

Comment améliorer l'efficacité des formations pour les demandeurs d'emploi grâce aux outils du Big Data?

Les formations des demandeurs d'emploi sont-elles efficaces pour améliorer leurs perspectives d'emploi et si, oui, pour quels chômeurs en particulier ? Ce numéro de Regards économiques résume les résultats d'une étude qui s'est penchée sur ces deux questions à partir d'un grand ensemble de données administratives sur des demandeurs d'emploi résidant en Flandre et en utilisant les outils de programmation et d'analyse statistique des Big Data.

Bart Cockx¹

En ces temps marqués par le coronavirus, le nombre de demandeurs d'emploi est en hausse. La question de savoir quelles mesures d'aide à l'emploi sont efficaces est dès lors d'une actualité brûlante. L'efficacité dont il est question ici ne se mesure pas simplement à l'aune de la proportion des bénéficiaires de l'aide ayant trouvé un emploi. Ces bénéficiaires ne sont en effet pas choisis au hasard et chacun d'eux a une chance de trouver un emploi sans l'aide en question. L'aide sera d'autant plus efficace qu'elle provoque une forte élévation de l'indicateur d'insertion en emploi retenu, par comparaison à une situation de référence : l'absence de cette aide.² On parlera ci-dessous de «gain» produit par une mesure d'aide si celle-ci a un tel «effet causal» favorable sur la trajectoire de ses bénéficiaires. Hiérarchiser des mesures d'aide à l'emploi selon leur «gain» en moyenne est donc un premier enjeu abordé dans ce numéro de *Regards économiques*. Toutefois, l'effet causal d'une mesure n'est a priori pas uniforme dans la population. L'enjeu étudié ensuite dans ce texte est dès lors celui de l'adéquation entre l'aide et le profil – les caractéristiques – du demandeur d'emploi : quelle mesure d'aide est-elle la plus efficace pour un profil donné ?

Fin 2018, une collaboration s'est mise en place entre le VDAB (qui est le service public pour l'emploi en Flandre) et les professeurs Bart Cockx (Université de Gand) et Michael Lechner (Université de Saint-Gall en Suisse). Elle visait à répondre à la question d'efficacité mentionnée ci-dessus dans le contexte de dispositifs de formation et d'orientation professionnelle. Pour y parvenir, les chercheurs

¹ L'auteur remercie Muriel Dejemeppe et Bruno Van der Linden pour leur aide à traduire le texte et à le rendre plus intelligible. Il remercie également Vincent Bodart pour ses commentaires. Toute erreur restant est la responsabilité de l'auteur.

² Le *Regards économiques* Numéro 124 de William Parienté est à ce propos une référence recommandée : https://www.regards-economiques.be/index.php?option=com_reco&view=article&cid=160.



ont utilisé les outils de programmation et d'analyse statistique des Big Data. Ils les ont appliqués à un grand ensemble de données administratives anonymisées sur des demandeurs d'emploi résidant en Flandre et renseignant quelques-unes de leurs caractéristiques personnelles, leur parcours dans le chômage et en emploi, ainsi que leur participation à des programmes de formation. Cette étude montre que si le VDAB attribuait les formations aux demandeurs d'emploi selon l'algorithme informatique développé par les chercheurs, on obtiendrait une augmentation de 20 % de la durée passée en emploi grâce aux formations.

L'efficacité des formations en terme de remise à l'emploi

L'efficacité de trois mesures de formation a été étudiée :

1. Des formations professionnelles courtes, c'est-à-dire d'une durée de moins de 6 mois (ci-dessous désignées par «FPC»);
2. Des formations professionnelles de durée moyenne, allant de 6 à 10 mois (ci-dessous désignées par «FPM»);
3. Des sessions d'orientation professionnelle (ci-dessous désignées par «SO»).

