

Distance entre métiers et transitions professionnelles des demandeurs d'emploi après une formation

Skill Distance Between Occupations and Post-Training Professional Transitions of Jobseekers

Kevin Michael Frick*, Yagan Hazard, Damien Mayaux*** et Thomas Zuber******

Résumé – La formation professionnelle contribue-t-elle à la résorption des déséquilibres structurels du marché du travail ? Nous construisons une nouvelle mesure d'écart de compétences entre métiers grâce à l'application d'un algorithme de traitement du langage à un échantillon d'offres d'emploi. En utilisant cette mesure, nous montrons que le différentiel de retour à l'emploi entre demandeurs d'emploi formés et non formés est tiré par les réallocations vers des métiers éloignés en termes de compétences par rapport à celui du dernier poste occupé. D'un point de vue purement réallocatif, le différentiel de retour à l'emploi lié à la formation professionnelle ne semble cependant pas être tiré par des redirections plus nombreuses vers des métiers en forte tension.

Abstract – Does vocational training help correct structural imbalances in the labour market? We propose a new measure of the skills distance between occupations, obtained by fine-tuning a large language model on a sample of job offers. Using this method, we demonstrate that the “return to employment” differential between jobseekers with and without training is driven by a reallocation of workers towards occupations that are very different from their previous posts in terms of the skills required. From a purely reallocative perspective, however, the return to employment differential associated with vocational training does not appear to be driven by more jobseekers moving to occupations where employers are struggling to recruit.

JEL : J62, J68, J24

Mots-clés : déséquilibre structurel, formation, compétences

Keywords: structural imbalance, training, skills

* Toulouse School of Economics ; ** Collegio Carlo Alberto et ESOMAS ; *** Paris School of Economics ; **** Banque de France.
Correspondance : thomas.zuber@banque-france.fr

Les opinions exprimées dans cette étude ne reflètent ni les positions de la Banque de France ni celles de l'Eurosystème. Cette recherche a bénéficié de la participation financière de la DARES, dans le cadre de l'appel à projets de recherche « La formation des personnes en recherche d'emploi – Vague 3 » qu'elle a lancé ; d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre du programme d'Investissements d'avenir portant la référence ANR-17-EURE-001 ; et d'un accès aux moyens de calcul de l'IDRIS au travers de l'allocation de ressources 2022-AD011013484R1 attribuée par GENCI. Nous tenons à remercier Cyril Verluse pour ses conseils avisés lors de la genèse de ce projet, Pauline Givord et Anne Bucher pour leurs nombreuses remarques et leur soutien dans l'accès aux données, ainsi que l'ensemble des participants aux différentes réunions de co-pilotage.

Reçu en mars 2024, accepté en septembre 2025.

Les jugements et opinions exprimés par les auteurs n'engagent qu'eux-mêmes, et non les institutions auxquelles ils appartiennent, ni a fortiori l'Insee.

Citation: Frick, K. M., Hazard, Y., Mayaux, D. & Zuber, T. (2025). Skill Distance Between Occupations and Post-Training Professional Transitions of Jobseekers. *Economie et Statistique / Economics and Statistics*, 547, 49–67. doi: 10.24187/ecostat.2025.547.2138

Les déséquilibres structurels d'offre et de demande sur le marché du travail contribuent à augmenter aussi bien le niveau que la persistance du taux de chômage. Parmi ces déséquilibres, l'hypothèse du *mismatch*, c'est-à-dire de l'inadéquation entre les compétences détenues par les travailleurs et celles demandées par les entreprises, est souvent mise en avant. Cette inadéquation a pour conséquence la coexistence de poches résiduelles de chômage et de marchés tendus dans lesquels les entreprises cherchant à recruter se heurtent à la rareté de l'offre de travail qualifié. Si en théorie ces déséquilibres pourraient se résorber grâce à une réorientation de l'offre de travail vers les marchés en tension, en pratique de nombreux facteurs empêchent ces mobilités professionnelles d'avoir lieu.

Les programmes de formation professionnelle ciblant les personnes en recherche d'emploi visent à apporter des éléments de réponse à ce problème. D'abord en facilitant l'accès à la formation des personnes en recherche d'emploi, puis en orientant autant que possible l'offre de formation existante vers les métiers en tension pour lesquels les difficultés de recrutement sont les plus importantes. La combinaison de ces deux leviers devrait permettre d'exploiter à leur maximum les gains que l'on peut attendre de la formation professionnelle sur le retour à l'emploi. En France, le Plan d'investissement dans les compétences (PIC) mis en place depuis 2018 s'inscrit dans cette logique.

Les effets directs de la formation sur le retour à l'emploi sont mieux connus que ceux, indirects, sur l'équilibre du marché du travail via les transitions professionnelles. De nombreuses études documentent l'effet de la formation sur le retour à l'emploi (voir par exemple la méta-analyse de Card *et al.*, 2018). En revanche, seules quelques rares études considèrent les effets indirects que la formation professionnelle peut avoir sur le marché du travail en contribuant à réduire les déséquilibres structurels d'offre et de demande de travail existant entre différents métiers (Şahin *et al.*, 2014 ; Barnichon & Figura, 2015 ; Marinescu & Rathelot, 2018). Une importante difficulté consiste à quantifier à quel point un facteur donné, comme la mobilité géographique ou les écarts de compétences entre métiers, est responsable des déséquilibres entre métiers. L'enjeu est pourtant de taille : en 2021 en France, ces déséquilibres structurels pourraient avoir été responsables d'environ 15 % du niveau de chômage observé (Fontaine & Rathelot, 2022).

Cet article cherche à déterminer si, et dans quelle mesure, les personnes en recherche d'emploi ayant suivi une formation professionnelle ont réalisé des transitions professionnelles (*i*) associées à un changement de compétences plus important que celui qu'elles auraient connu en l'absence de formation, et (*ii*) orientées vers des marchés du travail en plus forte tension que ceux dans lesquels elles auraient exercé en l'absence de formation. Pour ce faire, nous construisons une mesure de distance en compétences inter-métiers en utilisant le contenu textuel des offres d'emploi publiées par Pôle emploi. Nous développons à cette fin une nouvelle approche basée sur l'entraînement d'un réseau de neurones à partir de tâches dites « prétextes »¹ permettant d'isoler le contenu propre aux compétences dans le corpus des textes des offres d'emploi, de placer chaque offre et chaque métier dans un espace de compétences de dimension réduite et de définir une notion de distance inter-métiers comme l'angle entre deux métiers donnés dans cet espace. Les compétences qui forment les dimensions de l'espace ne sont pas interprétables individuellement, mais il est par construction possible de comparer le vecteur des compétences pour deux métiers donnés. Grâce à cette mesure de distance entre métiers, nous sommes capables de distinguer les transitions professionnelles qui relient entre eux des métiers dont les compétences sont relativement proches, et des transitions professionnelles dirigées vers des métiers dont les compétences s'écartent beaucoup plus nettement des compétences utilisées dans le métier d'origine du demandeur d'emploi. Dawson *et al.* (2021) utilisent une méthodologie similaire à la nôtre².

L'appariement de données administratives réalisé par le dispositif Formation, Chômage et Emploi (ForCE) de la Dares nous permet ensuite de suivre les trajectoires des personnes en recherche d'emploi ayant eu recours à la formation professionnelle entre 2018 et 2020. Notre étude se concentre sur un sous-échantillon de demandeurs d'emploi ayant exercé un emploi stable au cours de l'année écoulée pour lesquels les notions de métier d'origine et donc de transition professionnelle éventuelle sont bien définies. La comparaison des trajectoires professionnelles réalisées par les demandeurs d'emploi formés et non formés, associée à la présence d'un grand nombre de variables de contrôle présentes dans les données administratives, permet d'étudier

1. Une tâche prétexte est une tâche utilisée non car elle constitue l'objectif final de l'entraînement de l'algorithme mais parce qu'elle permet de contraindre celui-ci à acquérir certaines propriétés que l'on juge désirables.

