

# LES MIXTURE HIDDEN MARKOV MODELS (MODÈLES DE MARKOV CACHÉS À EFFETS MIXTES) DANS LA RECHERCHE EN SCIENCES SOCIALES

PAR | **STEF BOUWHUIS, DIMITRIS PAVLOPOULOS, MAURICIO GARNIER-VILLARREAL ET LAURA EBERLEIN**

Faculté des Sciences Sociales, Vrije Universiteit Amsterdam

## 1. INTRODUCTION

Ces dernières décennies, de plus en plus de données longitudinales de grande qualité sont devenues disponibles pour la recherche en sciences sociales. Les méthodes d'analyse des données basées sur les processus ont donc gagné en popularité (Abbott, 1983). Plutôt que de nous limiter à « seulement » examiner quelques transitions, une méthode basée sur les processus nous permet d'étudier une période entière et d'examiner l'ordre dans lequel les événements se produisent. Ces modes d'analyse des données offrent donc l'avantage de pouvoir dresser un tableau complet et détaillé des évolutions dans le parcours de vie des individus, par exemple, en termes de carrière (Fasang & Liao, 2014; Studer & Ritschard, 2016). Une méthode d'analyse basée sur les processus largement utilisés dans les sciences sociales est celle de l'analyse de séquences (AS). Dans l'AS, des séquences d'un phénomène donné sont élaborées pour chaque individu de l'étude, par exemple, des séquences de statuts sur le marché du travail. Ensuite, ces séquences sont généralement regroupées sur la base de leur similarité, à l'aide d'algorithmes de regroupement, principalement à l'aide du regroupement de Ward. Une extension de cette méthode est la *multi-channel SA* (analyse de séquences multicanaux) dans laquelle, on utilise des séquences combinées de deux phénomènes ou plus. La *Multi-channel SA* convient à l'étude de processus de différentes dimensions. Un exemple en est l'étude de Mattijssen et al. (2023), dans laquelle des données de registre néerlandaises ont été utilisées pour créer des types de séquences de sortants du système éducatif sur la base du type de contrat et des revenus (catégorisés). La *multi-channel SA*, a permis à cette recherche de décrire la riche variation des types de carrière des travailleurs entrant sur le marché du travail.

L'AS est une méthode déterministe, en ce sens que les séquences sont assignées de manière déterministe à des grappes, après quoi, il devient alors possible d'analyser les grappes auxquelles appartiennent les individus. Un autre type de modèles, dont les modèles de Markov, est de nature probabiliste (Liao et al., 2022; Vermunt, 2010a). Dans ce type de modèle, la probabilité qu'une situation se produise à l'instant  $t$  est estimée de manière modélisée, compte tenu de la situation à l'instant  $t-1$ . Un exemple en est la météo : quelle est la probabilité qu'il fasse beau lundi, en sachant qu'il faisait beau dimanche ? Le fait que cette question soit modélisée, signifie que le modèle corrige les erreurs de mesure. Le temps qu'il fait aujourd'hui est co-déterminé, mais pas parfaitement déterminé par le temps, qu'il faisait hier. Le modèle de Markov caché

(HMM) est une extension de ce modèle. *Hidden* fait référence, ici, à la nature latente du phénomène (social) dont question. Il est, en effet, possible d'observer directement s'il y a du soleil dehors. Toutefois, de nombreux phénomènes sociaux ne sont pas aussi faciles à observer directement et sont donc latents. Ce type de variables est observé au moyen d'indicateurs directement observables. Les HMM conviennent aux concepts multidimensionnels, notamment les opinions sur la démocratie. Les différentes dimensions de ce type de recherche comprennent, par exemple, une dimension libérale, utilisée pour évaluer la mesure dans laquelle les élections et la protection des minorités sont considérées comme faisant partie de la démocratie, et une dimension redistributive, qui évalue la mesure dans laquelle la redistribution des revenus est considérée comme faisant partie de la démocratie. Les individus qui obtiennent des résultats comparables au niveau des deux dimensions sont regroupés en grappes, appelées états dans les HMM. Les HMM peuvent être utilisés pour examiner comment un individu change d'état au fil du temps. Dans les HMM, les différents indicateurs sont fusionnés pour ainsi dire en une variable catégorielle latente, et nous étudions les mouvements entre les catégories de cette variable (les états). Cette méthode présente l'avantage de fournir des résultats plus clairs que l'AS multicanaux, dans laquelle les transitions sont étudiées séparément pour chacune des variables. Un autre avantage est qu'un HMM donne comme résultat une matrice de probabilités de transition, qui peut être utilisée pour comprendre la probabilité qu'un individu passe d'un état à un autre. L'analyse de séquences multicanaux n'offre pas de vue d'ensemble comparable pour pouvoir interpréter les transitions entre les états.

