

Techniques d'enquête

Une approche markovienne mixte à classes latentes pour estimer la mobilité sur le marché du travail au moyen d'indicateurs multiples et d'une interrogation rétrospective

par Francesca Bassi, Marcel Croon et Davide Vidotto

Date de diffusion : le 22 juin 2017



Statistique
Canada

Statistics
Canada

Canada

Comment obtenir d'autres renseignements

Pour toute demande de renseignements au sujet de ce produit ou sur l'ensemble des données et des services de Statistique Canada, visiter notre site Web à www.statcan.gc.ca.

Vous pouvez également communiquer avec nous par :

Courriel à STATCAN.infostats-infostats.STATCAN@canada.ca

Téléphone entre 8 h 30 et 16 h 30 du lundi au vendredi aux numéros sans frais suivants :

- Service de renseignements statistiques 1-800-263-1136
- Service national d'appareils de télécommunications pour les malentendants 1-800-363-7629
- Télécopieur 1-877-287-4369

Programme des services de dépôt

- Service de renseignements 1-800-635-7943
- Télécopieur 1-800-565-7757

Normes de service à la clientèle

Statistique Canada s'engage à fournir à ses clients des services rapides, fiables et courtois. À cet égard, notre organisme s'est doté de normes de service à la clientèle que les employés observent. Pour obtenir une copie de ces normes de service, veuillez communiquer avec Statistique Canada au numéro sans frais 1-800-263-1136. Les normes de service sont aussi publiées sur le site www.statcan.gc.ca sous « Contactez-nous » > « Normes de service à la clientèle ».

Note de reconnaissance

Le succès du système statistique du Canada repose sur un partenariat bien établi entre Statistique Canada et la population du Canada, les entreprises, les administrations et les autres organismes. Sans cette collaboration et cette bonne volonté, il serait impossible de produire des statistiques exactes et actuelles.

Signes conventionnels dans les tableaux

Les signes conventionnels suivants sont employés dans les publications de Statistique Canada :

- . indisponible pour toute période de référence
- .. indisponible pour une période de référence précise
- ... n'ayant pas lieu de figurer
- 0 zéro absolu ou valeur arrondie à zéro
- 0^s valeur arrondie à 0 (zéro) là où il y a une distinction importante entre le zéro absolu et la valeur arrondie
- ^p provisoire
- ^r révisé
- x confidentiel en vertu des dispositions de la *Loi sur la statistique*
- ^E à utiliser avec prudence
- F trop peu fiable pour être publié
- * valeur significativement différente de l'estimation pour la catégorie de référence ($p < 0,05$)

Publication autorisée par le ministre responsable de Statistique Canada

© Ministre de l'Industrie, 2017

Tous droits réservés. L'utilisation de la présente publication est assujettie aux modalités de l'[entente de licence ouverte](#) de Statistique Canada.

Une [version HTML](#) est aussi disponible.

This publication is also available in English.

Une approche markovienne mixte à classes latentes pour estimer la mobilité sur le marché du travail au moyen d'indicateurs multiples et d'une interrogation rétrospective

Francesca Bassi, Marcel Croon et Davide Vidotto¹

Résumé

Les erreurs de mesure peuvent provoquer un biais de l'estimation des transitions, donnant lieu à des conclusions erronées au sujet de la dynamique du marché du travail. La littérature traditionnelle sur l'estimation des mouvements bruts est basée sur la supposition que les erreurs de mesure ne sont pas corrélées au fil du temps. Cette supposition n'est pas réaliste dans bien des contextes, en raison du plan d'enquête et des stratégies de collecte de données. Dans le présent document, nous utilisons une approche basée sur un modèle pour corriger les mouvements bruts observés des erreurs de classification au moyen de modèles markoviens à classes latentes. Nous nous reportons aux données recueillies dans le cadre de l'enquête italienne continue sur la population active, qui est transversale et trimestrielle et qui comporte un plan de renouvellement de type 2-2-2. Le questionnaire nous permet d'utiliser plusieurs indicateurs des états de la population active pour chaque trimestre : deux recueillis au cours de la première interview, et un troisième recueilli un an plus tard. Notre approche fournit une méthode pour estimer la mobilité sur le marché du travail, en tenant compte des erreurs corrélées et du plan par renouvellement de l'enquête. Le modèle qui convient le mieux est un modèle markovien mixte à classes latentes, avec des covariables touchant les transitions latentes et des erreurs corrélées parmi les indicateurs; les composantes mixtes sont de type mobile-stable. Le caractère plus approprié de la spécification du modèle mixte est attribuable à des transitions latentes estimées avec une plus grande précision.

Mots-clés : Mouvements bruts; marché du travail; modèles mixtes; modèles à classes latentes.

1 Introduction

Les analystes peuvent exploiter les données par panels pour estimer les mouvements bruts de la population active, c'est-à-dire les transitions d'un état à un autre au fil du temps. Les mouvements nets mesurent les variations de divers états sur le marché au fil du temps, tandis que les mouvements bruts nous renseignent sur la dynamique du marché du travail.

Beaucoup de documents sur l'estimation des mouvements bruts sont basés sur l'hypothèse que les erreurs sont non corrélées au fil du temps, c'est-à-dire qu'il s'agit d'erreurs de classification indépendantes (ECI). L'hypothèse d'ECI suppose que : (i) les erreurs de classification désignant deux occasions différentes sont indépendantes l'une de l'autre en fonction des états réels, et (ii) les erreurs dépendent exclusivement de l'état réel actuel. Par conséquent, les erreurs de classification produisent de fausses transitions et provoquent donc une surestimation des changements.

Cependant, dans bien des contextes, l'hypothèse de l'ECI s'avère irréaliste, en raison du plan d'enquête et des stratégies de collecte de données. En pareilles circonstances, les erreurs de classification peuvent être corrélées : les états observés peuvent également dépendre des états réels à d'autres moments ou des transitions réelles, ou il peut y avoir des effets directs entre les états observés (Bound, Brown et Mathiowetz 2001).

1. Francesca Bassi, Department of Statistical Sciences, Université de Padoue, Italie, Via C. Battisti 241, 35121, Padoue, Italie. Courriel : francesca.bassi@unipd.it; Marcel Croon et Davide Vidotto, département de la méthodologie et des statistiques, Université de Tilburg, aux Pays-Bas.

Dans le présent document, nous utilisons une approche basée sur un modèle pour rajuster les mouvements bruts observés pour les erreurs de classification. Cette approche agence un sous-modèle structurel pour les taux de transition réels non observés et un sous-modèle de mesure corrélant les états réels aux états observés. Un cadre utile pour formuler notre modèle est fourni par l'analyse des classes latentes (CL).

Nous appliquons notre approche aux mouvements bruts observés au sein des trois états de la population active – Occupant un emploi (E), En chômage (C) et Inactif (I) – tirés de l'enquête italienne continue sur la population active (EICPA), une enquête trimestrielle reposant sur un plan de renouvellement de type 2-2-2 entraînant des panels à deux cycles à un trimestre, trois trimestres et un an d'intervalle. Nous considérons que les données sont recueillies de 2005 à 2009.

Le questionnaire nous permet d'utiliser plusieurs indicateurs des états de la population active pour chaque trimestre : (i) tous les répondants sont classifiés comme Occupant un emploi, En chômage ou Inactifs, d'après la définition du Bureau international du travail (BIT) en fonction des réponses données à une série de questions; (ii) on demande aux répondants de s'autodéclarer comme occupant un emploi, en chômage ou inactifs, ce qui constitue l'auto-évaluation de l'état; (iii) une question rétrospective porte sur l'état des répondants sur le marché du travail un an avant l'interview. Cette approche offre une façon d'estimer la mobilité sur le marché du travail en tenant compte des erreurs de mesure corrélées et du plan par renouvellement de l'enquête.

Plus précisément, le modèle qui convient le mieux est un modèle markovien à classes latentes (MMCL), dont les covariables touchant les transitions latentes et les erreurs corrélées font partie des indicateurs. On obtient le modèle mixte en présumant de l'existence de deux sous-populations non observables : les répondants mobiles, c'est-à-dire ceux qui changent d'état sur le marché du travail pendant la période à l'étude, et les répondants stables. Notre recherche a eu comme résultat secondaire que le modèle des répondants mobiles/stables et le MMCL estiment la même quantité d'erreur de mesure dans les données. Le meilleur ajustement de la spécification du modèle mixte est attribuable à des transitions latentes estimées avec une plus grande précision. Magidson, Vermunt et Tran (2007) ont également constaté que le modèle markovien mixte par CL convient mieux aux données que le modèle traditionnel. Toutefois, dans ce cas-ci, la différence pour ce qui est de l'ajustement était attribuable au fait que, comme l'hétérogénéité n'a pas été prise en compte, il s'en est suivi une surestimation de l'erreur de mesure.

Notre document examine les contributions récentes aux ouvrages scientifiques sur la question de l'estimation des mouvements bruts au moyen de la chaîne markovienne cachée et d'indicateurs multiples. Une description exacte du modèle se trouve dans Langeheine (1994). La méthode ne s'appliquait pas seulement à l'estimation des mouvements bruts sur le marché du travail, mais aussi à bien d'autres contextes, des données longitudinales étant disponibles. Paas, Vermunt et Bijmolt (2007), par exemple, ont estimé un MMCL pour étudier les tendances d'acquisition sur le marché des produits financiers; plusieurs indicateurs de détention de produits financiers ont été utilisés pour déterminer des segments du marché non observables directement, au sein desquels les clients pouvaient se déplacer à différentes reprises consécutives de mesure.

