

# Etude des déterminants de la sortie du chômage vers le travail

Lieven Desmet  
Direction Statistiques et Etudes  
ONEM

16 septembre 2011

## **Résumé**

Le but de notre recherche est de savoir si (et si oui, dans quelle mesure) certains facteurs peuvent influencer la sortie du chômage vers le travail. Nous prenons comme population cible les personnes entrées au chômage pour la première fois en 2007 (n'ayant pas d'antécédent au chômage). La base de données de recherche "Panel ONEM" nous permet d'en analyser un échantillon. Pour ces personnes, nous examinons si elles sont au travail un an après leur entrée au chômage et, par un modèle de régression logistique, nous cherchons le lien entre ce fait et les caractéristiques de la personne, son statut de chômeur et son entourage familial. L'analyse confirme quelques faits connus : les différences entre les régions, l'impact important du niveau d'études mais aussi de l'âge et de la catégorie d'indemnisation. Les différences entre les sexes peuvent s'expliquer par rapport à la situation familiale de la personne.

## Remerciements

L'auteur tient à remercier Kristel Bogaerts et Karel Van den Bosch de l'Université d'Anvers (Centrum voor Sociaal Beleid Herman Deleeck) et Joost Bollens de l'institut HIVA de la KULeuven pour les conseils avisés et les remarques.

La base de données "Panel ONEM" est le fruit d'une collaboration entre l'ONEM et la Politique scientifique fédérale. Son objet est de créer un échantillon de données qui répond aux besoins de connaissance de l'ONEM. Cette base de données est constituée avec l'appui de chercheurs de l'Université d'Anvers et de la KULeuven financés par la Politique scientifique dans le cadre de son programme de recherche AGORA. Les données elles-mêmes se basent essentiellement sur les informations collectées par la Banque Carrefour de la Sécurité Sociale. La bonne collaboration avec Chris Brijs de la Banque Carrefour a été également très appréciée.

Un grand merci également aux collègues de l'ONEM pour la relecture, les conseils et l'aide lors de la traduction.

# 1 But de la recherche et méthodologie

## 1.1 Introduction et question de recherche

Bien comprendre les facteurs qui influencent la sortie du chômage est d'une grande importance pour les gestionnaires politiques. Nous parlons alors de sortie du chômage, mais nous sommes évidemment surtout intéressés par la sortie du chômage vers le travail et non par les sorties vers une situation d'inactivité. La hausse du taux d'emploi est, par ailleurs, un des objectifs principaux de la stratégie Europe 2020 pour la croissance et l'emploi.

Afin d'examiner le pourcentage de sortie vers le travail, nous ne devons pas seulement disposer d'une vaste série de données relatives au chômage mais également de données relatives à l'occupation ainsi que de données personnelles et autres. Ceci est uniquement possible par le croisement des fichiers de paiement du chômage (ONEM) avec d'autres fichiers de la Sécurité sociale. Nous recourons, dès lors, à ce qu'on appelle le panel ONEM : une banque de données intégrée (basée sur un échantillon suffisamment grand) spécialement créée pour examiner les carrières des gens qui ont, à un moment ou un autre, été en contact avec l'ONEM (elle a été constituée dans le cadre d'une collaboration entre l'ONEM, la BCSS (Banque Carrefour de la Sécurité Sociale) et les chercheurs de l'Université Catholique de Louvain (KULeuven) et de l'Université d'Anvers, en un projet financé par la Politique Scientifique fédérale).

Le panel ONEM ne fournit pas uniquement des données sur le chômage, comme la catégorie d'indemnisation, la base d'admission, l'importance du montant journalier et la durée, mais également des données personnelles comme le domicile, le sexe, l'âge, la nationalité, ainsi que des données sur la situation familiale de l'intéressé comme son lien familial dans le ménage, le fait qu'il y ait des enfants dans le ménage (et l'âge du cadet) et la situation du partenaire éventuel. Des données importantes ont trait à la nomenclature de la position socio-économique qui indique si la personne est demandeuse d'emploi, si elle travaille ou si elle est inactive. Nous disposons en outre de données sur le niveau d'études. Toutes les données sont anonymisées.

Etant donné que les données sont disponibles sur une base trimestrielle (à l'exception des données personnelles), nous pouvons effectivement retracer les parcours individuels et détecter les transitions d'une position de chômeur à une position de travailleur. Nous pouvons donc ainsi constater une sortie du chômage vers le travail.

A présent, nous pouvons déjà formuler un peu plus précisément la question à approfondir : quelles caractéristiques (liées à la personne, au chômage, à la famille, etc.) influencent la sortie vers le travail ? Dans quel sens et dans quelle mesure ?

La réponse à cette question simple en apparence n'est ni brève, ni facile : il est clair que plusieurs facteurs interviennent en même temps. Notre approche consiste ensuite à construire un modèle, par le biais d'une analyse statistique (à savoir la régression logistique), de probabilité de sortie vers le travail, en fonction des différentes variables qui correspondent aux caractéristiques observées. Les résultats d'un tel exercice montrent quels effets sont importants, quels sont ceux qui le sont moins et dans quel sens l'effet agit. Le grand avantage de cette approche réside dans le fait que ces effets peuvent être examinés simultanément.

Pour rendre le texte accessible au grand public, les détails techniques se retrouvent dans l'Appendice. Tout d'abord, un certain nombre de définitions et de choix doivent toutefois être déterminés.

## 1.2 Population cible et concept de taux de sortie

En ce qui concerne le concept de chômage, nous précisons d'abord que nous nous intéressons aux personnes qui sont connues tant à l'ONEM comme chômeurs indemnisés (pour être précis "CCI-DE" : chômeur complet indemnisé et demandeur d'emploi) que dans la nomenclature de la position socio-économique (statut de demandeur d'emploi). D'autre part, le *timing* (la chronologie) est de la plus grande importance : quand l'intéressé est-il connu comme chômeur ? Et à quel moment allons-nous observer par la suite si la personne est sortie du chômage pour travailler ou non ?

Afin d'obtenir un groupe plus homogène, nous prenons seulement en considération les personnes qui n'ont pas d'antécédent en tant que chômeur. On sait que les chômeurs de longue durée ont une probabilité de sortie plus faible, mais la durée du chômage d'un chômeur n'est pas simple à mesurer de façon précise. On sait également que la durée du chômage est elle-même influencée par d'autres facteurs qui ne sont pas toujours mesurés (en effet, au fil du temps, une certaine sélection s'opère, entraînant une sortie des personnes ayant une situation plus favorable et une plus forte représentation des personnes ayant une situation moins favorable). On remarque qu'il s'agit d'un choix méthodologique important : il est également possible d'examiner l'ensemble de la population de tous les chômeurs, quels que soient leurs antécédents (et la durée apparaîtra alors comme un facteur important).

L'échantillonnage du panel ONEM essaie de se rapprocher autant que possible de la dynamique de la population des chômeurs. Cela signifie concrètement que, par trimestre, un nouvel échantillon est prélevé parmi les nouveaux arrivants : ce sont des chômeurs (dans ce trimestre) qui, dans les 5 dernières années (précédant ce trimestre), n'étaient pas connus à l'ONEM. Nous nous concentrons dans ce texte sur ce groupe.

Nous collectons les données sur les personnes devenues chômeuses dans les 4 trimestres de l'année 2007 (pour annuler les effets saisonniers) et obtenons ainsi un

échantillon de 14578 personnes. Pour chaque chômeur, nous examinons 4 trimestres plus tard si l'intéressé adopte une position " actif " (travailleur), voir Figure 1. Dans ce cas, nous considérons une sortie vers le travail.

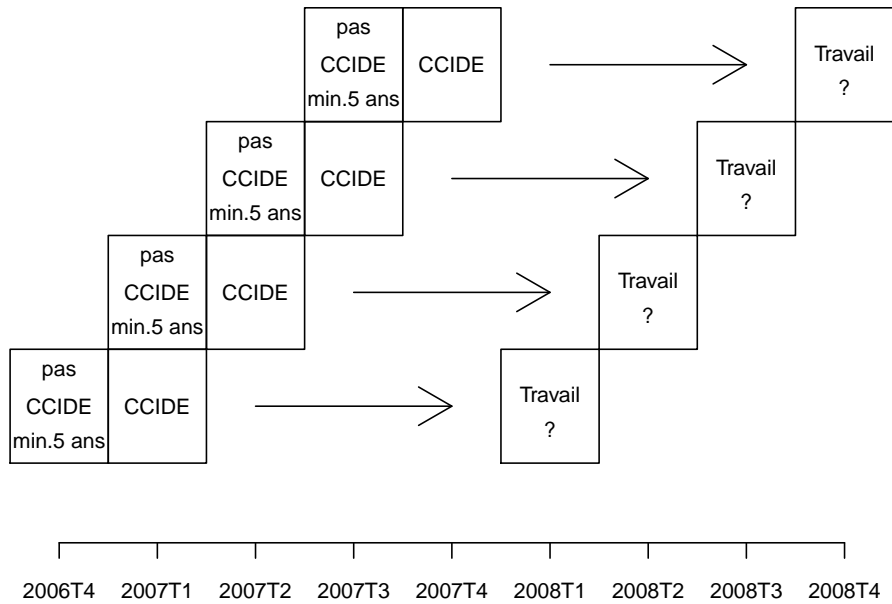


FIGURE 1 – Timing des entrées et sorties.