2. Voir également Bana *et al.* (2019) et Gentzkow *et al.* (2019).

la relation entre formation et retour à l'emploi, ainsi que la distance inter-métiers parcourue et les différentiels de tension entre métiers associés aux transitions professionnelles permises par la formation. Cette comparaison est réalisée grâce à une méthode de *machine learning* avec orthogonalisation en deux étapes (Chernozhukov *et al.*, 2018) permettant de corriger les différences observables entre le groupe traité et le groupe de contrôle³. Le résultat de cette comparaison n'est interprétable comme un effet causal de la formation professionnelle que sous l'hypothèse d'indépendance conditionnelle (CIA) du recours à la formation professionnelle – ce qui a peu de chance d'être le cas en pratique. Malgré cet écueil bien connu de la littérature sur l'évaluation des politiques de formation, nos résultats permettent d'apporter un premier éclairage sur la question peu étudiée du lien entre la formation professionnelle et la résorption des déséquilibres structurels du marché du travail.

Notre étude se concentre en effet sur la question plus spécifique de l'impact de la formation sur les transitions professionnelles des demandeurs d'emploi et certaines caractéristiques de l'emploi retrouvé. Nous montrons ainsi que la relation entre formation et retour à l'emploi passe presque entièrement par un retour à l'emploi vers des métiers éloignés des métiers d'origine des personnes formées. Du point de vue de la réallocation des emplois, ces transitions professionnelles plus nombreuses des personnes formées ne semblent cependant pas être systématiquement dirigées vers des métiers en plus forte tension que les métiers d'origine. Ceci laisse penser qu'un meilleur ciblage des métiers en tension par l'offre de formation professionnelle à destination des demandeurs d'emploi permettrait d'améliorer la réallocation des emplois retrouvés à la suite d'une formation. Notre méthodologie impose de nous restreindre aux demandeurs d'emploi ayant connu un épisode d'emploi stable au cours des douze derniers mois. Cette sous-population est plus jeune, plus diplômée, et plus proche de l'emploi que la moyenne des demandeurs d'emploi inscrits à Pôle emploi. En plus de la validité incertaine attachée à l'hypothèse d'indépendance conditionnelle, cette restriction nécessaire de notre échantillon circonscrit également le champ d'interprétation de nos résultats.

Nous contribuons à une abondante littérature qui montre que les transitions entre métiers sont associées à un fort coût en capital humain spécifique (Becker, 1964). Shaw (1984) est le premier à mettre en évidence le rôle spécifique joué par la trajectoire professionnelle accomplie entre

differents métiers, plutôt qu'entre secteurs, dans la détermination du revenu d'un travailleur. Dans le contexte du système d'apprentissage allemand, Eckardt (2022) montre que les personnes travaillant dans un domaine pour lequel elles n'ont pas été spécifiquement formées accusent une perte de revenu par rapport à leurs collègues bénéficiant d'une formation adéquate. Ces coûts sont positivement corrélés à une mesure de distance en compétences inter-métiers. Les coûts liés aux transitions professionnelles forcées ont par ailleurs été étudiés dans le cadre de l'évaluation des effets de la concurrence internationale sur les marchés du travail locaux. Traberman (2019) pour le Danemark et Basco *et al.* (2025) pour la France montrent que les changements de métiers causés par une forte concurrence internationale aboutissent à des pertes de revenus d'autant plus importantes que le nouveau métier du travailleur est éloigné de son métier d'origine. De plus, Hyman (2018) montre que les politiques de formation permettent aux travailleurs touchés de réaliser des transitions professionnelles vers des métiers moins exposés à la concurrence internationale. Enfin, reprenant l'intuition de Shaw (1984), une littérature abondante montre que la prise en compte du caractère multidimensionnel des compétences (par opposition à un simple indice linéaire) est cruciale pour la compréhension des transitions sur le marché du travail, du processus d'appariement entre travailleurs et entreprises et de la détermination du niveau de rémunération des salariés (Gathmann & Schönberg, 2010 ; Lindenlaub & Postel-Vinay, 2021 ; Guvenen *et al.*, 2020 ; Baley *et al.*, 2022).

Le reste de l'article est structuré de la manière suivante. Dans un premier temps, nous construisons une nouvelle mesure de distance inter-métiers à partir des données textuelles contenues dans les offres postées à Pôle emploi et la validons quantitativement et qualitativement. Dans un deuxième temps, nous présentons la base FORCE et les choix méthodologiques adoptés dans la sélection de notre échantillon d'étude et des différentes variables d'intérêt. Enfin, nous présentons les résultats issus d'une comparaison des trajectoires professionnelles réalisées par les demandeurs d'emploi ayant recours à la formation et celles des inscrits à Pôle emploi n'y ayant pas eu recours, avant de conclure.

3. Nous vérifions que l'utilisation de la méthode des appariements sur score de propension (Rosenbaum & Rubin, 1983) donne des résultats en tous points comparables.

1. Une nouvelle mesure de distance inter-métiers

Notre analyse de la relation entre formation et retour à l'emploi repose sur une nouvelle mesure de distance inter-métiers, qui traduit les obstacles liés aux différences de compétences entre métiers rencontrés lors d'une reconversion professionnelle.

Il existe déjà de nombreuses sources de données de qualité reliant métiers et compétences professionnelles. Aux États-Unis, le système O*NET fournit un répertoire détaillé des compétences, et pour chaque métier de la nomenclature américaine (*la Standard Occupational Classification*), un indicateur de la fréquence d'utilisation et du niveau de maîtrise attendu dans chacune des compétences. Il contient aussi une table des métiers connexes (*la Related Occupation Matrix*) qui donne, pour chaque métier de départ, la liste ordonnée des dix métiers vers lesquels il est le plus facile de se reconvertis eu égard aux compétences attendues. En France, la classification du Répertoire Opérationnel des Métiers et des Emplois (ROME V3) joue un rôle analogue : en plus de la définition de 532 métiers et 14 secteurs d'activité, elle fournit un registre de compétences qu'elle fait correspondre à la nomenclature des métiers et suggère une liste de métiers de reconversion possibles en partant d'un métier donné (rubrique Mobilités du ROME). Le ROME permet ainsi d'approcher la notion de distance en compétences inter-métiers soit en étudiant les intersections existant entre les compétences recensées pour chaque métier, soit en s'appuyant sur l'ensemble des trajectoires professionnelles suggérées par les transitions vers des métiers proches recensés dans la rubrique Mobilités.

Toutefois, l'usage du ROME V3 pour étudier l'effet de la formation professionnelle présente plusieurs limites. Tout d'abord, la liste des mobilités professionnelles suggérées ne vise pas à l'exhaustivité, ce qui fait qu'elle ne mentionne pas certaines transitions tout à fait possibles et pertinentes, et ne dit rien des transitions difficiles à réaliser. De plus, elle classe ces suggestions de mobilité professionnelle en deux catégories seulement, alors qu'on souhaiterait plutôt mesurer de manière continue la difficulté à se reconvertis d'un métier à un autre, en particulier pour des métiers assez lointains en termes de compétences. Enfin, les compétences identifiées dans le ROME sont souvent spécifiques à un ou deux métiers seulement, ce qui rend impossible la distinction entre des paires de métiers assez différents et des paires de métiers complètement

differents. Ces trois points sont autant d'obstacles à l'étude quantitative de l'effet de la formation sur les mobilités professionnelles, puisqu'on peut supposer que cette dernière permet justement des mobilités plus lointaines en termes de compétences.

Dans cette partie, nous construisons une mesure de distance inter-métiers en entraînant un réseau de neurones à partir de textes d'offres d'emploi. L'accès à un large corpus de données textuelles permet de produire une mesure continue et informative, même pour des paires de métiers éloignés. Après avoir détaillé les données utilisées, nous présentons la méthodologie de construction de la mesure et les différents exercices de validation.