Bartolucci, Lupparelli et Montanari (2009) ont estimé le même modèle pour les changements suivants de l'état de santé dans un échantillon de patients au fil du temps. Manzoni, Vermunt, Luijkx et Muffels (2010) ont appliqué un MMCL pour estimer les mouvements bruts sur le marché du travail suédois. Dans une étude plus récente, Pavlopoulos et Vermunt (2015) ont utilisé un modèle markovien pour estimer l'ampleur de l'erreur de mesure dans les renseignements de l'enquête sur la population active hollandaise et l'institut hollandais sur l'assurance pour les employés sur le type d'emploi (permanent ou temporaire).

Grâce à la contribution du présent document à la littérature scientifique au sujet de l'estimation des mouvements bruts, nous avons maintenant trois indicateurs, l'un d'eux étant recueilli rétrospectivement, sur l'état sur le marché du travail, et nous pouvons également tenir compte du plan par renouvellement de l'enquête. Le document contribue également à la littérature sur la qualité des données de l'EICPA (Bassi, Padoan et Trivellato 2012).

La présentation de l'article est la suivante. La section 2 présente le modèle traditionnel (ou standard) et le MMCL mixte. La section 3 décrit l'enquête et ses données. La section 4 compare les rendements du modèle traditionnel et du MMCL mixte. La section 5 présente les résultats, en s'appuyant sur le modèle convenant le mieux pour corriger les mouvements bruts sur le marché du travail à partir des erreurs de mesure. Les conclusions sont présentées à la section 6.

2 Le modèle markovien à classes latentes

L'analyse des classes latentes a été appliquée dans plusieurs études sur les données par panel pour séparer les changements réels des changements observés touchés par des mesures non fiables. Parmi les contributions relativement récentes, mentionnons celles de Bassi, Torelli et Trivellato (1998), de Biemer et Bushery (2000), de Bassi, Croon, Hageaars et Vermunt (2000) et de Bassi et Trivellato (2009).

L'état réel sur le marché du travail est traité comme une variable latente, et l'état observé est traité comme son indicateur. Ce modèle comprend deux parties :

- a) structure, décrivant la dynamique réelle parmi les variables latentes;
- b) mesure, appariant chaque variable latente à son indicateur ou ses indicateurs.

Examinons la formulation la plus simple des modèles markoviens à classes latentes (MMCL) (Wiggins 1973), qui suppose que les transitions non observables réelles suivent une chaîne markovienne de premier ordre. Comme pour toutes les spécifications du MMCL standard, l'indépendance locale parmi les indicateurs est présumée, c'est-à-dire que les indicateurs sont indépendants en fonction des variables latentes. Dans le MMCL comportant un indicateur par variable latente, la supposition de l'indépendance locale coïncide avec l'état des erreurs de classification indépendantes.

Si l'on suppose que X_{it} indique l'état réel sur le marché du travail au moment t pour une personne de l'échantillon générique $i, i = 1, \dots, n$; Y_{it} est l'état observé correspondant; $P(X_{i1} = l_1)$ est la probabilité de

l'état initial de la chaîne markovienne latente et $P(X_{it+1} = l_{t+1} | X_{it} = l_t)$ est la probabilité de transition entre l'état l_t et l'état l_{t+1} du moment t à $t + 1$, sachant que $t = 1, \dots, T - 1$, où T représente le nombre total de périodes consécutives, séparées par des intervalles identiques, pendant lesquelles une personne est observée. En outre, $P(Y_{it} = j_t | X_{it} = l_t)$ est la probabilité d'observer l'état j au moment t , sachant que la personne i au moment t est dans l'état réel l_t : on parle également ici de la composante de mesure du modèle.

Il en résulte que $P(Y(1), \dots, Y(T))$ est la proportion d'unités observées dans une cellule générique du tableau de contingence à T – entrées. Pour une personne de l'échantillon générique i , un MMCL se définit comme suit :

$$\begin{aligned}
 P(\mathbf{Y}_i = \mathbf{y}) &= \sum_{l_1}^K \dots \sum_{l_T}^K P(X_{i1} = l_1) \\
 &\quad \prod_{t=2}^T P(X_{it} = l_t | X_{it-1} = l_{t-1}) \\
 &\quad \prod_{t=1}^T P(Y_{it} = j_t | X_{it} = l_t)
 \end{aligned} \tag{2.1}$$

où \mathbf{y} est le vecteur renfermant les valeurs observées pour la personne i , l_t et j_t varient sur K classes (dans notre application, trois états de la population active). L'équation (2.1) précise la proportion d'unités dans la cellule générique d'un tableau de contingence à T – entrées comme produit de probabilités marginales et conditionnelles.

Dans un MMCL comportant des variables concomitantes, l'appartenance à des classes latentes et les transitions latentes sont exprimées comme des fonctions de covariables avec des distributions connues (Dayton et McReady 1988). $P(X_{i1} = l_1 | \mathbf{Z}_{i1} = \mathbf{z}_1)$, où \mathbf{z}_1 est un vecteur renfermant les valeurs des covariables pour le répondant i au moment 1, estime les effets des covariables sur l'état initial, et $P(X_{it} = l_t | X_{it-1}, \mathbf{Z}_{it} = \mathbf{z}_t)$, où \mathbf{z}_t est un vecteur renfermant les valeurs des covariables pour le répondant i au moment t , estime les effets des covariables sur les transitions latentes.

En fonction des composantes qui précèdent, le modèle complet pour la personne i est donné par :

$$\begin{aligned}
 P(\mathbf{Y}_i = \mathbf{y} | \mathbf{Z}_i = \mathbf{z}) &= \sum_{l_1}^K \dots \sum_{l_T}^K P(X_{i1} = l_1 | \mathbf{Z}_1 = \mathbf{z}_1) \\
 &\quad \prod_{t=2}^T P(X_{it} = l_t | X_{it-1} = l_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t) \\
 &\quad \prod_{t=1}^T P(Y_{it} = j_t | X_{it} = l_t)
 \end{aligned} \tag{2.2}$$

Lorsque plusieurs indicateurs (M) par variable latente sont observés, la formulation du modèle devient la suivante (Vermunt 2010) :

$$\begin{aligned}
P(\mathbf{Y}_i = \mathbf{y} | \mathbf{Z}_i = \mathbf{z}) &= \sum_{l_1}^K \dots \sum_{l_T}^K P(X_{i1} = l_1 | \mathbf{Z}_1 = \mathbf{z}_1) \\
&\quad \prod_{t=2}^T P(X_{it} = l_t | X_{it-1} = l_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t) \\
&\quad \prod_{m=1}^M \prod_{t=1}^T P(Y_{mit} = j_t | X_{it} = l_t)
\end{aligned} \tag{2.3}$$

Dans notre application, les indicateurs M sont donnés par les trois éléments d'information recueillis pour tous les répondants sur leur état sur le marché du travail.

Habituellement, les probabilités conditionnelles sont paramétrées et limitées par des modèles de régression logistique. Les paramètres sont estimés au moyen du maximum de vraisemblance (Vermunt et Magidson 2013). L'identification est un problème bien connu dans les modèles comportant des variables latentes et, bien que le nombre de paramètres indépendants ne doive pas dépasser le nombre de fréquences observées, ce n'est pas une condition suffisante. D'après Goodman (1974), une condition suffisante pour l'identifiabilité locale est que la matrice d'information soit définie positive. Le logiciel Latent Gold (Vermunt et Magidson 2008) fournit de l'information sur l'identification des paramètres. Un autre problème lié à l'estimation est celui des maxima locaux; pour y remédier, nous avons estimé nos modèles plusieurs fois avec différents ensembles de valeurs de départ.

Un MMCL mixte présume de l'existence dans la population de groupes non directement observables qui se déplacent au fil du temps, en suivant des chaînes latentes ayant différentes probabilités d'état initial et différentes probabilités de transition; on peut également présumer que les groupes ont des probabilités de réponse différentes (van de Pol et Langeheine 1990). Un tel modèle peut être élargi de manière à inclure des covariables variables dans le temps et des covariables constantes dans le temps (Vermunt, Tran et Magidson 2008). Un cas particulier d'un MMCL mixte à deux classes est le modèle mobile-stable : le groupe des mobiles a des probabilités positives de passer d'un état à un autre au fil du temps, et le groupe des stables ne change pas. Pour ce dernier groupe, les probabilités de transition entre les différents états sont imposées à zéro. Un MMCL mixte à deux classes comportant des variables concomitantes prend la forme suivante :

$$\begin{aligned}
P(\mathbf{Y}_i = \mathbf{y} | \mathbf{Z}_i = \mathbf{z}) &= \sum_{w=1}^2 \sum_{l_1}^K \dots \sum_{l_T}^K P(W = w) P(X_{i1} = l_1 | \mathbf{Z}_1 = \mathbf{z}_1, W = w) \\
&\quad \prod_{t=2}^T P(X_{it} = l_t | X_{it-1} = l_{t-1}, \mathbf{Z}_{it} = \mathbf{z}_t, W = w) \\
&\quad \prod_{j=1}^K \prod_{t=1}^T P(Y_{it} = j_t | X_{it} = l_t, W = w)
\end{aligned} \tag{2.4}$$

où W est une variable latente binaire. Le modèle mobile-stable est obtenu en présumant, pour $l_t \neq l_{t-1}$, que $P(X_{it} = l_t | X_{it-1} = l_{t-1}, W = 2) = 0$ et donc pour $l_t = l_{t-1}$ que $P(X_{it} = l_t | X_{it-1} = l_{t-1}, W = 2) = 1$.