Un modèle de Markov caché peut être étendu à un modèle de Markov caché à effets mixtes, qui ressemble un peu à l'AS. L'extension consiste à modéliser des états (trajets) en regroupant les individus qui présentent des schémas de transition comparables entre les états de la variable latente. L'hypothèse est que les trajets auxquels appartiennent les individus constituent également une variable latente, avec comme indicateurs l'état initial et les probabilités de transition. Le modèle estime la probabilité d'appartenir à chaque trajet. Un gros avantage de la MHMM par rapport à l'AS multicanaux réside dans le fait qu'on peut ajouter au modèle des prédicteurs de regroupement dans des états et des trajets. De ce fait, nous sommes en mesure de répondre à différents types de questions causales à l'aide des MHMM.

L'objectif de cet article est de familiariser le lecteur avec les MHMM. Dans la section consacrée aux méthodes, nous discutons des étapes à suivre en tant que chercheur pour réaliser un MHMM en mettant plus particulièrement l'accent sur les choix à opérer. Dans la section des résultats, nous présentons un exemple de l'application d'un MHMM pour illustrer les étapes à suivre et donner une idée de la manière dont les résultats de ces modèles peuvent être présentés et visualisés. Nous terminons par une conclusion dans laquelle nous réfléchissons brièvement à l'utilisation des MHMM dans la recherche en sciences sociales.

## 2. MÉTHODES

Les MHMM font partie de la famille des modèles à variables latentes. C'est pourquoi nous commencerons par une brève explication sur l'analyse des classes latentes (ACL), après quoi, nous examinerons comment les MHMM la développent.

L'ACL est une méthode analytique qui permet d'étudier des variables latentes discrètes sur la base de données transversales (Masyn, 2013; Nylund-Gibson et al., 2019; Nylund-Gibson & Choi, 2018). Les individus qui réalisent des scores similaires au niveau des indicateurs relatifs aux variables latentes sont regroupés dans un état. En ce sens, l'ACL est une méthode axée sur l'individu, par opposition à l'analyse factorielle qui est, elle, une méthode axée sur la variable (Collins & Lanza, 2010). Un choix important à faire dans la méthode ACL est celui du nombre optimal d'états. Ce choix se fait en deux étapes : (1) des modèles avec un nombre croissant d'états sont estimés ; et (2) un chercheur choisit le modèle le plus optimal. Ce choix repose à la fois sur des paramètres statistiques (critère d'information bayésien, critère d'information d'Aikake) et sur l'interprétabilité des états. En ce qui concerne la deuxième étape, l'hétérogénéité des classes latentes et la discriminabilité (également appelée séparation ou entropie) sont importantes. Cela implique de donner la préférence à un modèle avec des états homogènes et dans lequel les états sont suffisamment distincts les uns des autres. Les états du modèle choisi représentent en fait des catégories de la variable latente. Chacune des catégories doit être étiquetée en fonction des scores réalisés par les individus de cette catégorie au niveau des indicateurs de la variable latente. Ensuite, la probabilité que cet individu se trouve dans cet état est estimée pour chaque individu de chaque état. L'individu est attribué à l'état pour lequel la probabilité qu'il s'y trouve est la plus élevée. Le modèle ACL comprend deux parties : (1) la partie *mesure* ; et la partie (2) *structurelle*. La première partie fait référence à la manière dont les indicateurs sont liés à la variable latente, et la seconde partie à la distribution des individus dans les catégories de la variable latente.

Les Hidden Markov Models (HMM) sont une extension longitudinale de l'ACL (Nylund-Gibson et al., 2023; van der Nest et al., 2020; Vermunt, 2010b). Cela signifie que les états sont identifiés sur la base de données longitudinale et que le modèle modélise les mouvements (transitions) entre les états. Comme déjà mentionné, cela se fait sur la base d'une structure de transition de Markov : l'état dans lequel se trouve un individu au moment  $t$  est prédit exclusivement par l'état dans lequel il se trouvait au moment  $t-1$ . L'hypothèse est que les valeurs des indicateurs observés à différents moments sont indépendantes et conditionnées par les états latents (*cachés*). Cela signifie que la dépendance entre les moments doit s'expliquer par la structure d'autocorrélation des classes latentes. Tout comme l'ACL, les HMM comprennent une partie *mesure* et une partie *structurelle*. Au niveau des HMM aussi, la partie *mesure* fait référence à la relation entre les indicateurs et la variable latente sous la forme de *probabilités de réponse aux items* : la probabilité d'un score au niveau d'un indicateur lié à l'état auquel l'individu est rattaché. Dans les HMM, la partie *structurelle* fait référence à l'état initial, c'est-à-dire, l'état dans lequel un individu est classé à  $t=0$  et les probabilités de transition (*transition probabilities*) : c'est-à-dire, la probabilité qu'un individu passe d'un état à un autre état entre le moment  $t-1$  et  $t$ . L'intérêt du HMM (et de l'ACL) il est possible d'ajouter des prédicteurs, à la fois à la partie *mesure* et à la partie *structurelle*. Il est donc possible, par exemple, d'étudier si l'âge affecte l'état initial

et/ou les probabilités de transition d'un individu. Dans le modèle standard, on prend comme point de départ que ces probabilités de transition sont les mêmes à chaque moment. Si on imagine 40 moments, l'hypothèse est que les probabilités de transition du moment 1 au moment 2 sont identiques aux probabilités de transition du moment 39 au moment 40. Ce n'est pas probable dans tous les cas. Il est donc possible d'ajouter un élément temporel au modèle pour que des probabilités de transition puissent être calculées pour chaque moment.