En bref, nous considérons la population de tous les chômeurs qui sont entrés dans le chômage dans le panel ONEM au courant de l'année 2007 et nous examinons s'ils travaillent un an plus tard. Penchons-nous d'abord sur un certain nombre de caractéristiques de cette population.

### 1.2.1 Profil de la population cible

Lors de la constitution de l'échantillon pour le panel ONEM, on a veillé à ce que les rapports entre les différents statuts ONEM et les régions soient respectés. Cependant, compte tenu des choix dans la Section 1.2 il y a une différence entre notre population cible (nouveaux entrants en 2007) et par exemple la population de tous les CCI-DE, connus en 2007 (cf. rapport annuel ONEM de 2007).

L'histogramme de l'âge dans la population cible (Figure 2, à gauche) montre une distribution asymétrique, où tous les âges sont représentés mais l'âge qui revient le

plus se situe dans la "petite" vingtaine. Les jeunes sortant des études en 2007 sont évidemment pris en compte dans notre échantillon (et nous constatons effectivement que les groupes de nouveaux venus sont un peu plus nombreux dans les 2e et 3e trimestres que dans les autres trimestres). Par contre, l'histogramme de l'âge pour la population du rapport annuel (Figure 2, à droite) montre une distribution plutôt bimodale : il y a un nombre important de chômeurs âgés. En ce qui concerne la durée le contraste est encore plus grand : dans la population cible la médiane de la durée est d'environ 2 mois, par contre, dans la population du rapport annuel, elle se trouve dans la classe de 24 à 29 mois (calculs sur données ONEM).

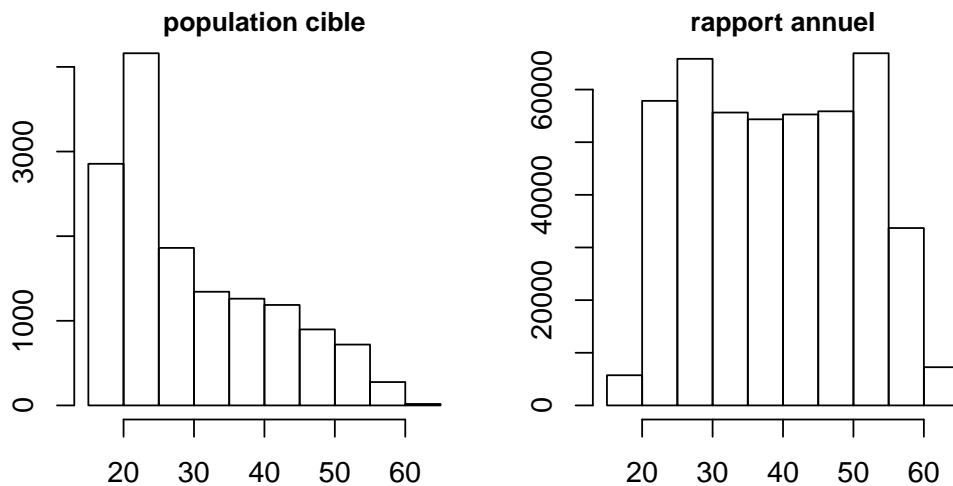


FIGURE 2 – Gauche : distribution de l'âge dans la population cible, Droite : idem pour la population des CCI-DE (rapport annuel 2007).

Des différences suivant le sexe et la région sont également constatées, voir Figure 3. Dans la population examinée, la région de Bruxelles-Capitale et la Région wallonne sont davantage représentées que dans la population de tous les CCI-DE. Cela peut s'expliquer par le fait que ces régions comptent un nombre relativement plus élevé de jeunes chômeurs. Par sexe, la population examinée est plus équilibrée que la population de tous les CCI-DE où les hommes sont davantage représentés.

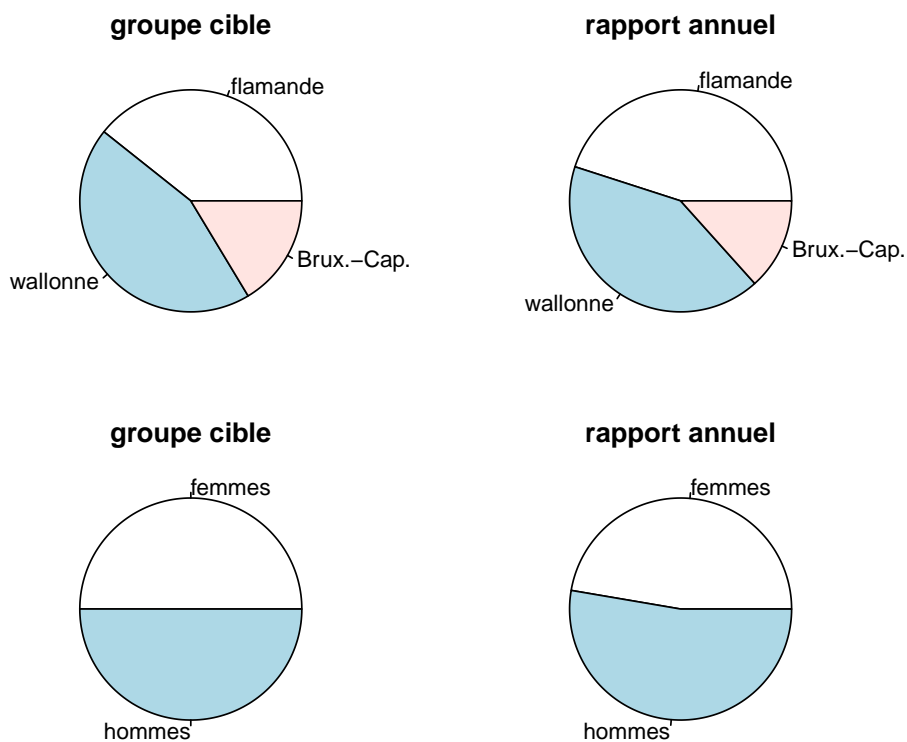


FIGURE 3 – Gauche : données rapport annuel 2007, droite : population cible. En haut : selon la région, en bas : selon le sexe.

### 1.2.2 Sortie du chômage vers le travail

Une année après leur entrée dans le chômage, 6170 des 14578 chômeurs travaillent, soit 42,32%, le taux de sortie global. Nous pourrions interpréter ce pourcentage comme la probabilité qu'une personne quelconque dans l'échantillon soit sortie du chômage, une année après son entrée, pour travailler (et ce pourcentage sera une bonne approximation de la probabilité de sortie dans toute la population cible, dont on ne connaît qu'un échantillon).

Par la suite, nous établirons un lien entre cette probabilité et les caractéristiques de la personne. Nous pouvons alors prévoir la probabilité de sortie, une année après son entrée, d'un chômeur quelconque sur la base de ces caractéristiques et donc examiner l'influence de ces caractéristiques.

Remarquons que le taux dépend fortement du délai entre le moment de l'entrée au chômage et le moment d'observation : si on regarde la situation de la personne un trimestre après son entrée au chômage, le taux sera nettement plus faible (environ 25 %) ; si on regarde après 8 trimestres, il sera plus élevé (environ 48 %).

Cela est illustré dans la Figure 4.

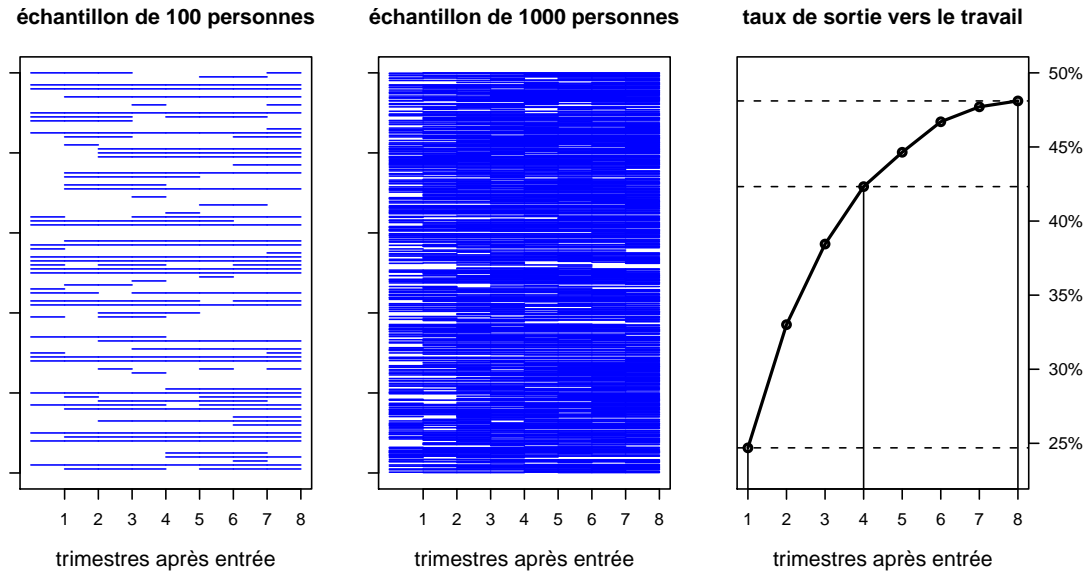


FIGURE 4 – Evolution du taux de sortie vers le travail au fil du temps de suivi. Gauche et milieu : les segments indiquent si la personne est au travail au fil des 8 trimestres à partir de l’entrée au chômage : si la personne n’est pas connue comme travaillant, le segment manque. Droite : le taux de sortie augmente au fil des trimestres.