1.1. Données

Le principal jeu de données utilisé est le texte de plus de 4 millions d'offres postées sur le site de Pôle emploi entre décembre 2018 et octobre 2020 ainsi que les codes métier ROME associés. Nous utilisons le texte de ces offres pour pré-entraîner notre modèle de langage, puis pour entraîner notre réseau de neurones à prédire le code métier à partir du texte de l'offre. Nous avons rééquilibré le nombre d'offres de chaque métier par rapport à l'échantillon initial afin de ne pas introduire de biais lors de l'entraînement. Nous présentons des statistiques descriptives sur le jeux de données d'offres d'emploi dans l'Annexe en ligne S1 (lien de l'Annexe en ligne à la fin de l'article).

Un de nos objectifs était de proposer une alternative aux métriques basées sur le ROME V3. Nous avons donc utilisé deux champs du ROME afin de produire des mesures de distance inter-métiers alternatives pour tester notre propre mesure :

- Le champ Mobilités associe à des couples de métiers la valeur 1 si une mobilité inter-métiers semble possible sans formation et 2 si une mobilité inter-métiers semble possible avec une formation minime des compétences sous-jacentes. Nous utilisons notamment ce champ pour produire une distance inter-métiers alternative, appelée *Graph distance*, qui est la distance sur le graphe orienté dont les sommets sont les codes ROME, dont la présence d'une arête indique qu'une mobilité est suggérée dans le champ Mobilités et dont la pondération de chaque arête correspond au niveau 1 ou 2 défini précédemment. Nous avons également utilisé le champ Mobilités dans la construction de notre mesure de distance inter-métiers.

- Le champ Compétences associe une liste d'items de compétences, générales ou spécifiques, à chacun des 532 métiers de la classification

ROME V3. Nous utilisons ce champ pour produire une distance inter-métiers alternative, appelée *Skill distance*, qui est le cosinus de l'angle entre les vecteurs représentant les métiers dans l'espace, dont la composante i vaut 1 si la compétence i est associée au métier et 0 sinon. Cela correspond à l'analogue de notre mesure de distance, mais à partir d'une représentation dans un espace de dimension égale au nombre total de compétences listées dans ROME et qui est établie principalement à dire d'expert.

En plus de ces champs issus de la classification du ROME, nous avons intégré à l'entraînement du réseau de neurones des informations sur le marché du travail français. Pour faire en sorte que notre mesure de distance inter-métiers reflète uniquement des écarts en matière de compétences requises, et non d'autres facteurs affectant les comportements des agents sur le marché du travail comme les stéréotypes de genre associés à des métiers, le prestige social de certaines professions ou encore les différenciels de tension entre métiers, nous nous sommes limités à des données qui nous semblaient clairement interprétables en termes de compétences professionnelles, en particulier le niveau moyen de formation des travailleurs dans un métier donné⁴, ainsi qu'une liste de couples de métiers de départ et d'arrivée pour lesquels nous avons observé des transitions fréquentes au sein d'une même entreprise, associées à des hausses de salaire, que nous avons interprétées comme des mobilités professionnelles verticales⁵.

1.2. Le réseau de neurones

La construction de la distance inter-métiers se fait en trois temps : nous nous appuyons tout d'abord sur un modèle de langage pour extraire une représentation sémantique riche du contenu des offres d'emploi, puis sur un réseau de neurones pour ne retenir de cette représentation sémantique que le contenu pertinent par rapport aux compétences et positionner les métiers dans un espace de grande dimension, et enfin nous utilisons la distance angulaire dans cet espace comme mesure de distance inter-métiers.

Notre contribution méthodologique principale est la construction d'une représentation spatiale en grande dimension des compétences mobilisées et exigées par les 532 métiers de la classification ROME V3 qui soit pertinente quantitativement pour analyser les mobilités professionnelles tout en restant géométriquement interprétable. La mesure de distance inter-métiers utilisée dans la suite de cet article est un sous-produit immédiat de cette construction. Le modèle de langage utilisé pour analyser le texte des offres d'emploi,

FlauBERT (Le *et al.*, 2020), est une référence pour la langue française et nous nous en servons d'une manière tout à fait standard dans la littérature en traitement automatique du langage.

Le choix d'utiliser un réseau de neurones pour obtenir notre représentation nous distingue de la littérature existante sur la multidimensionnalité des compétences. Celle-ci repose majoritairement sur des méthodes de construction d'indice ou de réduction de dimension (analyse en composante principale, analyse de correspondance), qui sont simples à définir mathématiquement, mais produisent une représentation des compétences dont la pertinence par rapport à la finalité souhaitée n'a rien d'évident – en particulier pour un usage dans des analyses quantitatives. Au contraire, nous avons fait le choix d'une approche supervisée dans laquelle les tâches prétextes permettant l'entraînement de la représentation sont facilement explicables, en lien étroit avec la finalité de la représentation et imposent une cohérence géométrique au résultat. Ce parti pris permet une meilleure interprétabilité de la représentation obtenue, au prix d'une complexité accrue dans sa construction.

Le fait que la construction de la représentation spatiale des compétences s'appuie sur des données textuelles fait courir certains risques. Il est possible que les textes des offres d'emploi contiennent des éléments sans lien avec les compétences, et ayant donc un faible pouvoir prédictif sur le métier, comme le style de rédaction de l'offre, le nom de l'entreprise ou le lieu d'exercice de l'activité, qui pourraient affecter la représentation. On peut également craindre que, selon les métiers, les offres aient tendance à demander plus de compétences que le métier n'en exige vraiment, ou à demander des compétences différentes selon la phase du cycle du marché du travail, comme observé dans (Deming & Kahn, 2018). La taille du corpus d'offres d'emploi et la faible profondeur du réseau de neurones permettent cependant d'espérer que ce dernier exploitera surtout les éléments du texte ayant un fort pouvoir prédictif sur le métier et parviendra à faire abstraction du bruit dans la description des compétences.

1.2.1. Objectifs

Nous utilisons un réseau de neurones afin de placer les 532 métiers de la classification ROME dans un espace de dimension 20 de telle sorte que le positionnement relatif de deux métiers

4. Pour ce faire nous utilisons le niveau de formation renseigné dans le Fichier historique de Pôle emploi pour les inscrits en 2018. Cette information est tirée de l'appariement ForCE.

5. Ces transitions verticales sont construites sur la base de données administratives de l'année 2019 (Base tous salariés, fichier « Postes », Insee).

nous renseigne sur la possibilité d'effectuer une transition professionnelle de l'un vers l'autre. Les dimensions de cet espace n'ont pas vocation à être considérées individuellement, comme si elles mesuraient l'importance d'une compétence spécifique (ou même facilement interprétable) pour un métier donné. En revanche, le vecteur formé de ces 20 dimensions contient une représentation synthétique des compétences requises par le métier et peut, quant à lui, être interprété. Une fois les métiers positionnés dans l'espace, nous utilisons la distance cosinus entre vecteurs comme mesure de la distance inter-métiers en termes de compétences.

Formellement, on appellera représentation R la fonction qui à chaque métier de la classification ROME associe le vecteur correspondant dans cet espace.

$$\begin{aligned} R : \{ \text{Codes ROME} \} &\rightarrow \mathbb{R}^{20} \\ x &\mapsto R(x) \end{aligned}$$

La représentation est construite de sorte que la géométrie dans cet espace ait une interprétation claire en matière de compétences et de mobilité professionnelle.

(i) L'angle entre les vecteurs $R(x)$ et $R(y)$ doit refléter à quel point les compétences mobilisées dans les métiers x et y sont similaires. Le cosinus de cet angle est notre mesure de distance inter-métiers.

(ii) La norme $\|R(x)\|$ doit refléter le niveau de maîtrise attendu dans les compétences mobilisées dans le métier x .

