures

Pour étudier deux types d'événements ou plus dans le même modèle, on peut encore utiliser l'équation (3.3). Par exemple, si l'on considère simultanément la probabilité conditionnelle de sortir d'une période de faible revenu λ_{it} et la probabilité d'entrer de nouveau dans une période de faible revenu q_{it} , il nous suffit de redéfinir le vecteur de variables indépendantes et le vecteur de paramètres. Soit $\tilde{\mathbf{x}}_{it}$ le nouveau vecteur de variables indépendantes de dimension $(2p \times 1)$ avec $\tilde{\mathbf{x}}_{it}^T = (\mathbf{x}_{it}^T, \mathbf{0}^T)$ pour la sortie de la période de faible revenu et $\tilde{\mathbf{x}}_{it}^T = (\mathbf{0}^T, \mathbf{x}_{it}^T)$ pour la réentrée dans une période de faible revenu, où $\mathbf{0}$ est un vecteur de 0 $(p \times 1)$. Soit $\boldsymbol{\beta}$ le nouveau vecteur de paramètres $(2p \times 1)$ où $\boldsymbol{\beta}^T = (\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T)$, $\boldsymbol{\beta}_1$ est le vecteur de paramètres associé à λ_{it} et $\boldsymbol{\beta}_2$ est le vecteur de paramètres associé à q_{it} . Nous pouvons écrire chaque probabilité en fonction de $\tilde{\mathbf{x}}_{it}$ et de $\boldsymbol{\beta}$, c.-à-d. $\text{logit}(\lambda_{it}) = \mathbf{x}_{it}\boldsymbol{\beta}_1 + \mathbf{0}\boldsymbol{\beta}_2 = p_{it}$ et $\text{logit}(q_{it}) = \mathbf{0}\boldsymbol{\beta}_1 + \mathbf{x}_{it}\boldsymbol{\beta}_2 = p_{it}$.

4. CONCLUSION

Nous démontrons dans le présent article comment modéliser les probabilités de transition en tenant compte de l'hétérogénéité inobservée grâce à un modèle de mélange de distribution. Nous appliquons la procédure à l'étude de la dynamique du revenu dans quatre pays, à savoir le Canada, la Grande-Bretagne, l'Allemagne et les États-Unis. Nous utilisons l'algorithme d'Espérance-Maximisation (EM) pour estimer les paramètres inconnus des fonctions économétriques de risque. Cet algorithme peut être appliqué à l'aide de tout progiciel statistique standard.

Bien que nous ayons ajusté les modèles pour tous les ensembles de données, nous soulignons qu'il faut établir un estimateur de la variance avant de pouvoir faire des inférences appropriées quant aux comparaisons entre pays et à l'intérieur des pays. En outre, à la section 3.4, nous proposons un estimateur de la variance qui nous permettra de faire les inférences nécessaires. Bien que l'estimateur par le Jackknife proposé soit utilisé et reconnu à grande échelle dans la littérature, nous devons poursuivre les travaux avant de pouvoir l'appliquer à notre situation particulière.

RÉFÉRENCES

- Bane, M. J. et Ellwood D. (1986). "Slipping Into and Out of Poverty: The Dynamics of Spells." *Journal of Human Resources*. Vol. 21 No. 1, pp. 1-23.
- Bradbury, B., Jenkins S. P. et Micklewright J. (2001). *The Dynamics of Child Poverty in Industrialized Countries*. Cambridge University Press.
- Burkhauser, R. V., Butrica B. A., Daly M. C., et Lillard D. R. (2000). "The Cross-National Equivalent File: A product of cross-national research." in *Social Insurance in a Dynamic Society*, Irene Becker, Notburga Ott, and Gabriele Rolf (Eds.)
- D'Ambrosio, C., Papadopoulos F. et Tsakoglou P. (2002). "Social Exclusion in EU Member States: A Comparison of two Alternative Approaches." Paper presented to the XVI Annual Conference of the European Society for Population Economics, Bilbao.38.
- Dempster, A. P., Laird N. M. et Rubin D. B. (1977). "Maximum likelihood Estimation from Incomplete Data via the EM Algorithm (with Discussion)." *Journal of the Royal Statistical Society*. Series B. Vol. 39, pp 1-38.

Jenkins, S. (2000). "Modelling Household Income Dynamics." *Journal of population Economics*. Vol. 73, pp. 13-22.

Small Area and Administrative Data Division (2002), Longitudinal Administrative Databank, Dictionary 2000, Catalogue no. 12-585-XIE, Statistics Canada.

Stevens, A. H. (1999). "Climbing Out of Poverty, falling Back In: Measuring the Persistence of Poverty Over Multiple Spells." *Journal of Human Resources*. Vol. 34 No. 3, pp. 557-588.

Stewart, K. (2002). "Measuring Well-Being and Exclusion in Europe's Regions." Paper presented to the XVI Annual Conference of the European Society for Population Economics, Bilbao.

OECD (2001). *Employment Outlook*. Paris: OECD.