La fonction de probabilité d'un modèle de CL peut également être estimée si l'information ne se trouve pas dans les variables de réponse. Nous profitons de cette occasion pour tenir compte des tendances de réponse générées par le plan d'enquête par renouvellement. Les ménages échantillonnés sont interviewés sur deux trimestres consécutifs, ne participent pas à l'enquête pendant les deux trimestres suivants et sont ensuite interviewés de nouveau, à deux autres reprises (voir le tableau 3.1). Nous avons présumé que l'information manquante en raison du plan d'enquête est manquante de façon aléatoire. Dans ce cas-ci, chaque unité contribue seulement à la fonction de vraisemblance d'après l'information disponible (Vermunt 1997).

3 Les données

L'enquête italienne continue sur la population active (EICPA), réalisée par l'ISTAT (*Italian Institute of Statistics*), est la principale source officielle de documentation statistique sur le marché du travail italien. L'EICPA est réalisée depuis 1969 et a été modifiée bien des fois. En 2004, une importante mise à jour a été effectuée, principalement en raison de l'obligation d'adapter l'enquête aux nouvelles normes de l'Union européenne (UE). Les principaux changements mettaient en cause des interviews réparties au fil des années de l'étude, de nouveaux critères pour classer l'état des répondants sur le marché du travail, des techniques de collecte de données assistée par ordinateur et des interviews avec rétroinformation. Chaque année, l'enquête recueille des renseignements sur environ 280 000 ménages, ce qui donne au total environ 700 000 personnes. La population de référence se compose de tous les membres des ménages résidant officiellement en Italie.

Le plan d'échantillonnage de l'EICPA comporte deux étapes : 1) les municipalités ont été désignées comme les unités primaires d'échantillonnage (UPE) avec stratification, et les ménages ont été désignés comme les unités finales d'échantillonnage (UFE) avec renouvellement. Les UPE ont été stratifiées en fonction de la taille de la population. Les grandes municipalités, dont la population dépasse un seuil donné (également appelées les municipalités auto-représentatives), étaient toujours incluses dans l'échantillon; les petites municipalités (non auto-représentatives) étaient regroupées en strate, de manière à ce qu'une municipalité par strate soit sélectionnée avec une probabilité proportionnelle à sa population; 2) les ménages étaient sélectionnés au hasard à partir des registres de population dans toutes les municipalités tirées à l'étape 1.

L'enquête a été réalisée tous les trimestres en fonction d'un plan par renouvellement de type 2-2-2. Les membres des ménages ont été interviewés pendant deux trimestres consécutifs. Après une pause de deux trimestres, ils ont été interviewés à nouveau, deux fois pendant les deux trimestres correspondants de l'année suivante. Par conséquent, chaque ménage a été inclus dans quatre cycles de l'enquête pendant une période de 15 mois. Ce système par renouvellement signifiait que la moitié de l'échantillon est demeuré inchangé au cours de deux trimestres consécutifs et des trimestres à un an d'intervalle, et que 25 % de l'échantillon est demeuré inchangé pendant trois trimestres.

Toutes les analyses statistiques suivantes sont faites sur la population longitudinale. L'EICPA n'est pas conçue comme une enquête par panel en bonne et due forme : la population initiale change pendant la période à l'étude en raison d'événements démographiques et de migrations. Bien que l'ISTAT ait proposé une procédure pour calculer les poids longitudinaux (Boschetto, Discenza, Lucarelli, Rosati et Fiori 2009), ces derniers ne sont pas à la disposition des chercheurs, ce qui fait que nous n'avons pas pu prendre en considération le plan d'échantillonnage complexe. Toutefois, nous considérons qu'il était raisonnable de présumer que les répondants appartenant aux mêmes ménages étaient indépendants.

L'information sur l'état sur le marché du travail dans un trimestre de référence a été recueillie trois fois : (i) chaque répondant était classifié comme occupant un emploi, en chômage ou inactif, conformément à la définition du BIT en fonction des réponses données à un groupe de questions en particulier; (ii) dans une section subséquente du questionnaire, tous les répondants devaient se classifier sur le marché du travail, afin de recueillir l'état « auto-évalué »; (iii) au bout d'un an, dans une question rétrospective, on demandait aux répondants quel était leur état sur le marché du travail un an avant la première interview.

D'après la définition du BIT, les répondants ont été classifiés comme occupant un emploi pendant le trimestre de référence s'ils avaient 15 ans ou plus pendant la période de référence et ont effectué un type quelconque de travail, pendant au moins une heure, contre rémunération, bénéfices ou gain familial, ou s'ils n'étaient pas au travail, mais ils avaient un emploi ou une entreprise, dont ils étaient absents temporairement en raison d'une maladie, de vacances, d'un conflit de travail ou d'études ou de formation. Les répondants ont été classifiés comme en chômage s'ils avaient 15 à 74 ans et étaient : (a) sans travail pendant la semaine de référence; (b) actuellement disponibles pour travailler dans les deux semaines suivant la semaine de référence; (c) activement à la recherche de travail, c'est-à-dire qu'ils avaient pris des mesures particulières pour trouver du travail, pendant la période de quatre semaines se terminant par la semaine de référence, ou s'ils n'étaient pas à la recherche de travail, mais avaient trouvé un emploi devant commencer plus tard, dans une période allant jusqu'à trois mois (Organisation internationale du travail (OIT) 2008).

L'auto-évaluation actuelle et la question rétrospective classifiaient les répondants dans huit catégories : occupant un emploi; en chômage et à la recherche d'un nouvel emploi; en chômage et à la recherche d'un premier emploi; s'acquittant de tâches ménagères; étudiants; retraités; ayant un handicap les empêchant de travailler; autres.

Le tableau 3.1 démontre le plan par renouvellement de l'enquête pour deux années civiles consécutives. Les lettres indiquent les groupes de renouvellement : quatre groupes de renouvellement ont été interviewés à chaque trimestre. En ce qui concerne une année civile, l'information sur l'état sur le marché du travail provenait de neuf groupes de renouvellement. Toutefois, le plan par renouvellement produit une tendance particulière de données manquantes. Par exemple, pour les unités du groupe de renouvellement A qui sont interviewées pour la quatrième fois au premier trimestre de l'année 1, seuls l'indicateur du BIT (B) et l'autoévaluation (A) de l'état sur le marché du travail au cours du premier trimestre de l'année 1 sont disponibles. Pour les unités du groupe de renouvellement F, qui ont été interviewées pour la première fois au premier trimestre de l'année 1, nous avons seulement de l'information sur l'état sur le marché du travail

en fonction de la définition du BIT, de l'auto-évaluation et de la question rétrospective (R) pour les deux premiers trimestres de l'année 1.

Tableau 3.1
Plan par renouvellement de l'EICPA

Groupe de renouvellement	Année 1				Année 2			
	Trimestre I	Trimestre II	Trimestre III	Trimestre IV	Trimestre I	Trimestre II	Trimestre III	Trimestre IV
A	B-A							
B	B-A	B-A						
C		B-A	B-A					
D			B-A	B-A				
E	B-A-R			B-A	B-A			
F	B-A-R	B-A-R			B-A	B-A		
G		B-A-R	B-A-R			B-A	B-A	
H			B-A-R	B-A-R			B-A	B-A
I				B-A-R	B-A-R			B-A
L					B-A-R	B-A-R		
M						B-A-R	B-A-R	
N							B-A-R	B-A-R
O								B-A-R

B = Indicateur du BIT, A = auto-évaluation de l'état sur le marché du travail, R = indicateur rétrospectif.

Nous avons examiné les données recueillies de 2005 à 2010. (Sont exclues de ces analyses les données recueillies en 2004, première année de la mise en œuvre de la nouvelle enquête sur la population active, parce que les données n'étaient pas complètement fiables; en ce qui concerne 2010, nous n'utilisons ici que l'information recueillie au moyen de la question rétrospective et portant sur l'état sur le marché du travail en 2009.) Le tableau 3.2 énumère la composition du marché du travail au premier trimestre à partir de données regroupées au cours de la période de cinq ans. L'indicateur du BIT compte manifestement un pourcentage plus faible de personnes en chômage et un plus fort pourcentage de personnes inactives que les deux autres indicateurs. Les deux mesures basées sur l'auto-évaluation donnent un plus haut taux de chômage, parce que le BIT applique une définition très stricte du chômage. Pour être réputés en chômage, les répondants de 15 à 74 ans ne doivent pas occuper un emploi au moment de l'interview, mais ils doivent être prêts à accepter des emplois convenables au cours des deux prochaines semaines si l'occasion se présente, et ils doivent avoir cherché activement des façons de décrocher un emploi dans les deux semaines précédentes. Le BIT fournit ces lignes directrices afin de faciliter les comparaisons du rendement sur le marché du travail au fil du temps et d'un pays à un autre (OIT 2008). Cependant, ce cadre a été établi à un moment où le type d'emploi prédominant était à temps plein et sous contrat permanent; depuis, la situation d'emploi est devenue plus souple, les emplois à temps partiel et temporaires s'étant multipliés, en particulier pour les personnes sur le point d'entrer sur le marché du travail.