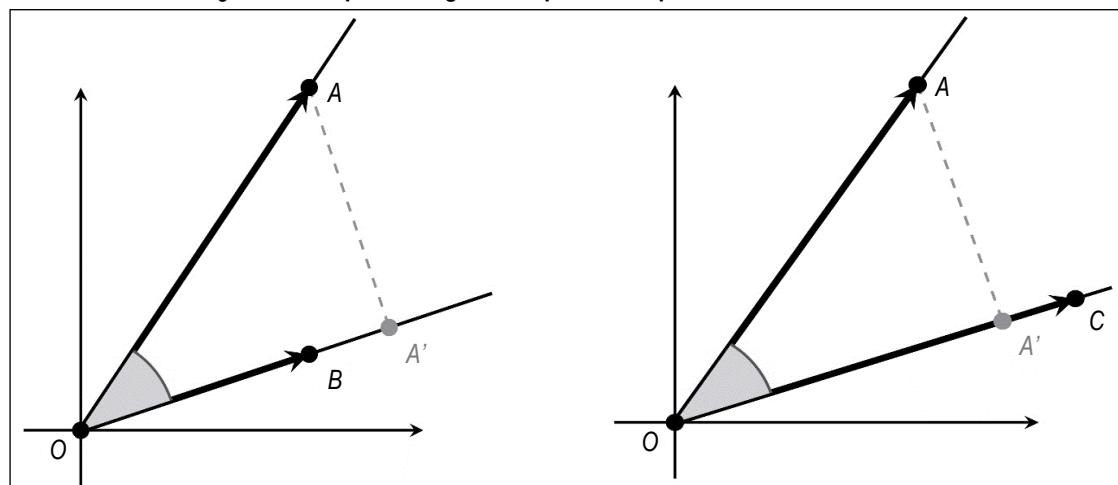
(iii) La projection du vecteur u sur la droite engendrée par $R(x)$ doit refléter le niveau de maîtrise dans les compétences mobilisées par le métier x d'une personne dont les compétences détenues sont représentées par u .

(iv) Toutes les composantes du vecteur $R(x)$ sont positives et représentent une certaine dimension des compétences mobilisées par le métier x .

La distinction entre quelles sont les compétences mobilisées (la direction du vecteur) et quel est le niveau de maîtrise dans les compétences mobilisées (la norme du vecteur) s'inspire des questions posées dans plusieurs jeux de données existants reliant les compétences et les métiers, comme ceux de la classification O*NET aux États-Unis. En pratique, nous utilisons la direction des vecteurs dans la représentation des métiers (voir infra), mais leur norme joue également un rôle dans le processus de construction de cette représentation.

La figure I illustre l'interprétation géométrique de notre représentation. Une conséquence de la troisième propriété, qui porte sur le transfert de compétences lors d'une transition professionnelle, est que la dimension de l'espace majore le cardinal du plus grand ensemble de métiers dont les compétences mobilisées n'ont deux à deux absolument rien en commun. Autrement dit, en passant de l'un de ces métiers à un autre, il faut absolument tout réapprendre de zéro. Nous avons choisi de travailler en dimension 20, ce qui impose à la représentation de capter des compétences partagées entre les 532 métiers de

Figure I – Interprétation géométrique de la représentation R des métiers



Note : représentation du passage d'un métier de départ A à un métier d'arrivée B dans l'espace des compétences. Notre mesure de distance inter-métiers, qui traduit l'écart entre les compétences mobilisées par les métiers A et B, correspond au cosinus de l'angle grisé. Le métier A nécessite un niveau de maîtrise des compétences qu'il mobilise plus élevé que le métier B, d'où une plus grande distance à l'origine. Si un individu dont les compétences sont exactement celles correspondant au métier A venait à se reconvertisir dans le métier B, son niveau de maîtrise des compétences mobilisées par B serait la distance à l'origine du point A'. Ici, le niveau de maîtrise est suffisant pour envisager une mobilité de A à B. À l'inverse, il n'est pas possible pour ce même individu de se reconvertisir dans le métier C, qui mobilise les mêmes compétences que le métier B mais nécessite un niveau de maîtrise plus élevé.

la classification ROME, tout en laissant plus de flexibilité pour représenter ces compétences que dans un modèle de très faible dimension.

Le cadre géométrique proposé possède deux propriétés qui peuvent paraître contre-intuitives. D'une part, la représentation introduit une forme de substituabilité imparfaite entre les compétences. Avoir un niveau de maîtrise très important dans un domaine donné assure un niveau minimum dans tout autre domaine ayant au moins une compétence en commun. En ce sens, la représentation gère mal les impossibilités absolues de reconversion. D'autre part, la relation binaire sur les métiers qui indique qu'une transition professionnelle est possible n'est en général pas transitive, autrement dit, il peut être possible d'aller d'un métier A à un métier B et d'un métier B à un métier C sans pouvoir aller directement de A à C. L'idée sous-jacente est que cette relation binaire capture la possibilité d'une mobilité professionnelle en un temps raisonnable sans avoir à suivre une formation spécifique. Or, la somme de deux temps raisonnables n'est pas forcément raisonnable.

Cette traduction géométrique de la faisabilité des transitions professionnelles dans l'espace des compétences n'est pas utilisée directement dans la construction de la mesure. Cependant, elle illustre le mécanisme de la réutilisation partielle des compétences déjà acquises lors d'une mobilité professionnelle. Plus on s'éloigne du métier initial en termes de compétences mobilisées, c'est-à-dire plus l'angle est grand, plus la norme du projeté est petite, c'est-à-dire moins on peut réutiliser ses compétences déjà acquises. Ainsi, pour un niveau de maîtrise des compétences donné dans le métier d'arrivée, il est plus faisable de se reconvertir dans un métier avec lequel on a une faible distance angulaire. Cela explique l'intérêt d'utiliser la distance angulaire comme mesure de la difficulté à se reconvertir.

En plus de cette représentation des métiers, notre réseau de neurones permet également d'associer à tout texte d'offre d'emploi une position dans ce même espace de dimension 20. On notera F la fonction associée.

$$\begin{aligned} F : \{ \text{Textes d'offres d'emploi} \} &\rightarrow \mathbb{R}^{20} \\ x &\mapsto F(x) \end{aligned}$$

Cette représentation des textes d'offre d'emploi a deux intérêts :

- Elle est utilisée pour apprendre la représentation R des métiers. La principale tâche d'entraînement du réseau de neurones consiste à comparer le texte t d'une offre d'emploi à un métier x , via leurs représentants $F(t)$ et $R(x)$, afin de prédire le code ROME du métier de l'offre d'emploi.

- Une fois l'entraînement terminé, elle peut être utilisée pour analyser qualitativement la représentation obtenue. On peut modifier le texte t d'une offre d'emploi en entrée et observer comment cela affecte son représentant $F(t)$.

1.2.2. L'architecture du réseau de neurones

Le réseau de neurones est composé de trois blocs, comme illustré sur la figure II.

Le bloc 1 est un modèle de langage. Il prend en entrée un texte et produit une représentation de ce texte en 768 dimensions qui permet de répondre à une large classe de problèmes sémantiques. Nous utilisons une version pré-entraînée de ce modèle et ses paramètres sont figés pendant l'entraînement des autres blocs.

Le bloc 2 est un réseau de neurones à trois couches. Il prend en entrée la représentation d'une offre d'emploi en sortie du bloc 1 et produit une représentation de cette offre en 20 dimensions. Cette étape permet de ne conserver des 768 dimensions initiales que l'information pertinente en termes de compétences et de mobilité professionnelle. Associé avec le bloc 1, il forme la fonction F qui associe au texte d'une offre d'emploi sa représentation en dimension 20.