D’autre part, dans notre définition, nous regardons uniquement la situation à un certain moment. Si, dans la même figure, nous considérons les trajets parcourus par un certain nombre de chômeurs choisis au hasard au cours des 8 trimestres suivant leur entrée, nous constatons qu’un certain nombre d’entre eux restent au travail de manière ininterrompue (ligne pleine) ou ne sortent pas vers le travail (aucune ligne tracée), mais également que bon nombre de chômeurs ont un parcours de carrière instable. Au fil du temps, nous voyons néanmoins que davantage de personnes sont au travail au même moment. Remarquons aussi qu’une photo instantanée, par exemple après 4 trimestres, ne donne pas une vision complète.

Les données avec lesquelles nous travaillons sont les données les plus récentes dans le panel ONEM : les personnes qui sont devenues chômeuses en 2007 et sont éventuellement sorties du chômage en 2008 ou 2009. L’objectif est évidemment d’actualiser l’étude, dès que des données plus récentes sont disponibles. En tout cas, le moment pour effectuer cette étude est intéressant dans la mesure où nous pouvons partir du principe que la crise n’a pas eu encore beaucoup d’influence ; c’est, en effet, une période favorable sur le plan de la conjoncture.



## 2 Analyse des facteurs de sortie vers le travail

Avant d'examiner les facteurs, nous devons disposer d'un modèle valable qui, d'une part, reflète bien les données et qui, d'autre part, n'est pas trop compliqué afin d'en faciliter l'interprétation. Il n'est pas judicieux de reprendre dans le modèle toutes les variables dont on dispose, déjà parce que certaines variables comprennent à peu près les mêmes informations ce qui perturbe les calculs.

Le choix du modèle demande une recherche en soi et dépend de plusieurs considérations comme d'une compréhension préalable des facteurs dont on connaît le caractère déterminant ou non, combinées aux statistiques qui reflètent dans quelle mesure le modèle est vraisemblable (compte tenu des données) et dans quelle mesure il explique les données.

Les effets interagissent parfois : il peut ainsi arriver que, par exemple, l'influence de la catégorie d'indemnisation soit différente chez les hommes et chez les femmes. Dans ce cas, il peut être opportun de prendre en compte un effet interactif.

Pour éviter trop de détails techniques, nous renvoyons à la Section 4.3 pour la liste des variables et la sélection de modèle et nous partons du principe que nous avons sélectionné un modèle. Nous regardons d'abord un modèle qui comprend autant que possible tous les effets, mais sans variables superflues.

### 2.1 Modèle de base

Le modèle qui est pris en compte intègre comme variables : niveau d'études, région, catégorie d'indemnisation, âge, hauteur du montant journalier, nationalité, sexe, le fait de savoir si le partenaire est actif, la base d'admission, la tranche d'âge du cadet (s'il y a des enfants) et enfin, l'effet d'interaction entre le sexe et la catégorie d'indemnisation.

Le Tableau 11 donne l'ordre de sélection des effets, en fonction d'une vraisemblance décroissante, ce qui en dit long sur l'importance relative des effets. Sans suivre cela au pied de la lettre, nous pouvons relever que le niveau d'études et la région constituent les principaux effets dans notre modèle, suivi de l'âge et de la catégorie d'indemnisation.

Sur la base des estimations des paramètres, il est possible d'interpréter les effets des différents facteurs : le Tableau 12 présente les estimations des paramètres ainsi que leur signification. Ce tableau nous montre, par exemple, l'influence hautement significative de la variable " région ", avec des coefficients négatifs pour les catégories Wallonie et Bruxelles (par rapport à la catégorie de référence Flandre). Sur la base des coefficients, nous pouvons toutefois aussi calculer la probabilité de sortie estimée (pour un exemple de calcul voir Section 4.3.3) pour un certain nombre de cas concrets, ce qui facilite l'interprétation des effets.

Pour examiner l'effet des déterminants les plus importants (le niveau d'études, la région, la catégorie d'indemnisation et la tranche d'âge) nous pouvons examiner un certain nombre de cas concrets. Dans ces exemples, on retrouve un certain nombre de combinaisons de valeurs que ces caractéristiques peuvent prendre et, ainsi, la probabilité de sortie est évaluée sur la base du modèle de régression logistique (voir Tableau 1). *Pour tous les exemples, nous donnerons les probabilités pour les trois régions. Les valeurs des autres variables qui ne figurent pas dans le tableau, par manque de place, seront abordées ultérieurement.*

niveau d'études	catégorie d'indemnisation	partenaire actif	âge (classe)	probabilité par région		
				flamande	wallonne	Brux.-Cap.
élevé	chef de ménage	non	≤ 25	66,9%	53,6%	50,0%
élevé	chef de ménage	non	> 50	35,3%	23,8%	21,3%
élevé	cohabitant	oui	≤ 25	81,0%	70,9%	67,8%
élevé	cohabitant	oui	> 50	53,5%	39,7%	36,3%
moyen	chef de ménage	non	≤ 25	52,7%	38,9%	35,5%
moyen	chef de ménage	non	> 50	23,1%	14,7%	13,0%
moyen	cohabitant	oui	≤ 25	70,1%	57,3%	53,7%
moyen	cohabitant	oui	> 50	38,8%	26,6%	23,9%
faible	chef de ménage	non	≤ 25	43,3%	30,4%	27,4%
faible	chef de ménage	non	> 50	17,1%	10,5%	9,3%
faible	cohabitant	oui	≤ 25	61,7%	47,9%	44,3%
faible	cohabitant	oui	> 50	30,3%	19,9%	17,7%

TABLE 1 – Effet des déterminants les plus importants

Certains exemples sont illustrés respectivement par les graphiques en Figure 5, 6 et 7.

Dans ces exemples, il s'agit - pour les autres caractéristiques - d'un homme belge, admissible sur la base du travail, bénéficiant d'une allocation moyenne et sans enfant (d'autres situations sont examinées plus loin dans le texte). Si la catégorie d'indemnisation ONEM est "cohabitant", on suppose également simultanément l'effet d'un partenaire actif.

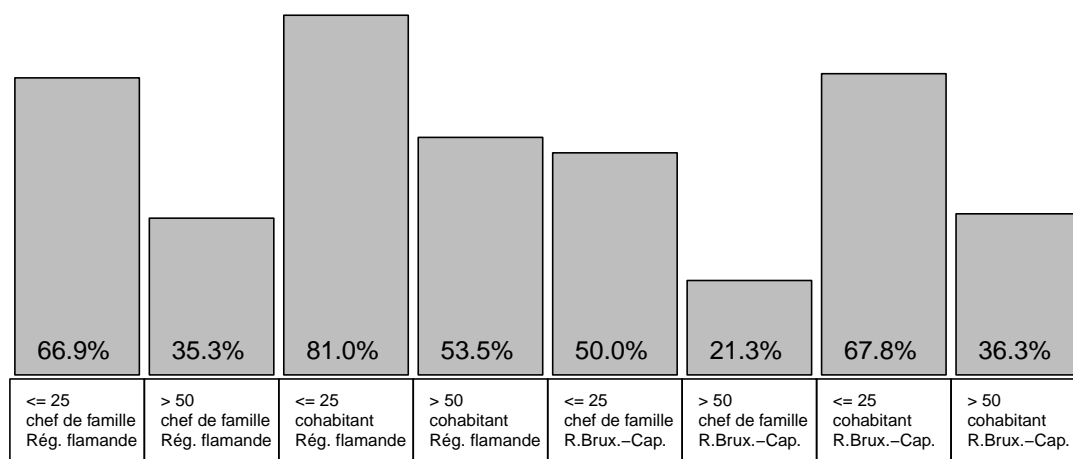


FIGURE 5 – Niveau d'études élevé. Estimations des probabilités de sortie pour différentes combinaisons de classe d'âge, région et catégorie d'indemnisation.