Tableau 3.2
Composition du marché du travail 2005 – Trimestre I de 2009, % – données regroupées

	E	C	I
BIT	43,07	3,60	53,33
A	41,73	6,73	51,54
R	41,55	6,49	51,96

E = Personnes ayant un emploi, C = Personnes en chômage, I = Personnes inactives.

D'autres études dans la littérature démontrent que la distinction entre les états sur le marché du travail n'est pas toujours évidente : les personnes ne connaissent peut-être pas les définitions officielles ou elles peuvent considérer leur état sur le marché du travail comme différent des critères standard (par exemple, voir Clark et Summer 1979; Flinn et Heckman 1983; Gonul 1992). Dans la plupart des cas, il est difficile de faire la distinction entre les personnes en chômage et les personnes inactives : la condition la plus importante semble être la recherche active d'un emploi, puisque les répondants peuvent se considérer comme en chômage même s'ils ne cherchent pas activement un emploi. Des incohérences peuvent donc survenir entre l'information recueillie dans le cadre d'enquêtes et le comportement réel. Une autre explication des différences entre les classifications du BIT et de l'auto-évaluation est que les répondants ayant un emploi temporaire pour ce qui est du nombre d'heures par semaine peuvent ne pas se considérer comme ayant un emploi.

Le tableau 3.3 énumère les incohérences, c'est-à-dire les différents états sur le marché du travail observés pour le même répondant avec deux indicateurs, parmi les trois indicateurs pour la période en question. Les données au fil des trimestres et des années ont été regroupées pour des motifs d'espace. Le nombre d'incohérences est manifestement plus élevé pour l'état de chômage que pour les deux autres états, et la plupart des mauvaises classifications ont tendance à désigner des personnes hors de la population active plutôt que des personnes occupant un emploi, comme l'ont démontré bien des études précédentes (par exemple, voir Poterba et Summers 1986). En comparant l'état sur le marché du travail conformément à la définition du BIT à celui déclaré d'après les réponses à la question rétrospective, on a obtenu le plus grand nombre d'incohérences. En examinant les cohérences au fil des trimestres et des années pour deux des trois indicateurs (non déclarés ici faute d'espace), nous constatons que la cohérence a tendance à augmenter légèrement au fil du temps, peut-être parce que tous les intervenants participant au processus d'enquête – intervieweurs, répondants, etc. – apprennent comment recueillir et fournir des renseignements de bonne qualité pendant qu'ils participent à l'enquête. Bien que nous n'ayons pas observé d'effets saisonniers pour ce qui est du nombre d'incohérences, le nombre d'incohérences indiquait une erreur de mesure non négligeable dans les données, ce qui veut dire qu'un des deux indicateurs a été mal déclarés, voire les deux.

Tableau 3.3
Incohérences 2005 – 2009, % – données regroupées

	EC	EI	CE	CI	IE	IC
BIT – Auto-évaluation	0,97	1,72	0,44	13,02	0,17	5,80
BIT – Rétrospective	1,14	2,06	5,22	16,76	1,00	5,76
Auto-évaluation – Rétrospective	0,92	1,62	6,03	8,73	1,00	0,89

EC = Considéré comme ayant un emploi avec le premier indicateur, mais en chômage avec le deuxième indicateur.

EI = Considéré comme ayant un emploi avec le premier indicateur, mais inactif avec le deuxième indicateur.

CE = Considéré comme en chômage avec le premier indicateur, mais ayant un emploi avec le deuxième indicateur.

CI = Considéré comme en chômage avec le premier indicateur, mais inactif avec le deuxième indicateur.

IE = Considéré comme inactif avec le premier indicateur, mais ayant un emploi avec le deuxième indicateur.

IC = Considéré comme inactif avec le premier indicateur, mais en chômage avec le deuxième indicateur.

Cependant, les incohérences ressorties des tableaux 3.2 et 3.3 peuvent également se produire parce que les trois indicateurs sont exposés à une erreur de mesure. Des études antérieures ont examiné les causes de la mauvaise perception de l'état sur le marché du travail et conclu que la perception est influencée par des facteurs sociaux, démographiques, économiques et institutionnels (par exemple Richiardi 2002). Les incohérences entre les deux auto-évaluations (réelle et rétrospective) peuvent être en grande partie

attribuables à l'érosion de la mémoire (Bound, Brown et Mathiowetz 2001). Enfin, la plus forte cohérence entre les indicateurs d'auto-évaluation suggère la possibilité d'erreurs de mesure corrélées.

Le tableau 3.4 indique les probabilités de transition trimestrielles parmi les trois états de la population active, du premier au deuxième trimestre des années 2005 à 2009 avec les trois indicateurs. L'indicateur du BIT décrit un marché du travail beaucoup plus dynamique, en particulier pour les répondants en chômage, que les indicateurs basés sur une auto-évaluation et une rétrospective. Cette différence est une autre preuve qui signale une erreur de mesure dans les données. D'après la littérature existante, nous savons que même de petits degrés d'erreur de classification peuvent entraîner un biais grave de l'estimation des probabilités de transition (Hagenaars 1994; Pavlopoulos, Muffles et Vermunt 2012). Si les erreurs ne sont pas corrélées au fil du temps, nous pouvons nous attendre à observer un marché du travail plus dynamique que le vrai marché, et le contraire si la corrélation des erreurs au fil du temps existe également.

Le tableau 3.5 compare les mouvements bruts observés, par exemple du premier au deuxième trimestre de 2005, selon le sexe et l'âge. On a obtenu les trois intervalles d'âge en divisant les échantillons en trois groupes de dimensions égales (c'est-à-dire 33^e et 66^e centiles). Plus précisément, pour l'année 2005, à l'âge 1, nous trouvons des répondants de 16 à 36 ans; à l'âge 2, ils ont de 36 à 55 ans, et à l'âge 3, ils ont de 56 à 75 ans. On constate que les femmes sont plus dynamiques que les hommes, en particulier pour ce qui est du chômage. Lorsqu'elles quittent une situation de chômage, les femmes ont tendance à quitter le marché du travail plus souvent que de commencer un emploi. Il y a également des variations importantes des mouvements bruts observés au fil du temps. Les répondants les plus âgés étaient plus stables lorsqu'ils étaient inactifs et avaient de plus fortes probabilités de quitter le marché du travail que de commencer un emploi après avoir occupé un emploi. Les jeunes répondants avaient de plus faibles probabilités de quitter une situation de chômage et d'inactivité en trouvant un emploi que ceux du deuxième groupe d'âge. Ces constatations portent à croire que le sexe et l'âge devraient être inclus comme covariables dans notre modèle, afin d'estimer les mouvements bruts corrigés sur le marché du travail.

Tableau 3.4

Mouvements bruts observés, du trimestre I au trimestre II de 2005 à 2009, %, Bureau international du travail (BIT), indicateur auto-évaluation (A) et rétrospective (R)

		EE	EC	EI	CE	CC	CI	IE	IC	II
2005	BIT	96,49	0,87	2,63	18,97	50,50	30,53	1,49	1,99	96,52
	A	96,99	1,33	1,69	15,32	69,85	14,83	1,29	1,50	97,21
	R	95,32	2,10	2,58	20,96	59,56	19,48	1,96	2,22	95,81
2006	BIT	96,13	0,78	3,09	20,40	45,21	34,39	2,42	1,74	95,84
	A	96,11	1,74	2,16	19,84	63,66	16,50	1,88	1,75	96,37
	R	95,55	1,72	2,73	17,93	66,57	15,50	2,00	1,75	96,25
2007	BIT	96,22	0,68	3,10	21,45	40,41	38,14	2,21	1,78	96,02
	A	96,08	1,74	2,16	19,84	63,66	16,50	1,88	1,75	96,37
	R	95,66	1,78	2,56	19,95	60,67	19,38	2,26	1,93	95,80
2008	BIT	97,05	0,80	2,16	19,82	48,50	31,68	1,87	1,87	96,26
	A	96,92	1,54	1,53	15,25	70,84	13,92	1,56	1,69	96,75
	R	95,76	2,13	2,11	19,04	62,60	18,36	2,02	2,26	95,72
2009	BIT	96,58	0,88	2,54	18,41	48,10	33,49	2,08	1,83	96,09
	A	96,14	1,76	2,10	15,17	70,09	14,75	1,59	1,61	96,80
	R	95,45	1,88	2,66	16,88	67,15	15,97	1,78	1,89	96,33

EE = Occupant un emploi au cours des deux trimestres.

EC = Occupant un emploi au premier trimestre, et en chômage au deuxième.

EI = Occupant un emploi au premier trimestre, et inactif au deuxième.

CE = En chômage au premier trimestre, et occupant un emploi au deuxième.