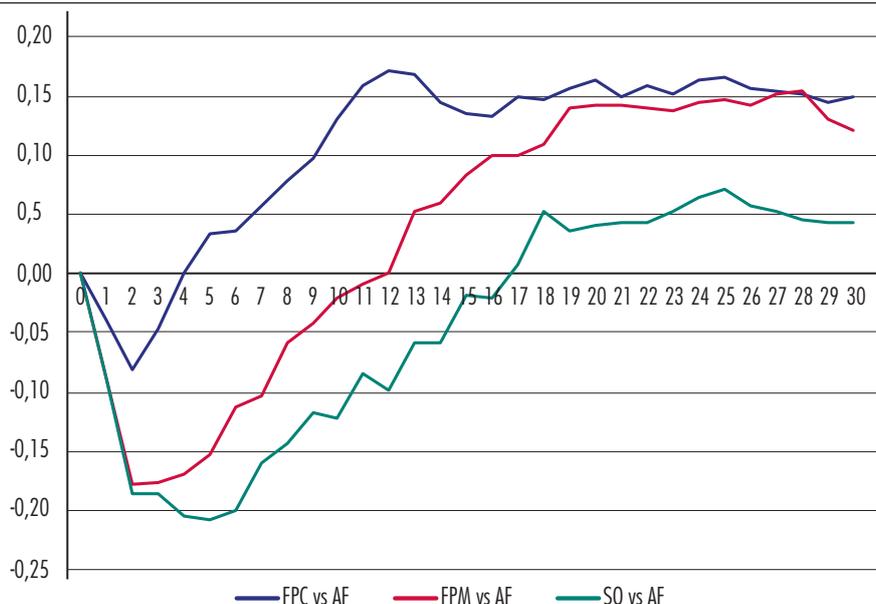
Les chercheurs concluent que, sur un horizon de temps de deux années et demi après le début de la formation,³ les bénéficiaires des formations professionnelles courtes passent en moyenne 19,5 mois en emploi. Pour les formations de durée moyenne et les sessions d'orientation, cette durée est respectivement de 17 et 15 mois. En l'absence de participation à l'une de ces formations (situation désignée ci-dessous par «AF»), ces mêmes personnes auraient en moyenne passé 16 mois en emploi.

Les *formations professionnelles courtes* sont donc *les plus efficaces* : elles augmentent en moyenne la durée passée en emploi de 3,5 mois (soit une augmentation de 21 %). La participation à une formation professionnelle de durée moyenne entraîne un gain d'1,1 mois (+ 7 %) tandis qu'une session d'orientation raccourcit la durée d'emploi de 1,3 mois (- 8 %). Selon les chercheurs, cet effet négatif doit être attribué à l'horizon temporel limité de l'étude. Pendant un laps de temps initial, l'effet sur l'accès à l'emploi des trois types de formations est négatif car durant leur participation au programme, les personnes ne recherchent pas de l'emploi ou le font moins intensément (voir le graphique 1). Dans le cas des participants aux sessions d'orientation professionnelle, cet effet négatif se prolonge davantage car il est fréquent que ces personnes participent ensuite à une formation plus complète.

Le graphique 1 décrit les différences moyennes entre le taux de retour à l'emploi de l'un des trois dispositifs de formation et le taux de retour à l'emploi en l'absence de participation à une formation. Ces différences sont mesurées de 2 à 30 mois depuis l'entrée dans le dispositif. Par exemple, cinq mois après l'entrée dans une formation professionnelle courte (FPC), les chances d'accès à l'emploi des bénéficiaires ont augmenté de 3 points de pourcentage par rapport à l'absence de formation (AF); par contre, à ce moment-là, ces chances d'insertion en emploi sont pour les bénéficiaires des formations de durée moyenne et des sessions d'orientation d'encore respectivement 15 et 21 points de pourcentage *plus faibles* qu'en l'absence de formation.

³ La période de 2,5 années nous semble suffisante pour effectuer une comparaison pertinente des «gains» générés par des formations de durées différentes. Nous constatons en effet que les effets des formations sur le taux d'emploi se stabilisent environ 18 mois après l'entrée dans le programme, et que le classement des différents programmes en terme d'efficacité de remise à l'emploi reste similaire tout au long de la période observée. Une analyse des effets sur une plus longue période ne pourrait pas mener à une conclusion différente par rapport à l'efficacité relative des programmes de formation étudiés.

Graphique 1. Efficacité des trois dispositifs de formation sur le taux de retour à l'emploi selon la durée écoulée depuis l'entrée en formation



Note de lecture : L'axe horizontal représente la durée écoulée en mois depuis l'entrée en formation, tandis que l'axe vertical donne la différence entre le taux de retour en emploi des participants à un certain type de formation (FPC, FPM ou SO) et celui des non-participants (AF). Ainsi, 7 mois après l'entrée dans une formation professionnelle courte (FPC), les bénéficiaires d'une telle formation ont un taux de retour à l'emploi supérieur de 0,05 (soit + 5 points de pourcentage) à celui des non-participants (AF). Par contre, à ce même moment, les bénéficiaires d'une formation professionnelle de durée moyenne (FPM) ont un taux de retour à l'emploi inférieur de 0,10 (soit - 10 points de pourcentage) à celui des non-participants (AF).