Les MHMM constituent une nouvelle extension. En plus de la structure des HMM, on ajoute aussi à ces modèles des *mixtures*, ou trajets (Vermunt, 2010b). Ces trajets peuvent également être considérés comme des catégories d'une variable latente constante dans le temps. Dans ce cas, la variable latente est constituée des schémas de développement des individus au fil du temps, ou en d'autres termes, des transitions entre les états de la variable latente précédente. Les individus sont regroupés dans des trajets, sur la base de leur état initial et des probabilités de transition. Ici aussi, le chercheur joue un rôle important dans le choix du nombre optimal de trajets. Plusieurs modèles sont d'abord estimés avec un nombre croissant de trajets, et ensuite, sur la base des critères statistiques mentionnés plus haut et de l'interprétabilité des trajets, le chercheur détermine le nombre optimal de trajets. Plus le nombre de trajets est élevé, plus l'estimation du modèle prend du temps. Pour réduire le temps nécessaire à l'estimation, on peut appliquer la procédure à deux étapes de Bakk et Kuha (2018). Dans cette procédure, les paramètres de la partie *mesure* estimés (à partir d'un modèle ne comprenant qu'un seul *mixture*) sont fixés au moment de la modélisation des trajets. Le résultat de l'estimation est donc, ici aussi, une variable (catégorielle) discrète. Les catégories sont une nouvelle fois étiquetées à l'aide des scores réalisés au niveau des indicateurs des trajets latents, à savoir l'état initial et les probabilités de transition. Des prédicteurs peuvent également être ajoutés aux trajets. C'est au chercheur qu'il appartient de décider à quel endroit du modèle (état initial, probabilités de transitions et/ou trajets) les prédicteurs sont ajoutés. Dans le cadre du choix de l'endroit optimal où placer les prédicteurs, des considérations tant statistiques (*model fit*) que théoriques peuvent jouer un rôle.

### 3. EXEMPLE D'UTILISATION DES MHMM

Dans cette section, nous allons expliquer l'utilisation des MHMM dans l'analyse des données longitudinales à l'aide d'une étude sur les carrières des jeunes travailleurs (Eberlein et al., 2024) Des informations sur les données et la population de cette étude sont disponibles dans l'encadré ci-contre.

#### **DONNÉES ET POPULATION DE L'ÉTUDE**

Cette étude a utilisé les données de registre du Centraal Bureau voor de Statistiek (CBS). Ces données contiennent des données mensuelles, entre autres, sur le type de contrat, les heures de travail et les revenus. Entre 2009 et 2013, cette étude a suivi les individus qui ont terminé une formation (avec ou sans diplôme) pendant 72 mois. Seuls les individus qui avaient terminé une formation (avec ou sans diplôme) en vue d'obtenir une qualification pour le marché du travail ont été inclus dans l'étude (niveaux universitaire, professionnel

secondaire et supérieur et CITE 353, 354, 645, 655, 747 et 757) Seuls les sortants du système éducatif qui avaient trouvé un emploi au cours d'un de ces mois ont été inclus et la période de suivi a commencé le premier mois au cours duquel le sortant du système éducatif avait un emploi. Au total 672.757 individus répondaient aux critères d'inclusion. En raison de la puissance de calcul nécessaire pour les MHMM, un échantillon aléatoire de 12.000 individus a été extrait de ce groupe.

L'objectif de cette étude était de comprendre comment la *qualité de l'emploi* évoluait les premières années de carrière des jeunes Néerlandais sortis du système éducatif : les individus qui ont terminé leur formation (avec ou sans diplôme) et sont entrés sur le marché du travail. Dans cette recherche, la variable latente est donc la *qualité de l'emploi*, un concept multidimensionnel qui comporte deux dimensions : (1) la sécurité d'emploi et (2) le niveau de revenu. Une première question à laquelle il faut répondre est la suivante : quels indicateurs allons-nous utiliser pour mesurer cette variable latente ? Dans cette étude, nous avons choisi de mesurer cette variable à l'aide des indicateurs suivants : (1) le statut de travail (salarié, indépendant, inactif) ; (2) le type de contrat (fixe, temporaire, intermittent /intérimaire, pas de contrat) ; (3) heures de travail ; et (4) revenu mensuel. Il est possible que les indicateurs relatifs au statut professionnel et au type de contrat se chevauchent. Il a cependant été choisi de les inclure en tant qu'indicateurs distincts dans le modèle, car il est possible, par exemple, qu'un individu soit à la fois indépendant et salarié.