CC = En chômage au cours des deux trimestres.

CI = En chômage au premier trimestre, et inactif au deuxième.

IE = Inactif au cours du premier trimestre, et occupant un emploi au deuxième.

IC = Inactif au cours du premier trimestre, et au chômage au deuxième.

II = Inactif au cours des deux trimestres.

Tableau 3.5

Mouvements bruts observés du trimestre I au trimestre II de 2005, selon le sexe et l'âge, %, Bureau international du travail (BIT), indicateurs auto-évaluation (A) et rétrospective (R)

		EE	EC	EI	CE	CC	CI	IE	IC	II
Hommes	BIT	97,20	0,78	2,02	22,73	51,60	25,68	1,93	2,07	96,00
	A	97,63	1,08	1,29	18,97	73,80	7,23	1,36	1,10	97,53
	R	96,13	1,84	2,03	26,14	65,27	8,60	2,13	1,50	96,37
Femmes	BIT	95,43	1,01	3,57	15,70	49,31	34,99	1,23	1,98	96,79
	A	96,00	1,69	2,31	11,93	65,73	22,34	1,26	1,81	96,93
	R	94,14	2,46	3,40	16,24	53,56	30,19	1,86	2,71	95,43
Âge 1	BIT	88,27	0,46	11,27	21,16	27,50	51,35	0,26	0,06	99,67
	A	89,66	0,56	9,78	10,20	60,09	29,71	0,31	0,10	99,60
	R	83,36	0,45	16,19	20,78	42,54	36,68	0,51	0,13	99,36
Âge 2	BIT	97,65	0,55	1,80	21,62	43,01	35,37	2,72	2,95	94,33
	A	97,87	0,92	1,20	16,83	64,65	18,52	2,52	2,61	94,87
	R	97,04	1,23	1,74	24,60	53,42	21,98	4,05	4,24	91,70
Âge 3	BIT	96,18	1,32	2,50	17,54	51,14	31,32	3,81	6,75	89,44
	A	96,83	1,89	1,28	14,77	71,97	13,27	3,17	4,82	92,01
	R	94,82	3,29	1,89	19,62	63,60	16,78	4,52	6,68	88,80

EE = Occupant un emploi au cours des deux trimestres.

EC = Occupant un emploi au premier trimestre, et en chômage au deuxième.

EI = Occupant un emploi au premier trimestre, et inactif au deuxième.

CE = En chômage au premier trimestre, et occupant un emploi au deuxième.

CC = En chômage au cours des deux trimestres.

CI = En chômage au premier trimestre, et inactif au deuxième.

IE = Inactif au cours du premier trimestre, et occupant un emploi au deuxième.

IC = Inactif au cours du premier trimestre, et en chômage au deuxième.

II = Inactif au cours des deux trimestres.

4 Résultats : Comparaisons des MMCL mixte et standard

Nous estimons plusieurs spécifications des MMCL standard et mixtes. Le modèle standard comporte deux composantes : structure, décrivant la dynamique réelle parmi les variables latentes (vrais états) par une chaîne markovienne de premier ordre; et mesure, qui apparie chaque variable latente à ses indicateurs (états observés sur le marché du travail). Certaines restrictions comportant de l'information a priori et/ou des suppositions sont imposées aux paramètres de la composante de mesure, en fonction de la preuve ressortie des données observées (incohérences et transitions) et des conclusions de la littérature sur la méthodologie de l'enquête et la psychologie cognitive sur le mécanisme entraînant des erreurs. Seulement quatre des neuf groupes de renouvellement fournissant de l'information visant une année civile ont été interviewés à tous les trimestres, et nous avons les trois indicateurs des états sur le marché du travail seulement pour deux de ces groupes (voir le tableau 3.1). Pour les deux autres groupes, nous ne recueillons pas l'information avec la question rétrospective. La tendance de l'information manquante en raison du plan par renouvellement de l'enquête est incluse dans les MMCL estimés comme des données manquantes sur une base aléatoire.

Tous les modèles estimés partagent les caractéristiques suivantes : les transitions réelles suivent une chaîne markovienne de premier ordre. En raison du plan d'enquête, aucune personne n'était observée pendant trois cycles consécutifs, c'est-à-dire qu'une chaîne markovienne de deuxième ordre ne peut pas être estimée, puisque les statistiques suffisantes relatives sont absentes. Cependant, bien que l'état sur le marché du travail au cours d'un trimestre donné risque fort de toucher l'état au cours du trimestre subséquent, il est beaucoup moins probable que l'incidence soit significative au bout de deux trimestres. On présume que les erreurs de classification sont constantes au fil du temps pour chaque indicateur; la supposition de l'ECI est incluse. L'ajustement du modèle est évaluée par l'indice BIC (pour *Bayesian information criterion*) en raison de la grande taille de l'échantillon (moyenne de 250 000 unités par année; voir le tableau 4.1).

La spécification d'un MMCL mixte est également recommandée par le fait que l'échantillon peut renfermer plusieurs groupes de répondants ayant des comportements différents sur le marché du travail. Comme susmentionné, la littérature récente démontre qu'en ne tenant pas compte de l'hétérogénéité non observée des transitions pour estimer les MMCL, on risque d'entraîner des estimations biaisées de l'erreur de mesure (Magidson et coll. 2007). De plus, un MMCL mixte peut améliorer l'ajustement des données.

Nous estimons un MMCL mobile-stable en supposant des erreurs de mesure constantes dans les deux groupes latents. Il convient de souligner que tous les modèles estimés ont été relevés, et que, pour réduire le risque de détection des maxima locaux, l'estimation a été effectuée à plusieurs reprises avec différents ensembles de valeurs de départ. Le logiciel Latent Gold 4.5 a été mis en œuvre (Vermunt et Magidson 2008).

Le tableau 4.1 compare les MMCL mixte et standard adaptés à nos cinq échantillons de données, qui visent les années de 2005 à 2009, et au moyen de l'indice BIC. Le modèle mixte démontre une meilleure correspondance pour tous les échantillons. Le tableau 4.2 indique les pourcentages de personnes mobiles et stables au premier trimestre de 2005, ainsi que la répartition de deux groupes non observés au premier trimestre de chaque année. De toute évidence, l'hétérogénéité non observée est fortement corrélée à l'état initial et, comme il fallait s'y attendre, les personnes stables occupent un emploi ou sont inactives, c'est-à-dire qu'un très faible pourcentage d'entre elles sont en chômage.

Tableau 4.1
Comparaison des MMCL standard et mixte : indice BIC

Année	n	Standard	Mixte
2005	220 051	650 241	649 401
2006	206 037	587 794	587 058
2007	274 484	748 788	748 654
2008	277 363	667 399	666 335
2009	274 723	747 997	746 991

Tableau 4.2
MMCL mixte : proportion de personnes mobiles et stables et répartition de l'état initial en 2005, trimestre I, %

	Proportion	E	C	I
Mobiles	10,23	39,85	39,09	21,06
Stables	81,79	41,79	3,36	54,85

E = Personnes ayant un emploi, C = Personnes en chômage, I = Personnes inactives.

Comme le démontrent les données dans les tableaux 4.3-4.5 (la composition du marché du travail, les transitions estimées et les erreurs de mesure estimées démontrent la même tendance au cours des trois autres trimestres de chaque année), la meilleure correspondance aux données du modèle mixte est entièrement attribuable aux différents taux de transition estimés; la composition du marché du travail et les erreurs de mesure estimées sont les mêmes dans les deux modèles. Ce résultat vient contredire celui qui avait été obtenu par Magidson et coll. (2007), qui ont comparé le modèle des personnes mobiles-stables et le MMCL standard appliqué aux transitions sur le marché du travail de l'enquête sur la population active. Les auteurs susmentionnés ont constaté que le MMCL mixte convient mieux aux données que le MMCL standard, et que ce dernier, qui ne tient pas compte de l'hétérogénéité non observée, surestime le degré d'erreur de mesure en ce qui concerne le modèle des personnes mobiles-stables. Plus précisément, les auteurs

susmentionnés ont utilisé des résultats simulés pour estimer une infraction aux probabilités de transition homogènes, de sorte que l'hétérogénéité corrélée avec l'état initial produise des estimations exagérées des erreurs de mesure dans un MMCL standard.

Tableau 4.3
Comparaison des MMCL standard et mixte : composition du marché du travail au trimestre I de 2005, %

		E	C	I
2005	Standard	41,67	7,00	51,33
	Mixte	41,59	7,02	51,39

E = Personnes ayant un emploi, C = Personnes en chômage, I = Personnes inactives.

Tableau 4.4
Comparaison des MMCL standard et mixte : transitions estimées du trimestre I au trimestre II de 2005 – 2009, %

		EE	EC	EI	CE	CC	CI	IE	IC	II
2005	Standard	97,36	1,32	1,32	15,59	76,18	8,23	0,57	0,74	98,69
	Mixte	96,46	1,68	1,86	19,61	69,65	10,74	0,91	1,09	98,00
2006	Standard	96,75	1,68	1,56	19,52	71,27	9,21	1,01	0,99	90,00
	Mixte	96,22	1,92	1,87	22,11	66,96	10,93	1,25	1,22	97,54
2007	Standard	96,69	1,67	1,64	18,84	70,56	10,60	1,01	0,99	98,00
	Mixte	96,42	1,80	1,78	20,22	67,80	11,98	1,10	1,45	95,45
2008	Standard	97,56	1,41	1,03	15,86	79,73	4,42	0,53	0,62	98,85
	Mixte	96,45	1,89	1,66	19,56	73,25	7,19	0,83	0,89	98,28
2009	Standard	96,85	1,71	1,44	14,04	75,33	9,63	1,04	1,01	97,95
	Mixte	96,27	1,95	1,78	17,09	71,16	11,75	1,30	1,22	97,48

EE = Occupant un emploi au cours des deux trimestres.