L'efficacité la plus grande pour les immigrés récents

L'adéquation entre le programme de formation et la personne sans emploi a de l'importance. D'un chômeur à l'autre, les chercheurs obtiennent en effet de grandes différences d'efficacité d'un dispositif donné. Ils établissent ainsi que les formations ont le plus d'impact sur le retour à l'emploi des immigrés récents dont la connaissance du néerlandais est limitée. Ceci s'explique sans doute par la plus grande ampleur des obstacles rencontrés par ces personnes sur le marché du travail : le gain de la formation en matière d'employabilité est par conséquent supérieur pour ce groupe.

Le supplément de gains d'emploi peut atteindre 20 % par comparaison avec l'affectation actuelle des chômeurs aux formations

L'outil de programmation et d'analyse statistique développé par les chercheurs permet de prévoir le gain d'emploi qu'un chômeur quelconque donné peut tirer de sa participation à un dispositif de formation déterminé. Si ces prévisions étaient utilisées pour orienter les demandeurs d'emploi vers les programmes de formation qui ont été évalués par les chercheurs, la durée passée en emploi augmenterait de quasi 20 % en moyenne. Ceci provient de ce que l'objectif, fréquent, d'un taux de placement élevé à l'issue d'une formation ne correspond pas toujours à un haut *gain* en matière de mise à l'emploi. Un conseiller du service public peut obtenir un haut taux de placement en sélectionnant les personnes au profil favorable qui ont déjà une chance élevée de trouver un emploi sans participer à une formation. A l'opposé, les chercheurs se sont intéressés au *gain* produit par la formation. Si le chômeur n'a guère de chance de trouver un emploi sans la formation, un gain d'emploi élevé peut découler d'un dispositif et ce, même s'il y correspond à un taux de placement modeste.

Un processus d'attribution transparent

Recourir à des outils informatiques pour attribuer les formations aux demandeurs d'emploi présente l'inconvénient de ne pas fournir la raison pour laquelle telle formation convient mieux qu'une autre à tel demandeur d'emploi. Sur base des caractéristiques disponibles sur le demandeur d'emploi, le programme informatique «prédit», telle une boîte noire, le type de programme de formation auquel devrait participer le demandeur d'emploi. Pour remédier à ce problème de manque de transparence dans un mode d'attribution qui dépendrait d'un seul programme informatique, les chercheurs ont développé un autre algorithme dans le but de trouver un mode d'attribution simple et transparent approchant au mieux celui décrit précédemment. Ce mode d'attribution s'appuie sur seulement trois caractéristiques du demandeur d'emploi : son âge, son pays de naissance et son degré de connaissance du néerlandais. Ce mode d'attribution simplifié aboutit à presque 80 % des gains d'emploi engendrés par le mode d'attribution recourant à toutes les informations disponibles sur la personne. Ceci signifie qu'en réallouant les demandeurs d'emploi selon ce mode d'attribution reposant sur trois caractéristiques seulement, le gain moyen en durée d'emploi par rapport à l'affectation actuelle des chômeurs dans les formations est de 15 %.

Restrictions légales ou politiques

Des raisons légales ou politiques pourraient interdire l'usage du pays de naissance comme critère d'attribution. La méthode décrite demeure néanmoins applicable. Dans ce cas, il suffit d'indiquer à l'algorithme que le pays de naissance ne peut servir au processus d'attribution.

Implications pour le VDAB ou d'autres services publics de l'emploi

Cette recherche démontre comment des outils informatiques et statistiques peuvent apporter un soutien au VDAB lors de l'attribution des politiques d'aide à l'emploi pour les chômeurs. Elle démontre que les gains d'emploi peuvent être conséquents et que pour les obtenir il n'y a pas lieu de recourir à des règles opaques qui sont susceptibles de cacher un traitement inéquitable des demandeurs d'emploi. Le VDAB a l'intention d'étudier davantage dans quelle mesure ce nouvel instrument peut être mis en œuvre.