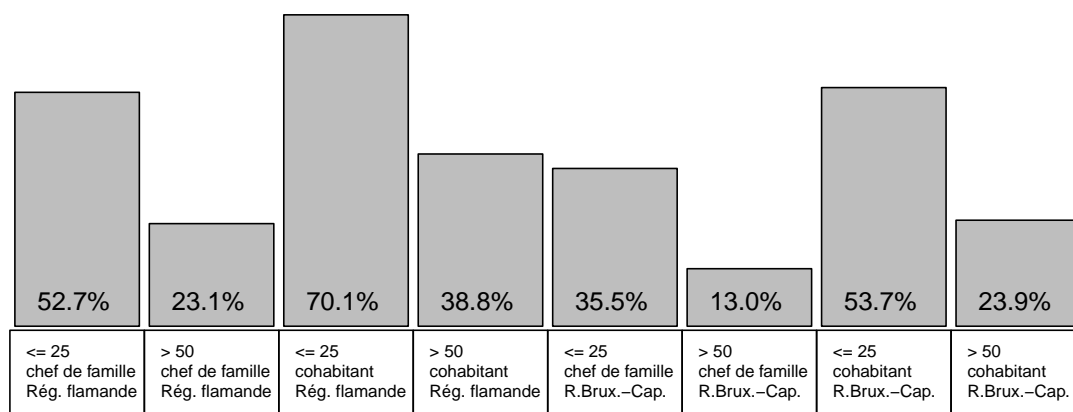


FIGURE 6 – Niveau d'études moyen. Estimations des probabilités de sortie pour différentes combinaisons de classe d'âge, région et catégorie d'indemnisation.

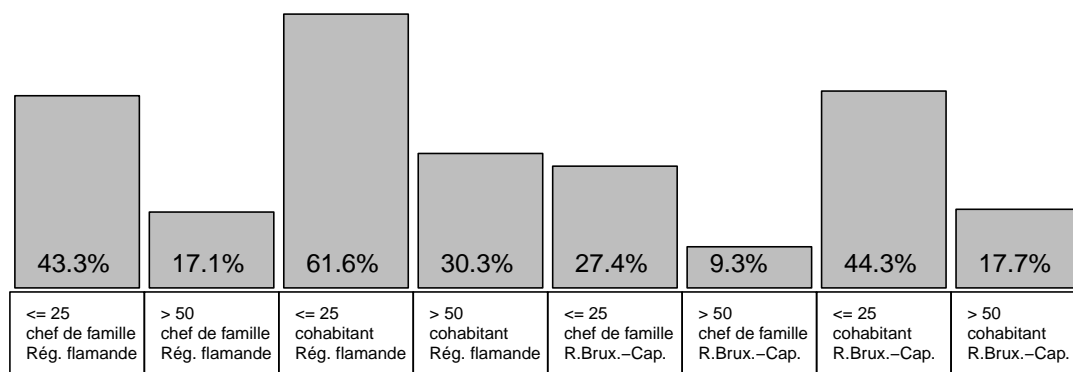


FIGURE 7 – Niveau d'études faible. Estimations des probabilités de sortie pour différentes combinaisons de classe d'âge, région et catégorie d'indemnisation.

Le niveau d'études a toute son importance : alors que la probabilité pour les personnes moins qualifiées se situe grosso modo entre 10 et 60%, elle se situe entre 20 et 80% pour les personnes hautement qualifiées. Si le code étude concerne un contrat d'apprentissage, la probabilité de sortie est un peu plus élevée que pour quelqu'un avec une formation de niveau secondaire.

De grandes différences sont ensuite notées entre les régions (voir Tableau 1) : en Région flamande, la probabilité de sortie est clairement plus élevée que dans la région de Bruxelles-Capitale. En Région wallonne, les probabilités sont un peu plus favorables que dans la Région de Bruxelles-Capitale, mais l'écart avec la Région flamande reste important. Si nous regardons au niveau de l'arrondissement, des différences au sein des régions apparaissent, mais nous disposons de trop peu de données à ce sujet.

Pour la Région de Bruxelles-Capitale, les éléments d'explication généralement avancés sont le décalage important qui existe entre le marché du travail dans la région (caractérisé par une demande de profils qualifiés et un grand nombre de navetteurs venant des autres régions) et le profil de ses habitants.

La catégorie d'indemnisation a un impact clair (voir Tableau 2) : la probabilité de sortir du chômage est la plus importante chez les cohabitants et la plus faible chez les chefs de ménage. Les isolés se positionnent à un niveau un peu plus haut que les chefs de ménage.

Malgré le fait que les cohabitants peuvent compter sur les revenus d'autres personnes au sein du ménage, on constate une probabilité de sortie plus élevée pro-

bablement liée au montant inférieur des allocations qu'ils perçoivent. Il est aussi possible que le fait de pouvoir partager les charges du ménage rende plus aisé l'accès au monde du travail. De surcroît, il faut prendre en compte le fait qu'un certain nombre de cohabitants sont des enfants (voir plus loin).

Il apparaît de plus que le fait, pour l'ayant droit, d'avoir un partenaire actif augmente les probabilités de sortie (voir plus loin dans le texte).

Dans le tableau il s'agit d'un ayant droit masculin. Du fait d'une interaction entre catégorie d'indemnisation et sexe, nous examinerons ces effets plus loin.

niveau d'études	catégorie d'indemnisation	partenaire actif	âge (classe)	probabilité par région		
				flamande	wallonne	Brux.-Cap.
faible	cohabitant	oui	$\leq 25$	61,7%	47,9%	44,3%
faible	isolé	non	$\leq 25$	43,4%	30,5%	27,5%
faible	chef de ménage	non	$\leq 25$	43,3%	30,4%	27,4%
faible	chef de ménage	non	$> 25, \leq 35$	40,1%	27,7%	24,9%
faible	chef de ménage	non	$> 35, \leq 50$	31,5%	20,8%	18,5%
faible	chef de ménage	non	$> 50$	17,1%	10,5%	9,3%

TABLE 2 – Effet de la catégorie d'indemnisation et de l'âge

L'âge joue également un rôle clair : au fur et à mesure que l'âge avance, les probabilités diminuent. Ce qui est frappant, c'est que la différence entre la tranche d'âge de moins de 25 ans et celle entre 25 et 35 est minime (il peut être financièrement intéressant d'engager un jeune, alors qu'un travailleur entre 25 et 35 ans a probablement plus d'expérience et est plus productif). Entre 35 et 50 ans, les probabilités sont déjà remarquablement plus faibles et, au-delà de 50 ans, même très faibles.

Dans un premier temps, on peut examiner l'influence du sexe de manière séparée. Dans ce cas, les femmes ont systématiquement un peu moins de chance de sortir du chômage. Etant donné que nous supposons que ceci peut être lié à la situation familiale, nous préférons examiner ce point en même temps que la catégorie d'indemnisation. Il ressort alors que l'impact de la catégorie d'indemnisation dépend aussi du sexe.

Le Tableau 3 montre clairement l'effet d'interaction entre le sexe et la catégorie d'indemnisation : chez les femmes, occuper la position " chef de ménage " signifie une probabilité de sortie beaucoup plus faible que chez les hommes (chez ces derniers, il n'y a presque pas de différence entre les positions " isolé " et " chef de ménage "). Par rapport à son homologue masculin, le taux de sortie d'une femme "cohabitante" est également plus bas. Les tableaux croisés comparatifs montrent également que les chefs de ménage de sexe féminin ont, relativement plus souvent que les chefs de ménage de sexe masculin, des enfants à charge (respectivement 82,5% et 55%).

niveau d'études	catégorie d'indemnisation	âge (classe)	partenaire		probabilité par région		
			actif	sexe	flamande	wallonne	Brux.-Cap.
élevé	chef de ménage	$> 25, \leq 35$	non	homme	63,9%	50,4%	46,8%
élevé	cohabitant	$> 25, \leq 35$	oui	homme	78,9%	68,1%	64,9%
élevé	isolé	$> 25, \leq 35$	non	homme	64,1%	50,5%	46,9%
élevé	chef de ménage	$> 25, \leq 35$	non	femme	53,6%	39,8%	36,4%
élevé	cohabitant	$> 25, \leq 35$	oui	femme	74,5%	62,5%	59,1%
élevé	isolé	$> 25, \leq 35$	non	femme	65,3%	51,9%	48,3%

TABLE 3 – Effet d'interaction entre le sexe et la catégorie d'indemnisation

Outre les effets déjà examinés ci-dessus, nous pouvons encore étudier des effets supplémentaires fondés sur le passé professionnel, l'importance du montant journalier, la nationalité et la situation familiale.