La première étape d'un MHMM consiste à déterminer le nombre optimal d'états (ou de catégories) de la variable latente, dans le cas qui nous occupe, donc la *qualité de l'emploi*. Dans la recherche décrite ici, il a été choisi de prendre le modèle à huit états, ce qui signifie que notre variable latente a huit catégories avec différents niveaux ou formes de la *qualité de l'emploi*. Ces catégories sont ensuite étiquetées en fonction des scores réalisés par les individus dans les états au niveau des indicateurs cités plus hauts (voir Tableau 1). La catégorie « faibles revenus, contrat temporaire » doit son nom au fait que 99 % des individus dans cet état ont des contrats temporaires et de faibles revenus. Nous avons distingué les huit états suivants (nous allons d'abord les citer en français et ensuite donner leur désignation initiale en anglais) : (1) faibles revenus, contrat temporaire (*low income, fixed-term*) ; (2) revenus élevés, contrat temporaire (*fortunate fixed term*) ; (3) intérimaires et intermittents (*TAW/on-call*) ; (4) indépendants (*self-employed*) ; (5) faibles revenus, contrat fixe (*low income permanency*) ; (6) revenus moyens, contrat fixe (*moderate permanency*) ; (7) revenus confortables, contrat fixe (*comfortable permanency*) ; et (8) ne travaillent pas (*not in employment*).

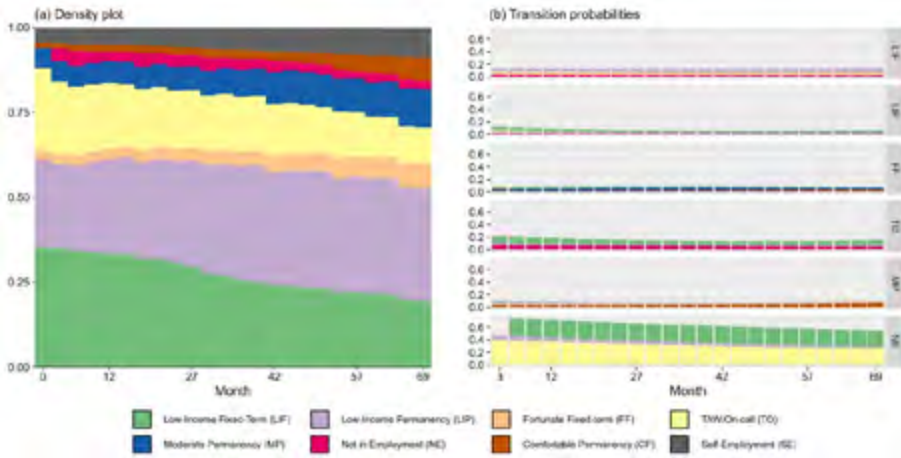
**TABEAU 1 : ÉTATS LATENTS DE LA QUALITÉ DE L'EMPLOI (POURCENTAGES)**

Indicateurs	Huit états de la variable latente <i>qualité de l'emploi</i>							
	Faibles revenus, contrat temporaire	Revenus élevés, contrat temporaire	Intérimaires et intermittents	Indépendants	Faibles revenus, contrat fixe	Revenus moyens, contrat fixe	Revenus confortables, contrat fixe	Ne travaillent pas
Taille de l'état	21	13	12	6	18	12	9	10
Type de contrat								
Fixe	0	0	1	0	100	100	100	0
Temporaire	99	100	2	0	0	0	0	0
Travail intérimaire	0	0	53	0	0	0	0	0
Travail intermittent	0	0	44	0	0	0	0	0
Pas de contrat	0	0	0	100	0	0	0	100
Revenus mensuels								
Pas de revenus	0	0	0	5	0	0	0	48
1-750	16	0	19	20	13	0	0	12
751-1500	32	0	32	20	26	1	0	31
1501-2250	51	3	32	14	61	5	0	7
2251-3000	1	70	13	11	0	95	0	2
3001-3750	0	19	3	10	0	0	66	0
> 3750	0	8	1	19	0	0	34	0
Statut sur le marché du travail								
Salarié	100	100	100	0	100	100	100	0
Indépendants	0	0	0	99	0	0	0	0
Ne travaillent pas	0	0	0	1	0	0	0	100
Heures de travail par semaine	30	39	23	-	30	38	40	0