Données et méthodes

Cette recherche s'appuie sur des données administratives rendues anonymes, que le VDAB a recueillies à propos de ses usagers. Ces données contiennent un éventail de caractéristiques socio-économiques au sujet des demandeurs d'emploi et une information détaillée à propos de leur historique sur le marché du travail et de leur participation passée à des dispositifs de formation et d'orientation professionnelle. Ces informations sont nécessaires pour pouvoir estimer un effet causal des dispositifs sur les chances d'accès à l'emploi selon la méthode de l'appariement («matching»).

La méthode d'appariement consiste à jumeler («appairer») chaque individu dans le groupe de participants aux formations avec un individu dans le groupe des non-participants sur la base d'une série de caractéristiques observables telles que le genre, l'âge, le niveau d'études, le degré de connaissance du néerlandais ou encore la durée passée en chômage. L'idée est d'appairer les participants avec des non-participants ayant exactement les mêmes caractéristiques sur les variables sélectionnées. L'effet causal des formations qui a été reproduit dans le graphique 1 est estimé en prenant la moyenne des écarts en termes de probabilité d'emploi entre chaque participant et son «semblable» non participant.

Les chercheurs ont réalisé leur analyse sur base de 60.000 personnes de 21 à 55

... **Données et méthodes**

ans qui ont perdu leur emploi entre décembre 2014 et juin 2016 et ont droit à des allocations de chômage. Parmi celles-ci, 1.305 personnes ont entamé une formation professionnelle courte, 1.220 une formation de durée moyenne et 1.115 une session d'orientation professionnelle endéans les 9 premiers mois d'inoccupation. La position occupée sur le marché du travail a été suivie durant 30 mois après le début de leur participation à un des trois dispositifs.

La méthode d'appariement standard permet de mesurer un effet *moyen* des formations (c'est-à-dire pour l'ensemble des participants) sur l'accès à l'emploi. Afin de déterminer l'hétérogénéité des effets des formations selon le profil des participants, les chercheurs ont utilisé la technique de l'apprentissage automatique (*machine learning*) qui est une application de l'intelligence artificielle. L'avantage de cette technique par rapport à une méthode d'appariement standard est qu'elle permet de sélectionner de manière automatique les variables pertinentes à la fois pour effectuer l'appariement entre participants et non participants, et pour tenir compte de l'hétérogénéité de l'effet. Un ingrédient essentiel de cette méthode est que la sélection des variables pertinentes et l'estimation des effets hétérogènes sont réalisés à partir de deux sous-échantillons tirés au sort parmi la population présente dans la base de données. La sélection des variables est réalisée par la minimisation d'un critère qui vise à arbitrer entre le meilleur appariement possible des individus et le plus grand degré de précision dans l'estimation des effets, tout en pénalisant les combinaisons de variables trop complexes. C'est ainsi que les chercheurs ont identifié que les immigrants récents dont la connaissance du néerlandais est limitée tirent le plus grand bénéfice des formations en terme de mise à l'emploi. Cette technique peut ensuite être utilisée pour prédire le type de formation le plus efficace selon les caractéristiques du demandeur d'emploi.

Bart Cockx est professeur d'économie à l'UGent et chercheur associé à l'IRES (LIDAM, UCLouvain).

Bart Cockx
bart.cockx@UGent.be

L'étude détaillée sur laquelle repose ce numéro de *Regards économiques* : Cockx Bart, Michael Lechner et Joost Bollens (2020). "Priority to Unemployed Immigrants? A Causal Machine Learning Evaluation of Training in Belgium." UGent, University of St Gallen et VDAB.

REGARDS ÉCONOMIQUES

Place Montesquieu, 3
B1348 Louvain-la-Neuve
regard-ires@uclouvain.be
tél. 010 47 34 26
www.regards-economiques.be



Directeur de la publication :
Vincent Bodart

Rédactrice en chef :
Muriel Dejemeppe

Secrétariat & logistique :
Virginie Leblanc

Graphiste :
Dominique Snyers

Comité de rédaction :

Paul Belleflamme
Vincent Bodart
Muriel Dejemeppe
Gilles Grandjean
Jean Hindriks
William Parienté
Sandy Tubeuf
Bruno Van der Linden
Frédéric Vrins