Comme on s'y attendait, la base d'admission joue aussi un rôle (voir Tableau 4) : l'admissibilité sur base du travail offre une plus grande probabilité de sortie que l'admissibilité sur base des études. Nous devons toutefois faire remarquer que l'âge peut jouer un rôle plus important. Etant donné que plus de 90% des chômeurs admis sur la base des études ont moins de 25 ans, leur probabilité de sortie est généralement assez élevée. Si l'on compare cependant des chômeurs de la même tranche d'âge, l'admissibilité sur la base du travail donne toujours une probabilité de sortie plus élevée, ce qui n'est pas surprenant, étant donné qu'il s'agit de personnes ayant une expérience professionnelle qui est, de plus, relativement récente.

niveau d'études	catégorie d'indemnisation	âge (classe)	sexe	passé professionnel	probabilité par région		
					flamande	wallonne	Brux.-Cap.
moyen	isolé	$\leq 25$	homme	après études	45,0%	31,9%	28,8%
moyen	isolé	$\leq 25$	homme	après travail	52,8%	39,0%	35,6%

TABLE 4 – Effet du passé professionnel

En ce qui concerne la nationalité, il est clair que les Belges ont une probabilité de sortie sensiblement plus élevée que les étrangers. On peut constater qu'il n'y a pas de grande différence entre les ressortissants européens (en dehors de la Belgique) et les autres étrangers. Nous pouvons supposer que le groupe des étrangers peut présenter des situations très divergentes, ce qui rend l'examen difficile. Nous pouvons remarquer que les étrangers (certainement en dehors de l'UE) sont moins bien placés que les Belges quand il s'agit du niveau d'études. C'est ce qui ressort également du Tableau 6.

niveau d'études	catégorie d'indemnisation	âge (classe)	sexe	nationalité	probabilité par région		
					flamande	wallonne	Brux.-Cap.
moyen	cohabitant	$> 25, \leq 35$	femme	belge	61,6%	47,9%	44,3%
moyen	cohabitant	$> 25, \leq 35$	femme	UE	53,0%	39,2%	35,8%
moyen	cohabitant	$> 25, \leq 35$	femme	autre	51,7%	38,0%	34,7%

TABLE 5 – Effet de la nationalité

	faible	moyen	élevé	
Belgique	13,6%	62,9%	23,6%	100,0%
Autre UE	23,9%	54,8%	21,3%	100,0%
Autre non UE	39,2%	49,2%	11,7%	100,0%

TABLE 6 – Niveau d'études selon nationalité

Une enquête plus approfondie sur la spécificité de la situation des allochtones a été réalisée par Okkerse et Termote (2004). On peut notamment y lire que les femmes allochtones ont un taux d'activité beaucoup plus bas que celui de leurs homologues masculins. De plus, l'enquête de Vertommen et Martens (2005) montre que les belges naturalisés d'origine allochtone revêtent toujours une position inférieure sur le marché de l'emploi.

En ce qui concerne la situation familiale, il peut y avoir un effet lié au partenaire actif et un effet lié au fait d'avoir des enfants. L'effet de la position familiale de l'ayant droit au sein de la famille est examiné plus loin.

Il est en effet certain qu'un rapport existe entre la catégorie d'indemnisation et le fait d'avoir un partenaire actif. Nous devons néanmoins être prudents : alors que la grande majorité des chômeurs dont on connaît le statut actif du partenaire a également la catégorie d'indemnisation " cohabitant " (plus de 91%), l'inverse est moins vrai puisque la plupart des cohabitants n'ont pas de partenaire actif (seulement 35,0% des chômeurs ayant la catégorie d'indemnisation " cohabitant " ont un partenaire actif, d'ailleurs de nombreux cohabitants occupent la position familiale " enfant ", soit quelque 50,7%).

Il est donc judicieux de considérer séparément l'influence qu'exerce un partenaire actif et il semble que le fait d'avoir un partenaire actif ait un effet clairement positif. D'une part, il s'agit donc de personnes qui ont vraisemblablement le statut moins favorable de cohabitant. D'autre part, le fait d'avoir un partenaire actif signifie probablement que l'intéressé est davantage responsable, en tant que partenaire dans une famille avec ou sans enfant, du revenu du ménage (à la différence de la situation où l'intéressé est, par exemple, un enfant habitant chez ses parents).

Pour ce qui concerne le fait d'avoir des enfants, l'effet dépend de l'âge de ces enfants. Alors que le fait d'avoir de très jeunes enfants semble réduire les possibilités de sortir du chômage, le fait d'en avoir de plus âgés (3-12 ans) a plutôt l'effet inverse (quoi que pas significatif), et ce toujours par rapport à une situation sans enfant.

Dans un souci de cohérence, la classe d'âge de la mère suit celle de l'enfant.

niveau d'études	catégorie d'indemnisation	âge (classe)	sexe	âge du cadet	probabilité par région		
					flamande	wallonne	Brux.-Cap.
moyen	cohabitant	$> 25, \leq 35$	femme	néant	61,6%	47,9%	44,3%
moyen	cohabitant	$> 25, \leq 35$	femme	0-2	55,5%	41,6%	38,2%
moyen	cohabitant	$> 25, \leq 35$	femme	3-11	62,2%	48,5%	44,9%
moyen	cohabitant	$> 35, \leq 50$	femme	12-17	57,1%	43,2%	39,7%
moyen	cohabitant	$> 35, \leq 50$	femme	18-24	51,9%	38,2%	34,9%

TABLE 7 – Effet de la charge d'enfant dans la famille, selon l'âge du cadet

Il reste encore un seul facteur à examiner : l'importance du montant journalier (voir Tableau 8). Il est clair qu'il interfère en partie avec d'autres caractéristiques et il est également difficile d'établir, de façon objective, une répartition en classes. Cependant, il y a un effet significativement positif qui se base sur la catégorie du montant journalier le plus élevé. Nous pouvons supposer que l'importance du montant journalier reflète l'importance du plafond salarial (première période d'indemnisation).

niveau d'études	catégorie d'indemnisation	âge (classe)	sexe	montant journ. (classe)	probabilité par région		
					flamande	wallonne	Brux.-Cap.
moyen	isolé	$> 35, \leq 50$	homme	le plus bas	38,8%	26,7%	24,0%
moyen	isolé	$> 35, \leq 50$	homme	moyen	40,4%	28,0%	25,3%
moyen	isolé	$> 35, \leq 50$	homme	le plus élevé	48,2%	34,8%	31,7%

TABLE 8 – Effet du montant journalier

L'interprétation peut en être la suivante : celui qui a eu un salaire élevé est probablement plus fort sur le marché de l'emploi (mieux formé, plus qualifié, expérimenté) et il sera confronté à une plus grande diminution salariale ce qui est un stimulant supplémentaire pour retourner sur le marché de l'emploi.

## 2.2 Modèle alternatif

Lors de la sélection de modèle, il est apparu que la variable " position familiale " n'est de préférence pas reprise dans le modèle, étant donné que les informations se retrouvent en grande partie dans d'autres variables, notamment dans la catégorie d'indemnisation. Il est cependant intéressant d'évaluer un modèle distinct où la position familiale peut bien jouer un rôle important en laissant de côté la catégorie d'indemnisation ainsi que les données sur le partenaire et les enfants. Les tableaux relatifs à ce modèle sont le Tableau 15 (analyse de la variance) et le Tableau 16 (estimation des paramètres).

Les exemples du Tableau 9 se basent sur cet autre modèle (plus de détails dans la Section 4.3) qui est un peu moins bon en termes de vraisemblance statistique et classification, mais qui a toutefois l'avantage de permettre une interprétation en fonction de la position familiale.



niveau d'études	position familiale	âge (classe)	sexe	probabilité par région		
				flamande	wallonne	Brux.-Cap.
moyen	isolé	> 25, ≤ 35	homme	54,7%	40,2%	36,3%
moyen	enfant	≤ 25	homme	55,9%	41,3%	37,4%
moyen	chef de fam. monoparentale	> 25, ≤ 35	homme	45,9%	32,1%	28,6%
moyen	partenaire avec enfants	> 25, ≤ 35	homme	68,1%	54,3%	50,1%
moyen	partenaire sans enfant	> 25, ≤ 35	homme	66,9%	52,9%	48,7%
moyen	isolé	> 25, ≤ 35	femme	55,4%	40,8%	36,9%
moyen	enfant	≤ 25	femme	56,0%	41,4%	37,5%
moyen	chef de fam. monoparentale	> 25, ≤ 35	femme	45,2%	31,5%	28,0%
moyen	partenaire avec enfants	> 25, ≤ 35	femme	54,9%	40,4%	36,5%
moyen	partenaire sans enfant	> 25, ≤ 35	femme	57,2%	42,7%	38,7%

TABLE 9 – Effet de la position familiale

Il est frappant de constater la faible probabilité de sortie pour la position ” chef de famille monoparentale ”, et ce tant pour les hommes que pour les femmes (nous remarquons toutefois que, dans 86% des familles monoparentales, le chef de ménage est une femme). Pour les isolés, peu de différence est notée entre les hommes et les femmes. Ce qui est néanmoins frappant, c’est que les partenaires de sexe féminin avec ou sans enfant ont une probabilité de sortie similaire, alors que celle chez les hommes est sensiblement plus élevée. Chez les hommes, le fait d’avoir des enfants augmente quelque peu la probabilité de sortie, alors qu’il la diminue chez les femmes. Ces constats correspondent plutôt aux schémas traditionnels. Enfin, les enfants sont un peu plus susceptibles de sortir du chômage que les isolés, mais remarquons que, dans l’exemple, on a tenu compte d’une classe d’âge inférieure.

## 3 Conclusions et futures recherches

### 3.1 Conclusions générales

Bien que la question de recherche ne soit pas nouvelle, elle reste d’actualité et, grâce au projet du panel ONEM, pour la première fois, l’ONEM dispose de données lui permettant de la réaliser.