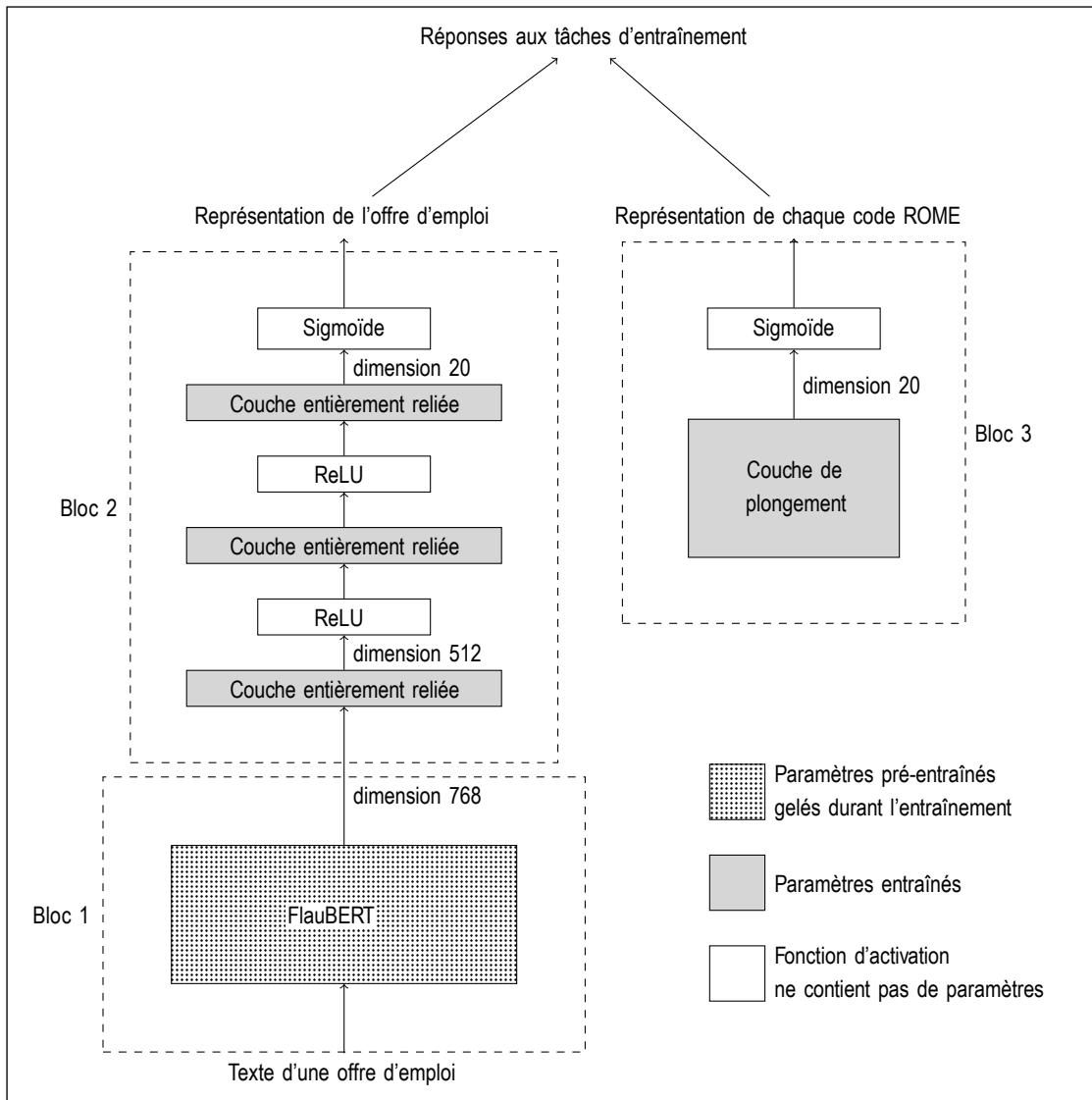
Le bloc 3 est une couche supplémentaire qui stocke la représentation R des 532 métiers de la classification ROME en 20 dimensions. La comparaison de la sortie du bloc 2 avec la sortie du bloc 3 permet d'évaluer la performance du réseau aux différentes tâches et d'entraîner ses paramètres en conséquence.

Les paramètres du bloc 1 (fond en pointillé sur la figure II) sont récupérés sur un modèle déjà entraîné (FlauBERT. Le et al., 2020), puis légèrement ré-entraînés de manière semi-supervisée sur notre corpus d'offres d'emploi séparément du reste du modèle, avant d'être gelés durant l'entraînement des blocs 2 et 3. Les paramètres des blocs 2 et 3 (couches en gris foncé sur la figure II) sont initialisés aléatoirement, puis entraînés grâce au texte des offres d'emploi tel que détaillé dans la sous-section suivante.

1.2.3. Tâches d'entraînement et pénalités associées

La représentation R des métiers fait partie des paramètres du réseau de neurones, qui évoluent à mesure que le réseau de neurones s'entraîne à réaliser certaines tâches. C'est le choix de tâches d'entraînement pertinentes et de pénalités associées qui permet d'imposer à la représentation les propriétés géométriques énoncées précédemment.

Figure II – Schéma du réseau de neurones utilisé



Nous utilisons les couples tâche/pénalité suivants pour l'entraînement du réseau de neurones :

Prédire le code ROME d'une offre d'emploi - *WARP loss*

Nous avons adapté la pénalité (*Weighted Approximate-Rank Pairwise*, WARP) de l'article Weston *et al.* (2011) à notre problème. Pour chaque texte t d'une offre d'emploi, nous calculons l'angle de son représentant $F(t)$ avec le représentant $R(x)$ de chaque métier x de la classification ROME. On s'attend à ce que l'angle entre $F(t)$ et le représentant $R(x_0)$ du métier correspondant à l'offre soit très faible. On pénalise donc le réseau pour chaque métier x telle que l'angle entre $F(t)$ et $R(x)$ est inférieur à celui entre $F(t)$ et $R(x_0)$.

$$l_{\text{WARP}} = \sum_{\text{ROME codes } x \neq x_0} |F(t) \cdot (R(x) - R(x_0))|_+$$

où $|\cdot|_+ = \max(0, \cdot)$.

Prédire les mobilités suggérées dans la classification ROME - *Triplet loss*

La *triplet loss* vient à l'origine de la reconnaissance d'image. Si l'on a trois photos de visage et que, contrairement à la troisième, les deux premières viennent d'une même personne, alors un algorithme de reconnaissance faciale devrait prédire que la première ressemble plus à la deuxième qu'à la troisième.

De la même manière, nous formons des triplets de métiers en prenant un métier de départ x_0 , un métier x vers lequel une transition depuis le métier de départ est suggérée dans le champ Mobilités de la classification ROME, et un autre y vers lequel aucune transition n'est suggérée. Nous prenons ensuite une offre au hasard dans chacun de ces métiers.

On s'attend à ce que la mobilité de x_0 à x soit un meilleur choix que celle de x_0 vers y . En particulier, la transition doit être faisable en termes de compétences sans que le niveau de maîtrise attendu n'ait trop diminué par rapport à l'emploi initial. Cela conduit à utiliser la métrique suivante pour dire à quel point une mobilité est susceptible d'être recommandée.

$$d(x_0, x) = \max \left(1 - \frac{\|R(x)\|}{\|R(x_0)\|}, 1 - \frac{R(x_0) \cdot R(x)}{\|R(x)\|^2} \right)$$

On définit alors la *triplet loss* comme l'écart, par rapport à cette métrique, entre les couples (x_0, x) et (x_0, y) .

$$l_{\text{Triplet}} = |d(x_0, x) - d(x_0, y)|_+$$

Prédire le niveau de maîtrise attendu dans les compétences mobilisées par le métier x - Norm loss

Pour chaque métier x , on compare la norme de son représentant $R(x)$ à une valeur e_x (normalisée entre 0 et 1) qui reflète le niveau de qualification moyen de travailleurs occupants ce métier⁶.

$$l_{\text{Norm}} = \sum_{\text{ROME codes } x} (\|R(x)\| - e_x)^2$$

Prédire le niveau de maîtrise - Vertical loss

Dans de nombreux secteurs, les mobilités verticales nécessitent d'avoir acquis les compétences des échelons inférieurs. On s'attend donc à ce que, parmi les couples de métiers de départ et d'arrivée souvent associés dans les données du marché du travail à des mobilités verticales, le premier nécessite moins de compétences que le second, dans toutes les dimensions que recouvre la notion de compétences.