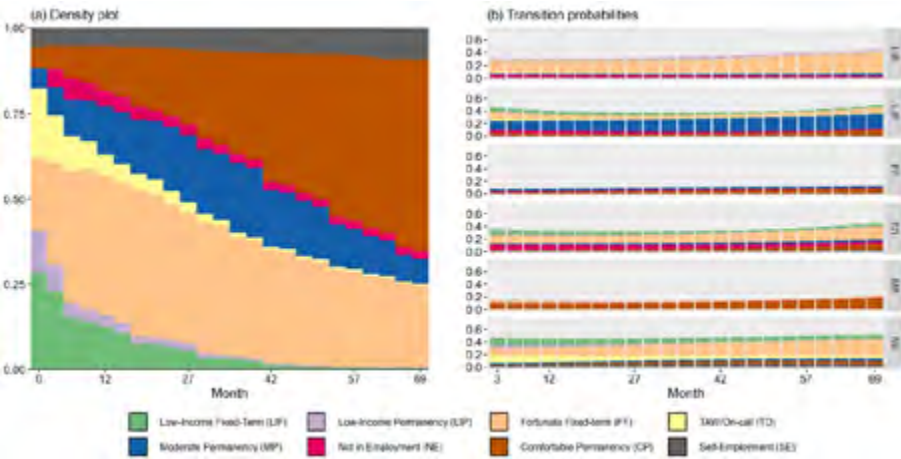
Outre la partie *mesure*, qui donne un aperçu de la relation entre les indicateurs et la variable latente, cette étape donne également un aperçu de la partie *structurelle* : la catégorie *qualité de l'emploi* dans laquelle les individus entrent sur le marché du travail (état initial) et la manière dont les individus évoluent entre les huit catégories. Comme déjà mentionné plus haut, on estime d'abord une matrice de transition qui reprend toutes les possibilités de transition sur l'ensemble de la période de l'étude. En ajoutant une variable temporelle à la partie structurelle du modèle, il devient possible d'examiner si ces probabilités de transition évoluent sur la période de l'étude.

Une fois la première étape réalisée, on peut passer à la deuxième étape. Dans cette étape, on choisit le nombre optimal de trajets ou *mixtures*. Ce choix est effectué sur la base de l'estimation de différents modèles contenant un nombre croissant de mixtures. Comme déjà mentionné, les mixtures regroupent des individus qui ont des états initiaux et des probabilités de transition comparables. Cette étude a distingué quatre trajets de ce type. Encore une fois, les trajets sont étiquetés sur la base des scores réalisés par les individus au niveau des états initiaux et surtout des probabilités de transition des trajets. Dans cette étude, le premier trajet est intitulé « piège à la stabilité » (*stable entrapment*) nommé ainsi parce que les individus se retrouvent coincés pendant de longues périodes dans une situation de « faibles revenus, contrat temporaire » et « faibles revenus, contrat fixe ». La Figure 1 montre le *density plot* (tracé de densité) et les probabilités de transition du trajet 1. Le *density plot* montre, pour chaque moment, comment les individus ont été répartis entre les huit états identifiés dans l'étude. On peut y voir que le nombre d'individus dans les catégories « faibles revenus, contrat temporaire » et « intérimaires et intermittents » diminue et que le nombre d'individus dans les catégories « faibles revenus, contrat fixe » et « revenus moyens, contrat fixe » augmente. Le *density plot* ne permet toutefois pas de voir comment les individus changent d'états. Les probabilités de transition (voir également la Figure 1) fournissent plus d'informations à ce sujet. Les abréviations utilisées pour les états, à l'extrême droit de la figure indiquent l'état à  $t-1$  et les barres indiquent le niveau de probabilité de passer d'un état aux autres états. On peut voir, par exemple, que les individus de la catégorie « faibles revenus, contrat temporaire » (LIF) passent principalement dans la catégorie « faibles revenus, contrat fixe » (LIP), « ne travaillent pas » (NE) et « revenus élevés, contrat temporaire » (FF). Les individus qui appartiennent à la catégorie « intérimaires et intermittents » (TO) passent principalement de la catégorie « faibles revenus, contrat temporaire » au début de la période de l'étude à la catégorie « ne travaillent pas » pendant toute la période de l'étude. Le deuxième trajet a été qualifié de « mobilité ascendante » (*upward mobility*) : de nombreux individus commencent dans un état qui se caractérise par de nombreux contrats temporaires, mais passent ensuite, pendant la durée de l'étude, à la catégorie « revenus confortables, contrat fixe ». Le troisième trajet est appelé « transitions vers un contrat fixe » (*moving to permanency*). Ici aussi, de nombreux individus commencent dans un état qui se caractérise par de nombreux contrats temporaires. Un grand nombre de ces individus passent ensuite dans la catégorie « revenus moyens, contrat fixe » : un état dans lequel on retrouve de nombreux contrats fixes, mais avec un revenu mensuel inférieur à celui de la catégorie « revenus confortables, contrat fixe ». Enfin, le quatrième et dernier trajet est celui appelé « sortie de l'emploi » (*moving out of employment*). Comme son nom le laisse supposer, dans ce trajet, de nombreux individus font une transition vers l'état « ne travaillent pas ».