EC = Occupant un emploi au premier trimestre, et en chômage au deuxième.

EI = Occupant un emploi au premier trimestre, et inactif au deuxième.

CE = En chômage au premier trimestre, et occupant un emploi au deuxième.

CC = En chômage au cours des deux trimestres.

CI = En chômage au premier trimestre, et inactif au deuxième.

IE = Inactif au cours du premier trimestre, et occupant un emploi au deuxième.

IC = Inactif au cours du premier trimestre, et en chômage au deuxième.

II = Inactif au cours des deux trimestres.

Tableau 4.5
Comparaison des MMCL standard et mixte : erreurs de mesure estimées au trimestre I de 2005 – 2009, %, indicateur du BIT

		EE	EC	EI	CE	CC	CI	IE	IC	II
2005	Standard	99,82	0,01	0,17	6,17	45,04	48,80	0,89	0,50	98,61
	Mixte	99,82	0,01	0,17	6,16	45,06	48,78	0,90	0,51	98,59
2006	Standard	99,83	0,01	0,16	6,50	41,92	51,58	0,75	0,45	98,80
	Mixte	99,87	0,01	0,13	5,17	37,28	57,55	0,68	0,40	98,92
2007	Standard	99,75	0,01	0,24	6,84	39,83	53,34	0,75	0,47	98,79
	Mixte	99,75	0,01	0,24	6,77	39,92	53,31	0,77	0,47	98,76
2008	Standard	99,83	0,01	0,17	3,81	42,45	53,74	0,61	0,38	99,02
	Mixte	99,83	0,01	0,17	3,82	42,41	53,76	0,62	0,38	99,00
2009	Standard	95,34	0,98	3,68	18,30	41,17	40,53	2,06	1,61	96,33
	Mixte	95,22	2,34	2,44	15,60	68,02	16,37	1,74	2,14	96,13

EE = Réellement occupé et réputé occupé par l'indicateur du BIT.

EC = Réellement occupé, mais réputé en chômage par l'indicateur du BIT.

EI = Réellement occupé, mais réputé inactif par l'indicateur du BIT.

CE = Réellement en chômage, mais réputé occupé par l'indicateur du BIT.

CC = Réellement en chômage et réputé en chômage par l'indicateur du BIT.

CI = Réellement en chômage, mais réputé inactif par l'indicateur du BIT.

IE = Réellement inactif, mais réputé occupé par l'indicateur du BIT.

IC = Réellement inactif, mais réputé en chômage par l'indicateur du BIT.

II = Réellement inactif et réputé inactif par l'indicateur du BIT.

Le modèle des personnes mobiles-stables décrit un marché du travail plus dynamique, en particulier pour les répondants en chômage : la probabilité de demeurer en chômage pendant le trimestre est plus faible que prévu par le modèle standard.

5 Résultats : MMCL mixte avec covariables et erreurs de mesure corrélées

Les résultats présentés à la section précédente indiquaient qu'un MMCL mixte convient mieux à nos données. Comme le MMCL standard, le MMCL mixte tient compte de la mauvaise classification et de la tendance des données manquantes à présumer de cette dernière de façon aléatoire, et il comprend également l'hétérogénéité non observée. La supposition que les données sont manquantes de façon aléatoire s'explique par le fait que chaque groupe de renouvellement est observé au cours de deux trimestres, mais pas au cours des deux trimestres subséquents, et que ces données sont manquantes en raison du plan et qu'elles ne dépendent pas de l'état réel ou déclaré des répondants ou d'autres variables non observées. Pour estimer nos modèles, nous avons utilisé simultanément de l'information de tous les groupes de renouvellement, c'est-à-dire une technique de calcul du maximum de vraisemblance à information complète. Les données des mouvements bruts observés, en particulier le fait que la mobilité observée est très différente entre les sexes et les âges (tableau 3.5), indiquaient que l'estimation d'un MMCL mixte avec ces deux covariables avait une incidence sur les transitions latentes.

Divers modèles ont été estimés avec la caractéristique commune suivante : les transitions mobile-stable et les transitions latentes suivent une chaîne markovienne de premier ordre. Afin de préciser le modèle de mesure, les facteurs suivants ont été pris en considération : (i) la réponse à la question sur l'état autoévalué sur le marché du travail est donnée pendant la même interview une fois que les répondants ont répondu aux questions sur lesquelles repose l'indicateur du BIT; (ii) toutefois, l'indicateur du BIT est déterminé par l'ISTAT en fonction des réponses données à une série de questions respectant les lignes directrices du BIT, tandis que A représente les auto-évaluations des répondants : il est plausible que les répondants ne soient pas au courant de la classification de l'ISTAT; (iii) l'indicateur A et l'indicateur découlant de l'interrogation rétrospective décrivent un marché du travail plus stable que celui du BIT et montrent le plus haut niveau de cohérence : les répondants peuvent être influencés par les réponses données au trimestre précédent; (iv) l'information pour R est recueillie un an après les réponses au BIT et A; (v) pour les personnes qui sont dans un état stable, la déclaration correcte de l'état sur le marché du travail est une tâche cognitive plus facile que pour les personnes qui traversent au moins un changement, et elle peut donc indiquer de plus fortes probabilités de donner des réponses incorrectes.

Parmi les diverses spécifications possibles, le modèle convenant le mieux, pour toutes les années analysées, consistait à présumer que les personnes stables déclarent leur état sur le marché du travail correctement et que, pour les personnes mobiles, les erreurs de mesure sont constantes au fil du temps et que les deux indicateurs basés sur l'auto-évaluation, A et R, sont corrélés, c'est-à-dire qu'un effet direct entre ces deux indicateurs est intégré à la spécification du modèle. (Tous les modèles estimés ont été cernés et, pour éviter les maxima locaux, une estimation a été effectuée plusieurs fois avec différents ensembles de valeurs de départ; pour estimer des modèles plus parcimonieux, les trois interactions des variables ont été

établies à 0.) À titre d'exemple, les tableaux 5.1 à 5.3 indiquent quelques-uns des résultats des estimations : la composition du marché du travail et les mouvements estimés pour l'ensemble de la population, les personnes mobiles et stables collectivement (l'ensemble complet des résultats des estimations peut être obtenu auprès des auteurs) et les erreurs de mesure estimées. En moyenne, au cours des cinq ans, le pourcentage de personnes mobiles était de 17,69.

Tableau 5.1
Composition estimée du marché du travail au trimestre I de 2005 – 2009, %

	2005	2006	2007	2008	2009
E	42,01	42,36	40,72	40,92	40,00
C	5,93	5,64	5,75	5,27	6,46
I	52,07	52,00	53,53	53,81	53,53

E = Personnes ayant un emploi, C = Personnes en chômage, I = Personnes inactives.

Tableau 5.2
Mouvements bruts estimés du trimestre I au trimestre II de 2005 - 2009, %, erreurs types entre parenthèses

	EE	EC	EI	CE	CC	CI	IE	IC	II
2005	96,70 (0,0017)	1,60 (0,0012)	1,61 (0,0012)	17,41 (0,0133)	71,80 (0,0142)	10,78 (0,0079)	0,97 (0,0013)	0,70 (0,0011)	98,29 (0,0017)
2006	96,10 (0,0027)	1,93 (0,0020)	1,93 (0,0020)	19,16 (0,0112)	67,04 (0,0150)	13,80 (0,0136)	1,71 (0,0011)	0,89 (0,0015)	97,41 (0,0018)
2007	96,30 (0,0023)	1,79 (0,0016)	1,89 (0,0017)	18,11 (0,0145)	67,95 (0,0158)	13,94 (0,0094)	1,42 (0,0018)	1,24 (0,0018)	97,34 (0,0025)
2008	96,88 (0,0037)	1,77 (0,0027)	1,35 (0,0028)	18,00 (0,0118)	74,57 (0,0157)	7,43 (0,0138)	1,61 (0,0013)	1,03 (0,0017)	97,37 (0,0020)
2009	96,50 (0,0024)	1,83 (0,0019)	1,62 (0,0016)	15,04 (0,0153)	71,62 (0,0168)	13,35 (0,0092)	1,55 (0,0019)	1,10 (0,0014)	97,35 (0,0024)

EE = Occupant un emploi au cours des deux trimestres.

EC = Occupant un emploi au premier trimestre, et en chômage au deuxième.

EI = Occupant un emploi au premier trimestre, et inactif au deuxième.

CE = En chômage au premier trimestre, et occupant un emploi au deuxième.

CC = En chômage pendant les deux trimestres.

CI = En chômage au premier trimestre, et inactif au deuxième.

IE = Inactif au cours du premier trimestre, et occupant un emploi au deuxième.