En notant $x \prec y$ le fait que beaucoup de mobilités verticales impliquent de passer du métier x au métier y , on définit la *vertical loss* par

$$l_{\text{Vertical}} = \sum_{x \prec y} \sum_{i=1}^{20} |R(x)_i - R(y)_i|_+$$

Les différentes pénalités associées à ces tâches sont normalisées, passées au logarithme puis sommées pour former la pénalité totale l , soit :

$$l = \log(1 + l_{\text{WARP}}) + \log(1 + l_{\text{Triplet}}) + \log(1 + l_{\text{Norm}}) + \log(1 + l_{\text{Vertical}})$$

1.3. Résultats

Nous avons validé qualitativement et quantitativement la mesure de distance obtenue à partir de la représentation spatiale des métiers. Nous détaillons dans l'annexe des premiers éléments sur la performance de l'algorithme sur les tâches prétextes ainsi qu'un exercice de validation qualitative.

Notre principale démarche de validation quantitative de la mesure obtenue a consisté à évaluer sa capacité à prédire les transitions professionnelles effectivement réalisées sur le marché du travail.

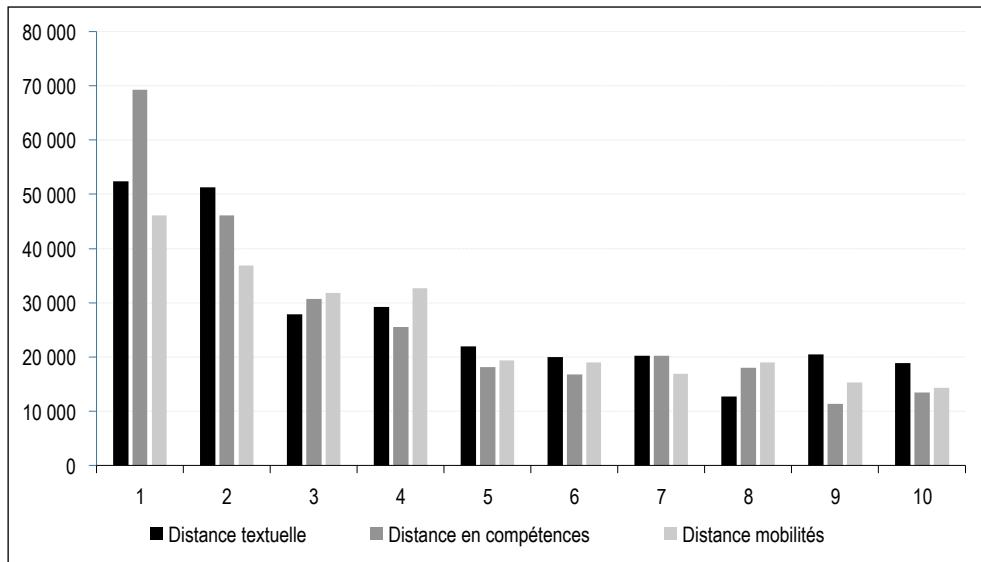
Prédire parfaitement les transitions observées n'est pas la vocation de la mesure, car ces dernières peuvent dépendre de nombreux facteurs qui n'ont rien à voir avec les compétences professionnelles, tels que les aspirations personnelles, les stéréotypes de genre ou encore le contexte local du marché du travail. Cependant, il semble raisonnable de supposer qu'il y ait en général plus de transitions professionnelles entre des métiers plus proches en termes de compétences mobilisées.

Nous avons comparé le pouvoir prédictif de notre mesure de distance inter-métiers par rapport à deux alternatives : d'une part, la distance sur le graphe des mobilités professionnelles suggérées dans la classification ROME, d'autre part, la distance cosinus entre les vecteurs de compétences associés au métier de départ et d'arrivée dans la classification ROME.

La figure III représente, pour chacune des trois mesures de distance inter-métiers et pour chaque rang k de 1 à 10, le nombre de transitions professionnelles observées qui vont d'un métier de départ vers le k -ième métier le plus proche selon la mesure. Cela nous permet d'évaluer la capacité à retrouver les transitions les plus probables et d'observer la qualité de la mesure pour des métiers plus éloignés. On observe que notre mesure de distance (appelée Distance textuelle dans le graphique) fait mieux que la distance construite à partir des suggestions de la classification ROME (Distance mobilités) pour 7 des 10 rangs. D'ailleurs, si l'on cumule les transitions expliquées par les k métiers d'arrivée les plus proches, notre mesure la surpassé pour tout k entre 1 et 10. Concernant la mesure fondée sur les compétences définies dans la classification ROME V3 (Distance en compétences), on remarque qu'elle a un très bon pouvoir prédictif pour les métiers très proches, mais un faible pouvoir prédictif pour les métiers plus lointains. Cela s'explique simplement par le fait que les notions de compétences utilisées dans la classification ROME V3 sont assez étroites et manquent de pouvoir de généralisation (ce constat a d'ailleurs orienté les travaux sur la refonte de la classification ROME vers sa version 4.0). À l'inverse, notre mesure construite à partir de données textuelles diverses et non

6. En pratique ce niveau est construit comme le niveau d'études moyen observé dans le métier, normalisé entre 0 et 1.

Figure III – Nombre de transitions expliquées par les mesures de distance inter-métiers



Note : cet histogramme représente le nombre de transitions réalisées sur le marché du travail en 2019 vers les 10 métiers d'arrivée les plus proches du métier d'origine selon différentes mesures de distance inter-métiers. Nous comparons la distance textuelle que nous construisons aux distances construites grâce aux rubriques Mobilités et Compétences du ROME V3. Les transitions utilisées dans ce graphique ne distinguent pas les transitions réalisées avec ou sans formation préalable.

Source : Base tous salariés, fichier « Postes », 2019, Insee ; calcul des auteurs.

structurées semble plus à même d'identifier des notions larges de compétence et de faire des généralisations puisque c'est elle la plus performante pour des rangs élevés.

Le tableau 1 donne les résultats de la régression, sur l'ensemble des couples de métiers pour lesquels au moins une transition a eu lieu, du logarithme du nombre de transitions réalisées par rapport à différentes mesures de distance inter-métiers normalisées. Cela nous permet de comparer la qualité de la mesure à prédire les transitions à longue distance, c'est-à-dire même pour des métiers de départ et d'arrivée qui peuvent être assez éloignés en termes de compétences, par opposition à la figure précédente qui se concentrat sur les transitions les

plus probables. La part de la variance expliquée par notre mesure de distance inter-métiers est plus grande que pour les deux autres mesures et le coefficient de corrélation est plus élevé. Il n'est pas surprenant que les mesures de distance inter-métiers centrées sur les compétences n'expliquent qu'une faible part des transitions observées, dans la mesure où elles ne prennent pas en compte les nombreux autres facteurs qui peuvent conduire à une mobilité professionnelle.

2. Formation professionnelle, retour à l'emploi et trajectoires professionnelles

Cette section étudie le devenir des demandeurs d'emploi primo-formés en termes de retour à l'emploi et de trajectoires professionnelles en

Tableau 1 – Pouvoir explicatif des mesures de distance inter-métiers sur les transitions réalisées

	log(transitions)	log(transitions)	log(transitions)	log(transitions)
Distance mobilités	-0,409 (0,0052)			-0,288 (0,0049)
Distance en compétences		-0,4501 (0,0051)		-0,287 (0,0049)
Distance textuelle			-0,609 (0,0052)	-0,495 (0,0052)
N	149 209	149 209	149 209	149 209
R ²	0,0409	0,0506	0,0935	0,1388

Note : cette table donne les résultats de quatre régressions dont la variable dépendante est le logarithme du nombre de transitions entre métiers. Les deux premières variables explicatives sont des mesures de distance construites à partir des rubriques Mobilités et Compétences du ROME V3, que nous comparons à la distance textuelle construite pour les besoins de cette étude. Chaque mesure de distance inter-métiers a été normalisée de sorte que le coefficient donne directement la corrélation entre variables expliquée et explicative.