**FIGURE 1 :** REPRÉSENTATION VISUELLE DU TRAJET 1 : « PIÈGE À LA STABILITÉ »

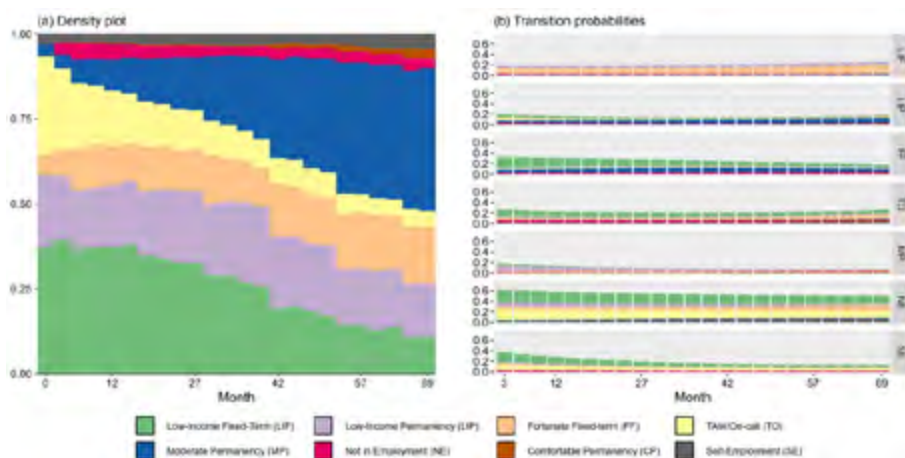


**FIGURE 2 :** REPRÉSENTATION VISUELLE DU TRAJET 2 : « MOBILITÉ ASCENDANTE »

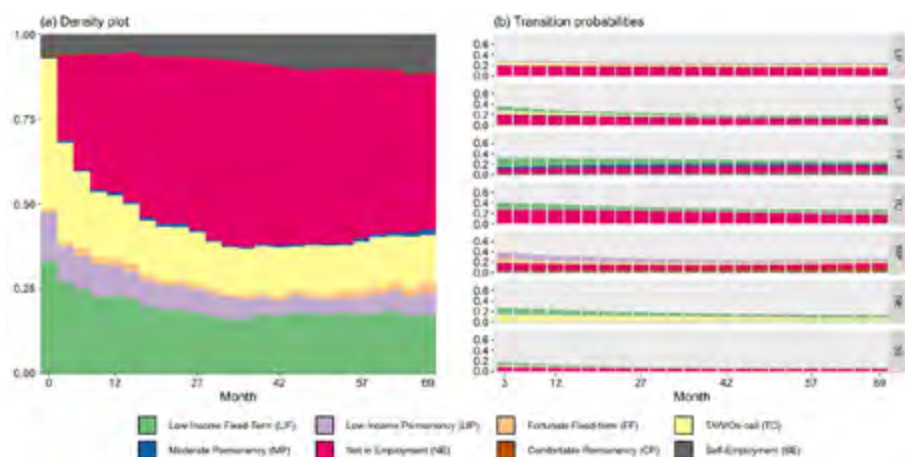




**FIGURE 3:** REPRÉSENTATION VISUELLE DU TRAJET 3 : « TRANSITIONS VERS UN CONTRAT FIXE »



**FIGURE 4:** REPRÉSENTATION VISUELLE DU TRAJET 4 : « SORTIE DE L'EMPLOI »



Une question importante est de déterminer comment visualiser au mieux les résultats d'un MHMM. Il est important, de faire remarquer, une fois encore, qu'il s'agit d'un modèle probabiliste : le modèle calcule la probabilité qu'un individu se trouve dans une des catégories de variables latentes et la probabilité qu'un individu se trouve dans un des trajets. Les visualisations des trajets réalisées à l'aide des *density plots* ci-dessous sont néanmoins, par nature, déterministes : nous y utilisons l'état prédit pour un individu à chaque moment. Cette visualisation déterministe des résultats s'avère légèrement problématique quand l'entropie (capacité de discernement) des états et des trajets est élevée. Dans ce modèle, l'entropie est de 0,70. C'est une valeur relativement

faible, mais qui reste acceptable. Il est important de mentionner que ce n'est pas le cas de la visualisation des probabilités de transition : ces dernières sont le résultat direct du modèle et ne reposent donc pas sur le statut prédit.