Les conclusions s’inscrivent généralement dans la lignée de faits connus, mais valent néanmoins la peine d’être mentionnées :

- On note des différences importantes entre les régions. Les probabilités de sortie sont les plus élevées en Région flamande, alors qu’elles sont les plus faibles dans la région de Bruxelles-Capitale.
- Le niveau d’études apparaît comme un facteur très important. Les personnes ayant un diplôme d’enseignement supérieur ou une bonne qualification ont une plus grande probabilité de sortie.

- La catégorie d’indemnisation joue un rôle important : la probabilité de sortir du chômage est la plus importante chez les cohabitants et la plus faible chez les chefs de ménage.
- La probabilité diminue avec l’âge, mais cela n’est pas l’effet le plus important.

Il reste naturellement difficile de distinguer tous les effets car ils interagissent souvent les uns avec les autres. Nous avons ainsi constaté que l’effet de la catégorie d’indemnisation n’est pas le même pour les hommes et pour les femmes : la probabilité de sortie est beaucoup plus faible chez les chefs de ménage féminins que chez les chefs de ménage masculins, et une femme cohabitante possède également une probabilité de sortie légèrement plus faible qu’un homme cohabitant.

Cette interaction a aussi été pertinente dans notre enquête sur la position familiale. Dans la position familiale ” partenaire ”, les femmes ont une probabilité de sortie nettement plus faible que les hommes, et avoir des enfants augmente la probabilité de sortie pour les hommes mais la diminue chez les femmes. Le chef d’une famille monoparentale, quelque soit le sexe, a aussi une probabilité de sortie clairement plus faible.

Le fait d’avoir un partenaire actif est toujours positif. Les autres effets s’ajoutent :

- L’effet du passé professionnel est positif.
- Les Belges ont une probabilité plus élevée que les non-Belges et il semble ne pas y avoir de grande différence entre les nationalités UE ou non UE (prises globalement).
- La hauteur du montant journalier a un effet positif ; il ne semble donc pas qu’il représente un frein à l’insertion.

### 3.2 Poursuite de la recherche

Cette étude a tenté de donner une image générale des facteurs qui déterminent si la population des chômeurs nouvellement entrés en 2007 travaillait un an après l’entrée dans le chômage. L’objectif essentiel est de réitérer l’analyse avec des données plus récentes dès qu’elles sont disponibles.

De plus, la méthodologie peut être affinée : comme expliqué dans la Section 1.2.2, notre modèle prend seulement en compte la position du chômeur comme étant actif ou non à un moment donné (à savoir après un an), alors que cette position est connue à 8 moments différents. Il existe des modèles de durée en temps discret (modèles également utilisés en biostatistique dans le cadre d’analyse de survie) qui tiennent compte des données durant les 8 trimestres où la personne est suivie. En outre, ceux-ci tiennent compte du fait que la sortie peut ne pas être observée (censoring). Pour plus de détails, voir Singer et Willett, 2007.

Il va de soi qu’une étude ne donne pas seulement des réponses ou des éléments de réponse, mais qu’elle suscite aussi de nouvelles questions. Nous pouvons dès lors

nous demander ce qu'il en est sur le plan de la qualité et de la durabilité de l'emploi (secteur, salaire, ...). Il ressort également de la Section 1.2.2 que de nombreuses personnes ont une carrière instable.

Une piste évidente serait d'élargir la population pour inclure les personnes ayant déjà un passé en tant que chômeurs. Dans ce cas la durée ressortirait comme facteur important. La prise en compte de la durée engendre quelques difficultés spécifiques. Voir par exemple Heylen, 2010 pour une discussion plus détaillée.

Une suite logique de l'étude pourrait concerner la ventilation de l'analyse par sexe et par région.

## 4 Appendice technique et méthodologique

### 4.1 Panel ONEM - échantillonnage

Le panel ONEM est basé sur un échantillon prélevé dans la population de tous les allocataires de l'ONEM. Pour cette population, des données supplémentaires en matière de sécurité sociale, d'emploi et de position familiale sont ajoutées en provenance du Datawarehouse Marché de l'Emploi et Protection Sociale de la Banque Carrefour de la Sécurité Sociale.

Sans entrer dans les détails, il est utile de donner une court descriptif du mécanisme d'échantillonnage car cela permet de mieux comprendre la méthodologie suivie dans cette étude.

Tout d'abord, le principe est que seuls les bénéficiaires ONEM peuvent être repris dans le panel.

Cela peut se faire à différents moments :

- Au premier trimestre de l'année 2000 (échantillon de départ) :
  - Echantillon aléatoire (33,33%) du *stock* de tous les bénéficiaires ONEM.
  - Stratifié selon l'arrondissement et le statut ONEM (regroupement).
- Un trimestre plus tard, à partir du 2ème trimestre de l'année 2000 (échantillon supplémentaire) :
  - Echantillon aléatoire (25%) issu du *flux* [nouvelle entrée : personnes qui, dans ce trimestre, ont été des bénéficiaires ONEM, mais qui n'étaient pas connues à l'ONEM dans la période précédant ce trimestre (antécédents sur une période de 5 ans mais à partir de l'année 2000)].
  - Stratifié selon la région et le statut ONEM.

Deuxièmement, chaque personne de l'échantillon est suivie au minimum 5 ans et minimum 36 mois après la cessation de tout lien avec l'ONEM. Après cette période, la personne sortira du panel (mais elle peut éventuellement être reprise en considération plus tard lors d'un échantillon supplémentaire), et il en va également ainsi lorsque la personne décède ou atteint l'âge de 65 ans.

Le suivi signifie que, par trimestre (ou par an si les données ne sont pas disponibles par trimestre), des données sont mises à jour à partir de l'entrée jusqu'à la situation la plus récente possible, selon la disponibilité dans le datawarehouse.

L'ayant droit ONEM est la personne clé sur laquelle portent toutes les variables par le biais d'un numéro (fictif) anonyme utilisé comme variable clé.

Pour pouvoir décrire la situation familiale, la méthodologie du datawarehouse est appliquée : un ménage est censé être constitué autour de l'identité de ladite personne de référence du ménage (connue en tant que telle par l'autorité administrative). Nous connaissons le lien familial du bénéficiaire ONEM par rapport à cette personne de référence (repris sous ce que l'on appelle " position familiale "), mais

les liens familiaux avec les autres membres du ménage restent inconnus. Dans la plupart des cas, en combinant suffisamment les données, nous pouvons néanmoins avoir une bonne idée du contexte socio-économique de l'intéressé. Outre la personne de référence du ménage, on déduit également du datawarehouse, de nouveau sur la base des données administratives enregistrées, si le bénéficiaire ONEM a un partenaire.

Afin d'établir le lien avec notre recherche, remarquons que :

- La population examinée se compose de bénéficiaires ONEM issus des 4 échantillonnages supplémentaires de 2007.
- Nous nous limitons aux bénéficiaires ONEM ayant un statut de chômeur complet indemnisé et demandeur d'emploi.

Toutes les variables du panel ONEM ne sont pas pertinentes pour la présente recherche. Le paragraphe suivant présente un aperçu des variables utilisées.

## 4.2 Description des variables pour la régression logistique

Toutes les variables utilisées proviennent du Datawarehouse Marché de l'Emploi et Protection Sociale ou en sont déduites. Pour plus de détails, nous renvoyons à la documentation fournie par la BCSS (voir par exemple le site Web : <http://www.ksz-bcss.fgov.be/>).

Bien qu'il soit possible de travailler avec des variables continues, nous préférons la construction de variables catégorielles (hypothèse moins stricte). A ce stade, nous pouvons donner un aperçu des variables utilisées et de la répartition en catégories.

La liste du Tableau 10 présente toutes les variables explicatives possibles, ainsi que la variable expliquée binaire " sortie ".

La nomenclature de la position socio-économique (pas reprise dans la liste) est la variable cruciale car elle donne un résumé assez détaillé de la position socio-économique de la personne au dernier jour du trimestre (avec une répartition principale entre " actif ", " demandeur d'emploi ", " inactif " et " autre "). On en déduit si la personne est demandeur d'emploi ou actif. Sur la base de la nomenclature, nous pouvons également déterminer si le partenaire éventuel est actif. Cela suppose bien entendu que la relation entre les partenaires est officielle et enregistrée. La fiabilité de cette donnée ne sera donc pas optimale. Etant donné qu'il n'y avait pas de différences significatives entre les situations où il n'y a pas de partenaire et les cas où le partenaire occupe une position " demandeur d'emploi ", " inactif " ou " autre ", une distinction est uniquement établie entre la situation où " il y a un partenaire actif " et tous les autres cas [c'est pourquoi nous définissons une variable binaire " Wptn " (partenaire actif)].