Source : Base tous salariés, fichier « Postes », 2019, Insee ; calcul des auteurs.

général, en les comparant à des demandeurs d'emploi jamais formés mais comparables du point de vue des caractéristiques observables. Nous décrivons d'abord les données administratives exploitées et leur traitement préalable, puis nous expliquons brièvement la stratégie empirique adoptée, avant de présenter les résultats obtenus.

2.1. Données et construction de l'échantillon

Nous utilisons les données du dispositif Formation, Chômage, Emploi (FORCE)⁷ qui regroupe notamment :

- le Fichier historique des demandeurs d'emploi (FH), qui rassemble pour chaque demandeur d'emploi des informations sur ses épisodes d'inscription à Pôle emploi sur les 10 années précédant le millésime de FORCE considéré ;
- la base régionalisée des stagiaires de la formation professionnelle (Brest), qui rassemble pour chaque demandeur d'emploi ayant été stagiaire de la formation professionnelle les caractéristiques des formations suivies depuis 2017 ;
- la base Mouvements de main-d'œuvre (MMO, dont les données sont issues de la Déclaration sociale nominative), qui rassemble depuis 2017 les informations sur les contrats de travail pour chaque salarié du privé⁸, pour l'ensemble des demandeurs d'emploi présents dans le FH du millésime de FORCE considéré.

Nous construisons notre base d'analyse en suivant les mêmes principes généraux que ceux adoptés par Chabaud *et al.* (2022) dans leur étude sur la relation entre formation professionnelle et retour à l'emploi des demandeurs d'emploi et mobilisant des données similaires. Nous créons d'abord pour chaque mois entre janvier 2018 et décembre 2020 une base contenant l'ensemble des demandeurs d'emploi inscrits durant ce mois m (hors catégorie E / catégorie administrative 5)⁹, après avoir préalablement abouté les épisodes administratifs de demande d'emploi séparés de moins de 30 jours. Afin de concentrer l'analyse sur les primo-formés, nous utilisons la table P2 du Fichier historique pour exclure les demandeurs d'emploi qui ont suivi une formation avant 2017. Cet échantillon est apparié à la base Brest afin de déterminer l'ensemble des demandeurs d'emploi entrant en formation pour la première fois au cours du mois m , qui constitue le groupe de traitement pour ce mois-ci – tout en excluant du groupe de contrôle l'ensemble des individus ayant suivi une formation avant le mois m . Nous excluons du champ des formations considérées celles qui sont directement associées à un projet de recrutement (POEI,

POEC, AFPR) – afin de ne pas considérer ces formations mécaniquement associées au retour à l'emploi dans le champ d'étude¹⁰. Enfin, l'appariement avec la base MMO permet de récupérer des informations sur les contrats détenus par les demandeurs d'emploi (*i*) préalablement au mois m et (*ii*) pour chacun des 24 mois suivant le mois m . Dans la mesure où cette étude se concentre sur les trajectoires professionnelles accomplies par les demandeurs d'emploi, notre échantillon principal restreint la population des demandeurs d'emploi aux personnes ayant détenu un emploi stable au cours des 12 mois précédant le mois m . Nous identifions ainsi pour chaque demandeur d'emploi un métier de référence qui, lorsque nous l'associons aux caractéristiques de l'emploi retrouvé après le mois m , nous permet de construire les principales variables d'intérêt de cette étude (distance inter-métiers parcourue, différentiel de tension, nature du contrat retrouvé, etc.).

Nous conservons un grand nombre d'informations sur les individus en provenance du FH, telles que leur âge, sexe, situation familiale, nombre d'enfants, niveau de formation, qualification, situation de handicap, habitation dans une zone rurale/urbaine, volonté de mobilité géographique, catégorie d'inscription, motif d'inscription, nationalité, métier recherché, salaire de réserve, ancienneté au chômage au mois m , temps passé au chômage et nombre d'épisodes de chômage avant l'épisode en question. Ces variables de contrôle nous permettent de redresser les différences observables existant entre les demandeurs d'emploi ayant recours à la formation professionnelle et ceux n'y ayant pas recours.

2.2. Stratégie empirique

Définissons une variable $D_i \in \{0,1\}$ indiquant si l'individu i est entré en formation durant le mois m ($D_i = 1$) ou non ($D_i = 0$). Pour une variable d'intérêt Y_i (par exemple, le retour à l'emploi dans les 24 mois suivant le mois m d'entrée potentielle en formation), on peut définir deux valeurs potentielles (selon le langage usuel de l'inférence causale) désignant la valeur prise par Y_i pour l'individu i dans le cas où i n'est pas

7. Millésime 2023T2.

8. Depuis 2022, les informations sur les contrats en cours d'une grande partie des salariés de la fonction publique sont également collectées.

9. La catégorie E / catégorie administrative 5 contient l'ensemble des demandeurs d'emploi inscrits et ayant un emploi à temps plein, n'ayant de ce fait aucune obligation de recherche d'emploi.

10. La base Brest permet de distinguer les Préparations Opérationnelles à l'Emploi Individuelles (POEI), les Préparations Opérationnelles à l'Emploi Collectives (POEC) et les Actions de Formation Préalables au Recrutement (AFPR) des Aides Individuelles de Formation (AIF) sur lesquelles nous nous concentrerons.

entré en formation durant le mois m ($Y_i(0)$) ou dans le cas où l'entrée en formation a eu lieu ($Y_i(1)$). Intuitivement, l'hypothèse d'indépendance conditionnelle suppose qu'un individu traité ($D_i=1$) aurait eu, en l'absence d'entrée en formation, un devenir similaire à un individu i' contrôlé ($D_{i'}=0$) ayant des caractéristiques observables X similaires à i . Formellement,

$$Y_i(0) \perp D_i | X.$$

Cette hypothèse, si elle est vérifiée, permet d'identifier l'*average treatment effect on the treated* (ATT), c'est-à-dire l'effet moyen de l'entrée en formation au mois m parmi les individus étant effectivement entrés. Il est important de garder à l'esprit, dans l'interprétation de nos résultats, que le traitement étudié ici est bien l'entrée en (première) formation *durant le mois m*, par opposition à tout autre scénario (incluant l'absence totale d'entrée en formation, *ou* une entrée postérieure au mois m).

Notons cependant que dans le cas présent, il est probable que les demandeurs d'emploi choisissant d'entrer en formation diffèrent des individus non formés selon des caractéristiques inobservables (motivation, aspiration à des changements de métier, etc.) susceptibles d'être corrélées significativement avec le retour à l'emploi et la trajectoire professionnelle en général. Ce genre de biais – empêchant l'interprétation causale d'analyses se basant sur l'hypothèse d'indépendance conditionnelle et ne contrôlant que pour les différences de caractéristiques observables – a été abondamment commenté dans la littérature (Lalonde, 1986). Il est donc plus prudent d'interpréter nos résultats comme des mesures de la corrélation existant entre formation professionnelle et retour à l'emploi, corrigées des différences de caractéristiques observables.

Concernant l'estimation, il existe plusieurs manières de contrôler pour des différences associées aux observables X , en particulier quand la dimensionnalité de X est élevée. La première consiste à avoir recours à du *matching* sur score de propension. Ces méthodes sont bien connues comme solutions au problème de dimensionnalité de X depuis Rosenbaum & Rubin (1983). La méthode que nous privilégions, dite de *Double Debiased Machine Learning* (DML) (Chernozhukov *et al.*, 2018), repose sur l'utilisation d'estimateurs non paramétriques de l'espérance conditionnelle de la variable d'intérêt et du score de propension qui sont ensuite combinés pour former un estimateur robuste à une mauvaise spécification de l'un des

deux termes¹¹. Bien que présentant de meilleures propriétés statistiques que l'estimateur classique d'appariement sur score de propension, le caractère biaisé ou non de l'estimateur DML est tout autant dépendant de la validité de l'hypothèse d'indépendance conditionnelle.