Une dernière possibilité des MHMM que nous allons aborder ici est celle d'ajouter des prédicteurs. Comme déjà mentionné plus avant, des prédicteurs peuvent être ajoutés dans différentes parties d'un MHMM : la partie *mesure*, la partie *structurelle* et/ou au niveau des trajets. Dans cette étude, sur la base de considérations théoriques et statistiques, nous avons choisi d'ajouter des prédicteurs au niveau des trajets (le niveau d'instruction et le fait qu'un sortant du système éducatif a ou non obtenu un diplôme au moment où il a quitté l'enseignement) et des états initiaux (temps écoulé avant de trouver un premier emploi et âge d'entrée sur le marché du travail). Les Tableaux 2 et 3 présentent les coefficients des prédicteurs, estimés à l'aide d'une analyse de régression multinomiale. Un coefficient positif indique un lien positif, un coefficient négatif indique un lien négatif).<sup>1</sup> Le Tableau 2 reprend les coefficients des prédicteurs de l'état initial. Le coefficient positif de la catégorie « revenus moyens, contrat fixe » (0,2329) dans le Tableau 2 montre que plus il lui faut de temps pour trouver un emploi, plus un individu est susceptible de commencer sa carrière dans cet état. On constate également que plus un individu est âgé lorsqu'il trouve son premier emploi, plus la probabilité est grande qu'il commence dans la catégorie « revenus confortables, contrat fixe » (0,3664).

**TABEAU 2 : COEFFICIENTS LOGIT AU NIVEAU DU CODAGE DES EFFETS, PRÉDICTEURS DE L'ÉTAT INITIAL**

		Temps nécessaire pour trouver un emploi	Âge au moment où l'individu trouve un emploi
État initial	Faibles revenus, contrat temporaire	0,0361	-0,1847
	Revenus élevés, contrat temporaire	-0,0216	-0,1143
	Intérimaires et intermittents	-0,0008	0,0704
	Indépendants	0,0689	-0,1762
	Faibles revenus, contrat fixe	-0,3147	0,1936
	Revenus moyens, contrat fixe	0,2329	-0,1259
	Revenus confortables, contrat fixe	-0,0899	0,3664
	Ne travaillent pas	0,0891	-0,0292

(1) Voici les coefficients des effets. Ils sont notés sur une échelle jusqu'à 0. Une autre possibilité est de choisir une catégorie de référence. Les coefficients permettent alors d'évaluer la probabilité de commencer dans un état initial donné par rapport à l'état initial de référence. On peut aussi procéder de la même manière pour les trajets.

Le Tableau 3 reprend les coefficients des prédicteurs des trajets. Dans ce tableau, on peut voir, entre autres, que les étudiants qui quittent l'université ont une plus grande probabilité de « mobilité ascendante » (1,5604) que les étudiants dont la scolarité s'est arrêtée à un autre niveau. Les étudiants qui ont obtenu un diplôme, ont aussi une probabilité plus élevée de « mobilité ascendante » (1,3799) que les étudiants qui ont arrêté leur scolarité sans avoir de diplôme.

**TABEAU 3 : COEFFICIENTS LOGIT AU NIVEAU DU CODAGE DES EFFETS, PRÉDICTEURS DES TRAJETS**

		Niveau d'instruction			Diplôme obtenu
		Université	Enseignement professionnel de niveau supérieur	Enseignement professionnel de niveau secondaire	
Trajet	Piège à la stabilité	-1,141	-0,0588	1,1998	-0,5133
	Mobilité ascendante	1,5604	0,1737	-1,7341	1,3799
	Transitions vers un contrat fixe	-0,0498	0,2986	-0,2488	0,6131
	Sorties de l'emploi	-0,3696	-0,4136	0,7831	-1,4797

#### 4. CONCLUSION

L'objectif de cet article était de familiariser le lecteur avec les MHMM. Pour cela, nous avons discuté de la distinction entre le MHMM et les autres méthodes d'analyse basées sur les processus tels que l'AS, nous avons donné les étapes à suivre dans le cadre de l'estimation du MHMM et mentionné les choix que cette méthode demande aux chercheurs de faire ainsi qu'illustré l'utilisation du MHMM à l'aide d'un exemple.

Le choix entre l'AS et le MHMM est principalement déterminé par notre perception du phénomène social que nous étudions ; devons-nous l'étudier comme probabiliste ou suffit-il de le considérer comme déterministe ? Étant donné que l'AS est une méthode moins complexe et moins chronophage, elle semble plus adaptée à l'étude des phénomènes unidimensionnels. Cependant, en ce qui concerne les phénomènes multidimensionnels, les résultats de l'AS multicanaux deviennent très complexes et donc difficiles à évaluer et à interpréter. Le MHMM, contrairement à l'AS, modélise la structure latente du phénomène et identifie donc les caractéristiques les plus pertinentes de l'état latent des trajets. La comparaison entre les résultats de l'AS multicanaux et du MHMM le montre immédiatement. Le MHMM est capable de plus facilement distinguer les trajets liés à des données très hétérogènes. Dans leurs articles Mattijssen et al. (2023) et Eberlein et al. (2024) ont utilisé des données de registre comparables sur une population comparable. Avec l'AS, Mattijssen et al. arrivent à 14 trajets de statut de revenus et de statut de contrat tandis qu'avec le MHMM, Eberlein et al. ne

trouvent que 4 trajets. Cela rend les résultats de l'étude plus facile à interpréter et à communiquer.