En principe, les variables explicatives sont déterminées au moment où la personne devient chômeuse. S'il s'agit de données mensuelles (comme l'ONEM), le dernier

Concept et origine	Variable	Valeurs
Catégorie d'indemnisation Fiche61 (ONEM) regroupement	"vergcats"	isolé cohabitant chef de ménage
Base d'admission (passé professionnel) Fiche7 (ONEM) regroupement	"arbeidsverleden"	1 : après travail 0 : après études
Hauteur du montant journalier Dgndmnd.klasse (ONEM)	"bedrk"	0 : le plus bas ( $\leq$ 25ème centile) 1 : groupe moyen 2 : le plus élevé ( $>$ 75ème centile)
Domicile (région) Arrondissement (registre national) regroupement	"gewest"	B : Région Bruxelles-Capitale Fl : Région flamande W : Région wallonne
Sexe (registre national)	"geslacht"	0 : homme 1 : femme
Age (en années) répartition	"leeftk4"	0 : $\leq$ 25 1 : $>$ 25, $\leq$ 35 2 : $>$ 35, $\leq$ 50 3 : $>$ 50
Nationalité regroupement code pays (registre national)	"nat"	0 : belges 1 : UE hors Belgique 2 : autres pays
Position familiale (par rapport à la personne de référence registre national) LIPRO	"hhpos"	isolé enfant chef de famille monoparentale partenaire avec enfants partenaire sans enfant
Age du cadet (ans) classes	"jongste_kind_leeftk"	0 : néant ou plus de 25 ans 1 : 0-2 2 : 3-11 3 : 12-17 4 : 18-24
Partenaire actif nomenclature en registre national	"Wptn"	0 : Partenaire inconnu/ non-actif 1 : Partenaire actif
Sortie vers travail (4 trimestres après l'entrée)	"Wout"	0 : non-actif 1 : actif
Niveau d'études (premier de 7 chiffres)	"stuck"	0 : faible (primaire + premier degré sec.) 1 : moyen (enseignement secondaire) 2 : élevé (enseignement supérieur) 3 : contrats d'apprentissage

TABLE 10 – Liste des variables

mois du trimestre sera alors considéré. S'il s'agit de données du registre national, c'est la situation à la fin de l'année 2007 qui sera prise en compte.

Tel que repris ci-dessus, il faut être prudent en interprétant les données de la situation familiale. Si l'ayant droit occupe la position familiale "partenaire avec des enfants", nous pouvons partir de l'hypothèse que les enfants présents au sein du ménage sont à charge de l'ayant droit (sachant que c'est vrai dans la plupart des cas). Si l'ayant droit est cependant lui-même connu comme étant "enfant" (par rapport à la personne de référence), nous ne pourrions pas en déduire que les jeunes enfants du ménage sont à charge de l'ayant droit, il s'agira souvent de petits frères ou de petites sœurs. Dans ce cas où l'ayant droit est dans la position "enfant",

nous partirons de l'hypothèse que les enfants du ménage ne sont pas à charge de l'ayant droit, à moins que cette personne ait la catégorie d'indemnisation " chef de ménage " (quelque 300 cas). (Concrètement : variable "jongste\_kind\_leeftk" = 0 si "hhpos" = "enfant", sauf si "vergcats" = "chef de ménage").

### 4.3 Sélection du modèle

Dans la liste des variables explicatives possibles, aucune ne peut être a priori exclue. Des recoupements sont néanmoins clairement possibles. C'est ainsi que la position familiale est nettement moins informative lorsque nous disposons de la catégorie d'indemnisation, de la présence d'enfants dans le ménage et de la présence ou non d'un partenaire actif.

Si ces quatre variables sont reprises ensemble dans un modèle, il apparaît que la position familiale n'est plus significative. C'est la raison pour laquelle la variable " hhpos " (position fam.) est retirée du modèle de base. Elle est toutefois examinée dans un deuxième modèle.

Les critères utilisés pour l'évaluation des modèles sont la *deviance* (vraisemblance statistique, voir par exemple Agresti, 2003) du modèle et la capacité de classification (prédiction) du modèle. Cette classification se déroule comme suit : nous utilisons le modèle pour calculer, au cas par cas, la probabilité de sortie et si celle-ci est (strictement) plus grande que la probabilité de sortie globale, le modèle prédit la sortie, si elle est moins importante, le modèle prédit la non-sortie (bien entendu quand on parle de sortie il s'agit toujours de sortie vers l'emploi). La capacité de classification du modèle est évaluée ensuite en comparant les prédictions avec les observations réelles (de sortie ou non).

Nous pouvons ensuite calculer le pourcentage de prédictions correctes, la spécificité (% de non-sortie prédit correctement) , la sensibilité (% de sortie prédit correctement), le taux de faux positifs et le taux de faux négatifs.

Une fois les effets principaux établis, il apparaît que l'introduction d'effets d'interaction n'améliore pas tant la qualité de prévision du modèle, que sa vraisemblance statistique. Les stratégies de choix de modèle automatique proposent l'ajout de plusieurs interactions, mais ils conduisent à un modèle qui est pratiquement illisible au niveau de l'interprétation. Nous préférons nous limiter à l'ajout d'une interaction significative qui a également une interprétation claire : l'interaction entre le sexe et la catégorie d'indemnisation.

#### 4.3.1 Modèle de base

Les variables qui sont reprises dans le modèle de base sont (la catégorie de référence est indiquée entre parenthèses) :

région ("flamande"), niveau d'études ("faible"), catégorie d'indemnisation ("isolé"), classe d'âge ( $\leq 25$ ), nationalité ("belge"), importance du montant journalier ("le plus bas"), partenaire actif ("néant"), âge du cadet ("néant"), sexe ("homme"), base d'admission ("après études").

La vraisemblance statistique indique la probabilité de l'effet, plus elle est élevée, plus l'effet est important. Cette information est reprise dans le tableau d'analyse de la variance, voir Tableau 11 (3ème colonne). Il en résulte que toutes les variables ont un effet significatif, comme indiqué par les codes repris dans la dernière colonne (seuil de signification) :  $0 \leq '***' < 0.001 \leq '**' < 0.01 \leq '*' < 0.05 \leq '.' < 0.1$

*Remarque générale : les noms qui apparaissent dans la première colonne de certains tableaux sont les noms techniques des variables. Pour leur signification, le lecteur pourra se référer au Tableau 10.*

	Df	Deviance	Resid. Df	Resid. Dev	$P(>  Chi )$	
NULL			14577	19865		
factor(gewest)	2	404.83	14575	19460	< 2.2e-16	***
factor(leefk4)	3	224.26	14572	19235	< 2.2e-16	***
geslacht	1	14.77	14571	19221	0.0001212	***
factor(bedrk)	2	30.84	14569	19190	2.011e-07	***
factor(vergcat)	2	283.09	14567	18907	< 2.2e-16	***
factor(nat)	2	71.21	14565	18836	3.451e-16	***
arbeidsverleden	1	7.05	14564	18828	0.0079084	**
factor(stuck)	3	294.73	14561	18534	< 2.2e-16	***
Wptn	1	30.14	14560	18504	4.018e-08	***
factor(jongste_kind_leefk)	4	22.30	14556	18481	0.0001744	***
geslacht :factor(vergcat)	2	16.64	14554	18465	0.0002435	***

TABLE 11 – Analyse de variance dans le modèle de base

En ce qui concerne l'interprétation des effets, nous nous intéressons bien entendu surtout aux estimations des paramètres données dans le Tableau 12.

Mais rappelons d'abord le modèle de régression logistique :

$$\log\left(\frac{p}{1-p}\right) = \alpha + \beta_1 x_1 + \dots + \beta_k x_k$$

où  $p$  représente (formellement) l'espérance mathématique de la variable "Wout" (représentant la sortie vers le travail) considérée comme variable aléatoire binaire. Dans l'équation ci-dessus, l'expression à gauche représente la transformation *logit* de  $p$ , l'expression à droite est un modèle linéaire, fonction des variables explicatives  $x_i$  ( $i = 1, \dots, k$ ) qui est censé être valable. L'estimation du modèle consiste alors à estimer les valeurs des paramètres  $\alpha$  (constant) et  $\beta_i$  ( $i = 1, \dots, k$ ) (coefficient devant l'effet correspondant  $x_i$ ). Les estimations des paramètres se trouvent dans la 2ème colonne du tableau. Pour les variables binaires, le coefficient donné



correspond à la valeur 1 de la variable (0 étant la référence). Pour les variables à plusieurs catégories, on retrouve une ligne pour chaque valeur, sauf à nouveau pour la valeur de référence (tous les effets des catégories de référence contribuent au coefficient constant ("intercept")). Le signe (+ ou -) du coefficient indique si la catégorie amène à un logit (et donc une probabilité de sortie) plus élevé, ou moins élevé (respectivement). Dans la 3ème colonne, on trouve les erreurs standard de l'estimation : plus elles sont petites, plus l'effet sera significatif (indiqué dans la dernière colonne avec les codes comme introduit auparavant).