IC = Inactif au cours du premier trimestre, et en chômage au deuxième.

II = Inactif pendant les deux trimestres.

Tableau 5.3a
Erreurs de mesure estimées de 2005 – 2009, indicateur du BIT, %, erreurs types entre parenthèses

	EE	EC	EI	CE	CC	CI	IE	IC	II
2005	99,75 (0,0002)	0,02 (0,0001)	0,23 (0,0001)	0,93 (0,0028)	89,72 (0,0050)	9,36 (0,0051)	0,97 (0,0004)	1,04 (0,0003)	98,00 (0,0005)
2006	99,75 (0,0007)	0,01 (0,0004)	0,24 (0,0005)	1,17 (0,0025)	89,39 (0,0042)	9,44 (0,0035)	0,55 (0,0003)	0,99 (0,0002)	98,46 (0,0004)
2007	99,82 (0,0002)	0,01 (0,0001)	0,24 (0,0002)	0,84 (0,0028)	88,28 (0,0050)	10,88 (0,0051)	0,58 (0,0004)	0,87 (0,0003)	98,55 (0,0005)
2008	99,44 (0,0007)	0,10 (0,0004)	0,46 (0,0005)	1,16 (0,0025)	89,36 (0,0042)	9,48 (0,0035)	0,57 (0,0003)	1,38 (0,0002)	90,05 (0,0004)
2009	99,77 (0,0001)	0,01 (0,0000)	0,22 (0,0001)	0,43 (0,0025)	88,98 (0,0038)	10,57 (0,0039)	0,33 (0,0003)	0,86 (0,0002)	98,79 (0,0003)

EE = Réellement occupé et réputé occupé par l'indicateur du BIT.

EC = Réellement occupé, mais réputé en chômage par l'indicateur du BIT.

EI = Réellement occupé, mais réputé inactif par l'indicateur du BIT.

CE = Réellement en chômage, mais réputé occupé par l'indicateur du BIT.

CC = Réellement en chômage et réputé en chômage par l'indicateur du BIT.

CI = Réellement en chômage, mais réputé inactif par l'indicateur du BIT.

IE = Réellement inactif, mais réputé occupé par l'indicateur du BIT.

IC = Réellement inactif, mais réputé en chômage par l'indicateur du BIT.

II = Réellement inactif et réputé inactif par l'indicateur du BIT.

Tableau 5.3b
Erreurs de mesure estimées de 2005 – 2009, indicateurs A et R, %, erreurs types entre parenthèses

État réel		AR								
		EE	EC	EI	CE	CC	CI	IE	IC	II
2005	E	94,83 (0,0008)	1,17 (0,0006)	2,28 (0,0005)	0,22 (0,0002)	0,18 (0,0001)	0,11 (0,0002)	0,44 (0,0003)	0,07 (0,0004)	0,70 (0,0003)
	C	0,01 (0,0001)	0,00 (0,0001)	0,00	0,97 (0,0006)	97,16 (0,0008)	1,11 (0,0004)	0,09 (0,0009)	0,31 (0,0004)	0,35 (0,0003)
	I	0,00	0,00	0,01 (0,0001)	0,12 (0,0005)	0,70 (0,0009)	0,70 (0,0008)	0,78 (0,0004)	0,98 (0,0006)	96,72 (0,0008)
2006	E	94,86 (0,0052)	0,96 (0,0006)	2,21 (0,0005)	0,16 (0,0001)	0,11 (0,0002)	0,10 (0,0009)	0,45 (0,0001)	0,06 (0,0004)	1,06 (0,0003)
	C	0,00	0,01 (0,0001)	0,00	0,86 (0,0001)	97,98 (0,0006)	0,50 (0,0001)	0,11 (0,0002)	0,32 (0,0003)	0,22 (0,0003)
	I	0,01 (0,0001)	0,00	0,01 (0,0001)	0,13 (0,0006)	0,82 (0,0005)	0,74 (0,0004)	0,71 (0,0004)	0,74 (0,0001)	96,83 (0,0005)
2007	E	95,17 (0,0009)	1,06 (0,0003)	1,06 (0,0005)	0,16 (0,0002)	0,11 (0,0004)	0,10 (0,0005)	0,45 (0,0006)	0,06 (0,0004)	0,82 (0,0004)
	C	0,00	0,01 (0,0001)	0,00	0,90 (0,0005)	97,74 (0,0009)	0,73 (0,0003)	0,09 (0,0005)	0,31 (0,0004)	0,21 (0,0002)
	I	0,01 (0,0001)	0,01 (0,0001)	0,01 (0,0001)	0,15 (0,0005)	0,59 (0,0006)	0,66 (0,0008)	1,10 (0,0004)	0,89 (0,0004)	96,59 (0,0020)
2008	E	94,65 (0,0006)	1,48 (0,0009)	1,83 (0,0005)	0,16 (0,0003)	0,02 (0,0006)	0,14 (0,0004)	0,72 (0,0003)	0,04 (0,0004)	0,96 (0,0002)
	C	0,00	0,03 (0,0001)	0,00	1,32 (0,0002)	97,39 (0,0010)	0,82 (0,0009)	0,05 (0,0005)	0,33 (0,0004)	0,05 (0,0004)
	I	0,01 (0,0001)	0,02 (0,0001)	0,01 (0,0001)	0,17 (0,0009)	0,45 (0,0005)	1,34 (0,0003)	1,05 (0,0006)	1,50 (0,0004)	95,45 (0,0003)
2009	E	96,11 (0,0004)	0,65 (0,0002)	1,21 (0,0001)	0,12 (0,0002)	0,24 (0,0003)	0,10 (0,0008)	0,42 (0,0009)	0,10 (0,0008)	1,04 (0,0009)
	C	0,01 (0,0001)	0,01 (0,0001)	0,00	0,59 (0,0004)	98,23 (0,0004)	0,55 (0,0002)	0,08 (0,0005)	0,26 (0,0006)	0,25 (0,0006)
	I	0,01 (0,0001)	0,00	0,01 (0,0001)	0,08 (0,0004)	0,76 (0,0002)	0,52 (0,0002)	0,74 (0,0004)	0,78 (0,0003)	97,08 (0,0008)

E = Personnes ayant un emploi, C = Personnes en chômage, I = Personnes inactives.

EE = Classifié comme occupant un emploi par les indicateurs auto-évaluation et rétrospective.

EC = Classifié comme occupant un emploi par l'indicateur auto-évaluation et en chômage par l'indicateur rétrospective.

EI = Classifié comme occupant un emploi par l'indicateur auto-évaluation et en inactif par l'indicateur rétrospective.

CE = Classifié comme en chômage par l'indicateur auto-évaluation et occupant un emploi par l'indicateur rétrospective.

CC = Classifié comme en chômage par les indicateurs auto-évaluation et rétrospective.

CI = Classifié comme en chômage par l'indicateur auto-évaluation et inactif par l'indicateur rétrospective.

IE = Classifié comme inactif par l'indicateur auto-évaluation et occupant un emploi par l'indicateur rétrospective.

IC = Classifié comme inactif par l'indicateur auto-évaluation et en chômage par l'indicateur rétrospective.

II = Classifié comme inactif par les indicateurs auto-évaluation et rétrospective.

La composition estimée du marché du travail au premier trimestre, comparativement à la composition observée (tableau 3.2), démontre un pourcentage de chômage légèrement inférieur à celui obtenu au moyen des deux indicateurs auto-évaluation, et plus élevé que celui de l'indicateur du BIT.

Les transitions estimées décrivent un marché du travail plus stable que celui qui a été observé avec les trois indicateurs, sauf deux transitions (voir le tableau 3.4). Les mouvements bruts estimés sont beaucoup plus semblables à ceux observés avec les questions reposant sur l'auto-évaluation et la rétrospective que ceux observés avec l'indicateur du BIT. Ce phénomène se manifeste également dans le cas de l'erreur de mesure estimée (tableau 5.3). Une objection immédiate à ce résultat serait que nous avons utilisé deux indicateurs très semblables (les deux auto-évaluations), ainsi qu'un troisième très différent (BIT). En fait, un résultat semblable – des erreurs de mesure plus faibles pour l'auto-évaluation que pour l'indicateur du BIT – a été obtenu en estimant un MMCL avec seulement deux indicateurs par variable latente : le BIT et l'auto-évaluation.

6 Conclusion

Le présent document propose une approche basée sur les classes latentes pour corriger les mouvements bruts à partir des erreurs corrélées. On met l'accent sur la capacité de tenir compte des erreurs de classification à l'étendue des données sur les panels, en raison du plan par renouvellement de l'enquête, qui entraîne des tendances aux données manquantes et à une hétérogénéité non observée.