Enfin, le MHMM convient mieux à la recherche sur les relations causales. Étant donné que le MHMM est un modèle, il est possible d'étudier les relations causales en ajoutant des prédicteurs aux différentes parties du MHMM.

---

## LITTÉRATURE

- Abbott, A. (1983). Sequences of Social Events: Concepts and Methods for the Analysis of Order in Social Processes. *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 16(4), 129–147. <https://doi.org/10.1080/01615440.1983.10594107>
- Bakk, Z., & Kuha, J. (2018). Two-Step Estimation of Models Between Latent Classes and External Variables. *Psychometrika*, 83(4), 871–892. <https://doi.org/10.1007/s11336-017-9592-7>
- Collins, L. M., & Lanza, S. T. (2010). *Latent Class and Latent Transition Analysis: With Applications in the Social, Behavioral, and Health Science*. John Wiley & Sons, Inc.
- Eberlein, L., Pavlopoulos, D., & Garnier-Villarreal, M. (2024). Starting flexible, always flexible? The relation of early temporary employment and young workers employment trajectories in the Netherlands. *Research in Social Stratification and Mobility*, 89, 100861. <https://doi.org/10.1016/j.rssm.2023.100861>
- Fasang, A. E., & Liao, T. F. (2014). Visualizing Sequences in the Social Sciences. *Sociological Methods & Research*, 43(4), 643–676. <https://doi.org/10.1177/0049124113506563>
- Liao, T. F., Bolano, D., Brzinsky-Fay, C., Cornwell, B., Fasang, A. E., Helske, S., Piccarreta, R., Raab, M., Ritschard, G., Struffolino, E., & Studer, M. (2022). Sequence analysis: Its past, present, and future. *Social Science Research*, 102772. <https://doi.org/10.1016/j.ssresearch.2022.102772>
- Masyn, K. E. (2013). Latent Class Analysis and Finite Mixture Modeling. In P. E. Nathan & T. D. Little (Eds.), *The Oxford Handbook of Quantitative Methods: Vol. Volume 2: Statistical Analysis* (p. 63). Oxford University Press.
- Mattijssen, L., Pavlopoulos, D., & Smits, W. (2023). Does it pay off to specialize? The interplay between educational specificity, level and cyclical sensitivity. *Social Science Research*, 109, 102782. <https://doi.org/10.1016/j.ssresearch.2022.102782>
- Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, 4(4), 440–461. <https://doi.org/10.1037/tps0000176>
- Nylund-Gibson, K., Garber, A. C., Carter, D. B., Chan, M., Arch, D. A. N., Simon, O., Whaling, K., Tarrt, E., & Lawrie, S. I. (2023). Ten frequently asked questions about latent transition analysis. *Psychological Methods*, 28(2), 284–300. <https://doi.org/10.1037/met0000486>
- Nylund-Gibson, K., Grimm, R. P., & Masyn, K. E. (2019). Prediction from Latent Classes: A Demonstration of Different Approaches to Include Distal Outcomes in Mixture Models. *Structural Equation Modeling: A Multidisciplinary Journal*, 26(6), 967–985. <https://doi.org/10.1080/10705511.2019.1590146>

Studer, M., & Ritschard, G. (2016). What matters in differences between life trajectories: A comparative review of sequence dissimilarity measures. *Journal of the Royal Statistical Society, Series A: Statistics in Society*, 179(2), 481–511. <https://doi.org/10.1111/rssa.12125>

van der Nest, G., Lima Passos, V., Candel, M. J. J. M., & van Breukelen, G. J. P. (2020). An overview of mixture modelling for latent evolutions in longitudinal data: Modelling approaches, fit statistics and software. *Advances in Life Course Research*, 43, 100323. <https://doi.org/10.1016/j.alcr.2019.100323>

Vermunt, J. K. (2010a). Longitudinal Research Using Mixture Models. In K. Montfort, J. H. Oud, & A. Satorra (Eds.), *Longitudinal Research with Latent Variables* (pp. 119–152). Springer. [https://doi.org/10.1007/978-3-642-11760-2\\_4](https://doi.org/10.1007/978-3-642-11760-2_4)

Vermunt, J. K. (2010b). Longitudinal Research Using Mixture Models. In K. Montfort, J. H. Oud, & A. Satorra (Eds.), *Longitudinal Research with Latent Variables* (pp. 119–152). Springer. [https://doi.org/10.1007/978-3-642-11760-2\\_4](https://doi.org/10.1007/978-3-642-11760-2_4)

## TABLE DES MATIERES

**LES MIXTURE HIDDEN MARKOV MODELS (MODÈLES DE MARKOV CACHÉS À EFFETS MIXTES) DANS LA RECHERCHE EN SCIENCES SOCIALES**

<b>1. INTRODUCTION</b>	429
<b>2. MÉTHODES</b>	431
<b>3. EXEMPLE D'UTILISATION DES MHMM</b>	432
<b>4. CONCLUSION</b>	439
<b>LITTÉRATURE</b>	441