	Estimate	Std. Error	z value	$Pr(>  z )$	
(Intercept)	-0.626718	0.091206	-6.871	6.35e-12	***
factor(gewest)B	-0.702862	0.054511	-12.894	< 2e-16	***
factor(gewest)W	-0.558702	0.039613	-14.104	< 2e-16	***
factor(leefk4)1	-0.129889	0.059044	-2.200	0.027815	*
factor(leefk4)2	-0.507433	0.067229	-7.548	4.43e-14	***
factor(leefk4)3	-1.309116	0.093739	-13.966	< 2e-16	***
geslacht	0.055753	0.081495	0.684	0.493896	
factor(bedrk)1	0.048225	0.076459	0.631	0.528216	
factor(bedrk)2	0.366653	0.102735	3.569	0.000358	***
factor(vergcat)chef de famille	-0.004719	0.089224	-0.053	0.957821	
factor(vergcat)cohabitant	0.428101	0.072959	5.868	4.42e-09	***
factor(nat)1	-0.355707	0.075428	-4.716	2.41e-06	***
factor(nat)2	-0.406126	0.081452	-4.986	6.16e-07	***
arbeidsverleden	0.312315	0.078005	4.004	6.23e-05	***
factor(stuck)1	0.377522	0.054813	6.887	5.68e-12	***
factor(stuck)2	0.973637	0.063128	15.423	< 2e-16	***
factor(stuck)3	0.448160	0.114261	3.922	8.77e-05	***
Wptn	0.312938	0.052873	5.919	3.25e-09	***
factor(jongste_kind_leefk)1	-0.253844	0.067751	-3.747	0.000179	***
factor(jongste_kind_leefk)2	0.025561	0.072405	0.353	0.724063	
factor(jongste_kind_leefk)3	0.187604	0.096445	1.945	0.051752	.
factor(jongste_kind_leefk)4	-0.018996	0.103630	-0.183	0.854559	
geslacht :factor(vergcat)chef de famille	-0.485615	0.126349	-3.843	0.000121	***
geslacht :factor(vergcat)cohabitant	-0.303691	0.092354	-3.288	0.001008	**

TABLE 12 – Estimation de paramètres dans le modèle de base

Remarquons que l'effet du sexe n'est pas significatif, ce qui s'explique par l'introduction d'un effet d'interaction entre le sexe et la catégorie d'indemnisation. Cette interaction est un effet significatif, par exemple un effet négatif pour les chefs de famille féminins.

Pour avoir une idée de la qualité de classification, le lecteur trouvera dans le Tableau 13 les nombres de sorties ou non, tant dans les valeurs observées que dans les prédictions. Le Tableau 14 reprend les indicateurs introduits dans la Section 4.3.

observation 1	observation 0	prédiction 1	prédiction 0
6170	8408	6834	7734

TABLE 13 – Prédictions de sortie ou non dans le modèle de base

sensibilité	specificité	taux de faux positif	taux de faux négatif	prédiction correcte
61,5%	63,9%	44,4%	30,6%	62,9%

TABLE 14 – Indicateurs de classification dans le modèle de base

### 4.3.2 Modèle alternatif

Comme expliqué précédemment, le modèle alternatif contient la variable concernant la position familiale (variable "hhpos", avec valeur de référence "isolé") mais pas les variables concernant la catégorie d'indemnisation, l'effet d'un partenaire au travail ni la variable concernant l'âge du cadet. Comme dans le modèle de base nous introduisons l'effet d'interaction entre sexe et position familiale.

Les Tableaux 15 et 16 donnent respectivement l'analyse de la variance ANOVA et l'estimation des paramètres du modèle. Les Tableaux 17 et 18 concernent la classification avec ce modèle.

	Df	Deviance	Resid. Df	Resid. Dev	$P(>  Chi )$	
NULL			14577	19865		
factor(gewest)	2	404.83	14575	19460	<2.2e-16	***
factor(leefk4)	3	224.26	14572	19235	< 2.2e-16	***
geslacht	1	14.77	14571	19221	0.0001212	***
factor(bedrk)	2	30.84	14569	19190	2.011e-07	***
factor(hhpos)	4	121.26	14565	19069	< 2.2e-16	***
factor(nat)	2	91.72	14563	18977	< 2.2e-16	***
arbeidsverleden	1	49.40	14562	18927	2.088e-12	***
factor(stuck)	3	329.32	14559	18598	< 2.2e-16	***
factor(jongste_kind_leefk)	4	32.10	14555	18566	1.825e-06	***
geslacht :factor(hhpos)	4	47.49	14551	18519	1.205e-09	***

TABLE 15 – Analyse de variance dans le modèle alternatif

	Estimate	Std. Error	z value	Pr(>  z )	
(Intercept)	-0.43537	0.08990	-4.843	1.28e-06	***
factor(gewest)B	-0.75275	0.05443	-13.829	< 2e-16	***
factor(gewest)W	-0.58694	0.03956	-14.836	< 2e-16	***
factor(leeftk4)1	-0.10911	0.06158	-1.772	0.076430	.
factor(leeftk4)2	-0.54137	0.07000	-7.734	1.04e-14	***
factor(leeftk4)3	-1.41851	0.09887	-14.347	< 2e-16	***
geslacht	0.02653	0.07962	0.333	0.738919	
factor(bedrk)1	-0.24013	0.06648	-3.612	0.000304	***
factor(bedrk)2	-0.08258	0.08663	-0.953	0.340454	
factor(hhpos)chef de famille monoparentale	-0.35342	0.22755	-1.553	0.120393	
factor(hhpos)enfant	-0.06168	0.09179	-0.672	0.501604	
factor(hhpos)partenaire avec enfants	0.56884	0.12217	4.656	3.22e-06	***
factor(hhpos)partenaire sans enfants	0.51369	0.08978	5.722	1.05e-08	***
factor(nat)1	-0.41245	0.07545	-5.467	4.59e-08	***
factor(nat)2	-0.53558	0.08127	-6.590	4.40e-11	***
arbeidsverleden	0.55140	0.07227	7.630	2.36e-14	***
factor(stuck)1	0.42217	0.05468	7.720	1.16e-14	***
factor(stuck)2	1.04222	0.06309	16.519	< 2e-16	***
factor(stuck)3	0.49819	0.11413	4.365	1.27e-05	***
factor(jongste_kind_leeftk)1	-0.21294	0.11575	-1.840	0.065826	.
factor(jongste_kind_leeftk)2	0.04341	0.10026	0.433	0.665055	
factor(jongste_kind_leeftk)3	0.17094	0.09487	1.802	0.071578	.
factor(jongste_kind_leeftk)4	0.21727	0.08566	2.536	0.011200	*
geslacht :factor(hhpos)chef de famille monoparentale	-0.05378	0.23123	-0.233	0.816073	
geslacht :factor(hhpos)enfant	-0.02163	0.09819	-0.220	0.825682	
geslacht :factor(hhpos)partenaire avec enfants	-0.58695	0.11020	-5.326	1.00e-07	***
geslacht :factor(hhpos)partenaire sans enfants	-0.43753	0.12370	-3.537	0.000404	***

TABLE 16 – Estimation de paramètres dans le modèle alternatif

observation 1	observation 0	prédiction 1	prédiction 0
6170	8408	6866	7712

TABLE 17 – Prédictions de sortie ou non dans le modèle alternatif

sensibilité	specificité	taux de faux positif	taux de faux négatif	prédiction correcte
61,4%	63,4%	44,8%	30,9%	62,5%

TABLE 18 – Indicateurs de classification dans le modèle alternatif

### 4.3.3 Exemple de calcul de probabilité

Nous prenons le premier exemple de la Figure 7 et reprenons dans le Tableau 19 une liste des caractéristiques avec les coefficients correspondants dans la dernière colonne : quand il s'agit de la référence on indique 0. La somme de tous les coefficients (y compris la constante) nous donne le logit de  $p$  qui s'élève ici à  $-0.2709$  ; on isole  $p$  par la transformation inverse :  $p = \frac{\exp(-0.2709)}{1+\exp(-0.2709)} = 0.4327$

Variable	valeur	paramètre
Catégorie d'indemnisation	chef de famille	-0.0047
Base d'admission	1 : après travail	0.3123
Niveau du montant journalier	1 : groupe moyen	0.04822
Région	flamande	0 (référence)
Sexe	homme	0 (référence)
Age	0 : $\leq 25$	0 (référence)
Nationalité	0 : belge	0 (référence)
Age du cadet	néant	0 (référence)
Partenaire actif	néant	0 (référence)
Niveau d'études	faible	0 (référence)
Intercept		-0.6267
logit		-0.2709

TABLE 19 – Caractéristiques de l'exemple et calcul de la probabilité de sortie

## Références

- [1] A. Agresti, *An introduction to Categorical Data Analysis* (New York : Wiley, 2007).
- [2] V. Heylen, *Duration dependence in Flemish unemployment data* (Leuven, WSE Report, 2010).
- [3] L. Okkerse en A. Termote, *Hoe vreemd is vreemd op de arbeidsmarkt, over de allochtone arbeidskrachten in België* (NIS, Brussel : Statistische studie 111, 2004).
- [4] J. Singer en J. Willett, *Applied Longitudinal Data Analysis* (New York : Oxford University Press, 2003).
- [5] S. Vertommen en A. Martens, *Allochtone werknemers op lokale arbeidsmarkten* (Leuven, Over.Werk, Tijdschrift van het Steunpunt WAV, 2 (2005)).