L'approche des classes latentes a été appliquée aux transitions sur le marché du travail italien en fonction des trois états habituels (occupant un emploi, en chômage et inactif). Les données portent sur les années 2005 à 2009 et ont été recueillies dans le cadre de l'enquête italienne continue sur la population active auprès d'un échantillon de ménages italiens avec un plan par renouvellement de type 2-2-2 d'un trimestre à un autre. L'information sur l'état de la population active au cours d'un trimestre de référence a été recueillie à trois reprises : (i) les répondants ont été classifiés comme occupant un emploi, en chômage ou inactifs, conformément à la définition du Bureau international du travail en fonction des réponses à un groupe de questions en particulier; (ii) on a demandé aux répondants de se classifier comme occupant un emploi, en chômage ou inactifs (c'est-à-dire l'état auto-évalué); (iii) une question rétrospective visait à déterminer l'état sur le marché du travail un an plus tôt. Autrement dit, trois indicateurs de l'état sur le marché du travail étaient disponibles. Les trois indicateurs donnaient des descriptions très différentes du marché du travail italien, révélant un degré d'incohérence significatif. Ce phénomène indique une erreur de mesure dans les données.

Le modèle convenant le mieux était un MMCL reposant sur les personnes mobiles-stables, où les transitions latentes sur le marché du travail suivent une chaîne markovienne de premier ordre, les personnes stables déclarent toujours correctement leur état sur le marché du travail; pour les personnes mobiles, les erreurs de mesure étaient constantes au fil du temps et corrélées aux deux indicateurs d'auto-évaluation; le sexe et l'âge des répondants étaient inclus comme covariables; le plan par renouvellement de l'enquête était traité comme de l'information manquante sur une base aléatoire. Le modèle corrige les mouvements bruts observés vers un marché du travail plus stable et estime que l'indicateur de l'état sur le marché du travail basé sur la définition du BIT est touché par le plus grand degré d'erreur de mesure.

Deuxième conclusion : en cas d'hétérogénéité non observée, un MMCL mixte convient mieux aux données que le MMCL standard. Cette conclusion cadre avec d'autres rapports (par exemple Magidson et coll. 2007). Cependant, dans notre cas, les deux modèles estiment la même quantité d'erreur de mesure, la différence d'ajustement étant attribuable aux mouvements estimés. Au lieu de cela, les auteurs susmentionnés ont découvert une surestimation de l'erreur de mesure lorsque l'hétérogénéité non observée n'était pas prise en compte.

Un dernier facteur pris en considération a trait au plan d'échantillonnage de l'enquête, qui comporte deux degrés, comme indiqué à la section 3. Dans nos analyses, nous n'avons pas tenu compte du plan d'échantillonnage complexe, mais avons estimé les mouvements bruts de la population longitudinale fournie par l'*Italian Institute of Statistics*. Dans les recherches à venir, il serait intéressant de comparer comment les résultats peuvent être touchés par l'intégration de méthodes d'enquêtes à des échantillons complexes au moyen de notre stratégie d'estimation. Lu et Lohr (2010) ont déjà abordé le sujet d'une perspective intéressante.

Bibliographie

- Bartolucci, F., Lupporelli, M. et Montanari, A. (2009). Latent Markov model for longitudinal binary data: An application to the performance evaluation of nursing homes. *Annals of Applied Statistics*, 3, 611-636.
- Bassi, F., et Trivellato, U. (2009). A latent class approach for estimating gross flows in the presence of correlated classification errors. Dans *Methodology of Longitudinal Surveys*, (Éd., P. Lynn), Chichester: Wiley, 367-380.
- Bassi, F., Padoan, A. et Trivellato, U. (2012). Inconsistencies in reported characteristics among employed stayers. *Statistica*, 1, 93-109.
- Bassi, F., Torelli, N. et Trivellato, U. (1998). Stratégies de collecte de données et de modélisation dans l'estimation de flux bruts relatifs à la population active entachés d'erreurs de classification. *Techniques d'enquête*, 24, 2, 117-132. Article accessible à l'adresse <http://www.statcan.gc.ca/pub/12-001-x/1998002/article/4348-fra.pdf>.
- Bassi, F., Croon, M., Hagenaars, J.A. et Vermunt, J.K. (2000). Estimating true changes when categorical panel data are affected by correlated and uncorrelated classification errors. An application to unemployment data. *Sociological Methods and Research*, 29, 230-268.
- Biemer, P.P., et Bushery, J.M. (2000). Validité de l'analyse markovienne de structure latente pour l'estimation de l'erreur de classification des données sur la population active. *Techniques d'enquête*, 26, 2, 157-171. Article accessible à l'adresse <http://www.statcan.gc.ca/pub/12-001-x/2000002/article/5534-fra.pdf>.
- Boschetto, B., Discenza, A.R., Lucarelli, C., Rosati, S. et Fiori, F. (2009). Longitudinal data for the analysis of Italian labor market flows. *Italian Journal of Applied Statistics*, 22, 129-150.
- Bound, M., Brown C. et Mathiowetz N.A. (2001). Measurement error in survey data. Dans *Handbook of Econometrics*, (Éds., J.J. Heckman et E. Leamer), Amsterdam: Elsevier, 3705-3843.
- Clark, K., et Summers, L.H. (1979). Labour market dynamics and unemployment: A reconsideration. *Brooking Papers on Economic Activity*, 1, 13-69.
- Dayton, C.M., et McReady, G.B. (1988). Concomitant-variable latent-class models. *Journal of the American Statistical Association*, 83, 173-178.
- Flinn, C.J., et Heckman, J.J. (1983). Are unemployment and out of the labour force behaviourally distinct market states? *Journal of Labour Economics*, 1, 28-42.
- Gonul, F. (1992). New evidence on whether unemployment and out of the labour force are two distinct states. *Journal of Human Resources*, 27, 329-361.
- Goodman, L. (1974). Exploratory latent structure analysis using both identifiable and unidentifiable models. *Biometrika*, 61, 215-231.
- Hagenaars, J.A. (1994). Latent variables in log-linear models of repeated observations. Dans *Latent Variable Analysis. Applications for Developmental Research*, (Éds., A. von Eye et C. Clogg), Thousand Oaks (CA): Sage, 329-352.

- Langeheine, R. (1994). Latent variable Markov models. Dans *Latent Variable Analysis. Applications for Developmental Research*, (Éds., A. von Eye et C. Clogg), Thousand Oaks (CA): Sage, 373-395.
- Lu, Y., et Lohr, S. (2010). L'estimation des flux bruts dans les enquêtes à base de sondage double. *Techniques d'enquête*, 36, 1, 13-24. Article accessible à l'adresse <http://www.statcan.gc.ca/pub/12-001-x/2010001/article/11248-fra.pdf>.
- Magidson, J., Vermunt, J.K. et Tran B. (2007). Using a mixture of latent Markov model to analyze longitudinal U.S. employment data involving measurement error. Dans *New Trends in Psychometrics*, (Éds., K. Shigemasu, A. Okada, T. Imaizumi et T. Hoshino), Tokyo: Universal Academy Press, 235-242.
- Manzoni, A., Vermunt, J.K., Luijkx, R. et Muffels, R. (2010). Memory bias in retrospectively collected employment careers: A model-based approach to correct for measurement errors. *Sociological Methodology*, 40, 39-73.
- Organisation internationale du travail (OIT) (2008). Résolution sur la mise à jour de la classification internationale type des professions, Genève. Article accessible à l'adresse <http://www.ilo.org/public/french/bureau/stat/isco/docs/resol08.pdf>.
- Paas, L.J., Vermunt, J.K. et Bijmolt, T.H. (2007). Discrete-time discrete-state latent Markov modelling for assessing and predicting household acquisitions of financial products. *Journal of the Royal Statistical Society, Series A*, 170, 955-974.
- Pavlopoulos, D., et Vermunt, J.K. (2015). Mesure de l'emploi temporaire. Les données d'enquête ou de registre disent-elles la vérité? *Techniques d'enquête*, 41, 1, 205-224. Article accessible à l'adresse <http://www.statcan.gc.ca/pub/12-001-x/2015001/article/14151-fra.pdf>.
- Pavlopoulos, D., Muffels, R. et Vermunt, J.K. (2012). How real is mobility between low pay, high pay and non-employment? *Journal of the Royal Statistical Society, Series A*, 170, 749-773.
- Poterba, J.M., et Summers, L.S. (1986). Reporting errors and labour market dynamics. *Econometrica*, 54, 1319-1338.
- Richiardi, M. (2002). What does the ECHP tell us about labour status misperception? A journey in less known regions of labour discomfort. *LABORatorio Revelli*, document de travail n° 69.
- van de Pol, F., et Langeheine, R. (1990). Mixed Markov latent class models. *Sociological Methodology*, 33, 231-247.
- Vermunt, J.K. (1997). *Log-Linear Models for Event History*. Thousand Oaks (CA): Sage.
- Vermunt, J.K. (2010). Longitudinal research using mixture models. Dans *Longitudinal Research with Latent Variables*, (Éds., K. van Montfort, J.H.L. Oud et A. Satorra), Heidelberg: Springer, 119-152.
- Vermunt, J.K., et Magidson, J. (2008). *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*. Belmont, MA: Statistical Innovations Inc.
- Vermunt, J.K., et Magidson, J. (2013). *Technical Guide for Latent Gold 5.0. Basic, Advanced, and Syntax*. Belmont, MA: Statistical Innovations Inc.

Vermunt, J.K., Tran, B. et Magidson, J. (2008). Latent class models in longitudinal research. Dans *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, (Éd., S. Menard), Burlington, MA: Elsevier, 375-385.

Wiggins, L.M. (1973). *Panel Analysis: Latent Probability for Attitude and Behavior Processes*. New York: Elsevier Scientific.