2.3. Caractéristiques de l'échantillon étudié

Cette sous-partie décrit l'échantillon de demandeurs d'emploi mobilisé lors de nos analyses. Comme nous l'avons déjà souligné, notre population d'intérêt se restreint à des personnes pour lesquelles nous sommes en mesure de définir un métier de référence et qui sont, par conséquent, relativement proches de l'emploi. Comme indiqué ci-dessus, nous excluons par ailleurs du champ de l'étude les demandeurs d'emploi ayant recours à des formations directement associées à un projet de recrutement (POEI, POEC et AFPR).

Le tableau 2 compare les demandeurs d'emploi pour lesquels un contrat de travail stable (CDI ou CDD de plus de 6 mois) terminé il y a moins de 12 mois est observé (notre population d'intérêt), à ceux pour lesquels ce n'est pas le cas. La première sous-population est la seule pour laquelle nous pouvons étudier l'association entre formation et changement de compétences ainsi que le différentiel de tension entre métier d'origine et métier d'arrivée. Cette sous-population apparaît plus jeune, avec une proportion plus faible de femmes, davantage de diplômés du supérieur ; elle est inscrite depuis moins longtemps au chômage et se forme davantage.

Le tableau 3 présente les caractéristiques des demandeurs d'emploi composant notre population d'intérêt selon le type de formation suivie. Nous distinguons les demandeurs d'emploi n'entretenant aucune formation des demandeurs d'emploi entretenant une formation de plus de 30 heures selon le type (certifiante ou non) et la durée de celle-ci (supérieure ou inférieure à 420 heures, soit 3 mois à temps complet)¹².

Comparée aux demandeurs d'emploi n'entretenant pas une formation de plus de 30 heures, la sous-population des demandeurs d'emploi débutant une première formation de plus de 30 heures est en moyenne plus jeune, au chômage depuis moins longtemps, et contient une proportion plus importante de femmes et de diplômés du supérieur. Ces observations sont toutes encore

11. Concernant l'exécution concrète de cette méthode, des packages sont disponibles sous Python et R (langage utilisé pour cette étude) à l'adresse suivante : <https://docs.doubleml.org/stable/index.html>.

12. Il existe une saisonnalité importante dans le nombre d'entrées en formation parmi les demandeurs d'emploi. Les mois de septembre correspondent à un pic du nombre d'entrées. Nous choisissons de présenter les statistiques pour les inscrits entre janvier 2018 et décembre 2020.

Tableau 2 – Statistiques descriptives selon l’observation (ou non) d’un contrat stable dans les 12 mois précédent le mois d’étude

	Sans emploi stable dans l’année écoulée	Avec un emploi stable dans l’année écoulée
Femme (%)	53,4	48,9
Âge (%)		
Moins de 25 ans	10,2	12,4
De 25 à 50 ans	60,1	64,5
Plus de 50 ans	25,1	18,0
Niveau d’études (%)		
Inférieur au baccalauréat	50,6	43,2
Baccalauréat	22,1	22,6
Supérieur au baccalauréat	27,3	34,3
Ancienneté d’inscription (mois)	31,7	12,6
Formation (%)	5,2	8,8
Formation certifiante (%)	2,1	4,0
Formation > 420 heures (%)	2,4	3,7
Observations (milliers)	4 118	908

Source et champ : données ForCE, Dares. Ensemble des demandeurs d’emploi inscrits entre janvier 2018 et décembre 2020 à l’exclusion des demandeurs d’emploi ayant recours à des formations directement associées à un projet de recrutement (POEI, POEC et AFPR).

Tableau 3 – Statistiques descriptives selon le statut de formation pour la population des demandeurs d’emploi avec un emploi stable au cours de l’année écoulée

	Sans formation	Formation	Formation certifiante	Formation > 420 heures
Femme (%)	48,9	49,8	51,3	55,7
Âge (%)				
Moins de 25 ans	12,4	12,4	13,3	18,6
De 25 à 50 ans	64,4	69,3	71,6	69,3
Plus de 50 ans	18,1	14,0	10,7	7,7
Niveau d’études (%)				
Inférieur au baccalauréat	43,3	38,5	33,7	32,7
Baccalauréat	22,5	26,5	29,2	31,9
Supérieur au baccalauréat	34,2	35,0	37,1	35,4
Durée d’inscription (mois)	12,7	8,3	8,3	8,4
Changement de métier désiré (%)	67,5	69,2	69,9	70,8
Distance si changement	0,51	0,53	0,53	0,54
Observations (milliers)	900	8,0	3,7	3,4

Source et champ : données ForCE, Dares. Ensemble des demandeurs d’emploi inscrits entre janvier 2018 et décembre 2020 pour lesquels un contrat de travail (CDI ou CDD de plus de 6 mois) terminé il y a moins de 12 mois est observé, à l’exclusion des demandeurs d’emploi ayant recours à des formations directement associées à un projet de recrutement (POEI, POEC et AFPR).

plus prononcées dans les sous-populations des individus débutant une formation certifiante ou une formation longue de plus de 420 heures. Par ailleurs, un fait surprenant est l’absence de différence notable d’aspiration à changer de métier entre ces différentes populations. Nous mesurons cette aspiration comme la part des demandeurs d’emploi déclarant rechercher un métier différent du dernier métier exercé de

manière stable¹³. On observe que presque les trois quarts des individus indiquent chercher un métier différent de celui précédemment exercé – cette proportion n’évoluant pas ou très

13. Nous retenons dans cet exercice un métier exprimé dans la version la plus détaillée de la classification FAP, soit 225 postes. Un passage à un niveau moins granulaire, par exemple la FAP 87, ne change pas le résultat. Le passage à la classification FAP permet de faire concorder le ROME utilisé par le FH et la PCS utilisée par les MMO.

peu quel que soit le statut de traitement. De plus, l'ampleur des changements en compétences envisagés – tels que mesurés par la distance en compétence entre le métier recherché et le métier précédemment exercé – est comparable entre individus débutant ou non une formation, quel que soit le type de celle-ci.

Les stratégies d'identification et d'estimation associées reposent sur la correction des différences observables entre demandeurs d'emploi contrôles et traités grâce à la méthode DML (voir supra). En pratique nous incluons parmi les variables de contrôle utilisées par l'algorithme le niveau de qualification, l'âge, le sexe, le type de contrat recherché, le niveau de formation, l'expérience sur le marché du travail, la résidence dans un quartier prioritaire de la ville, la situation matrimoniale, la catégorie administrative d'enregistrement à Pôle emploi, la nationalité, la résidence dans une zone urbaine sensible, le dernier motif d'inscription à Pôle emploi, le caractère obligatoire ou non des démarches de recherche d'emploi, le niveau du salaire de réserve déclaré, les préférences de mobilité, le métier recherché, la zone d'emploi de résidence, le nombre d'épisodes de chômage distincts ainsi que le temps total d'inscription à Pôle emploi au cours des dix dernières années.

2.4. Résultats

Tous les résultats présentés dans cette partie sont obtenus grâce la technique du *Double Debiased Machine Learning* (voir supra) et sont largement comparables à ceux obtenus en appliquant une méthode plus classique d'appariement sur score de propension.

Le tableau 4 synthétise nos résultats sur la relation entre la formation professionnelle et le retour à l'emploi à différents horizons temporels (3, 6, 12 et 24 mois après l'entrée en formation), pour l'ensemble des demandeurs d'emploi primo-formés entre janvier 2018 et juin 2020 pour lesquels nous observons un emploi stable au cours de l'année précédant leur entrée en formation. On distingue ici la relation entre l'ensemble des formations et le retour à l'emploi de la relation entre les seules formations certifiantes et le retour à l'emploi en général puis, dans un second temps, en restreignant la variable d'intérêt au retour à l'emploi dans un emploi stable (CDI ou CDD de plus de 6 mois). Comme largement documenté dans la littérature (Card *et al.*, 2018), nous retrouvons un phénomène dit d'« enfermement » (*lock-in*), c'est-à-dire une relation négative entre l'entrée en formation et le retour à l'emploi à court terme liée au temps

consacré à la formation. La relation entre la formation et le retour à l'emploi est positive 12 mois après l'entrée en formation et reste positive mais devient plus forte 24 mois après l'entrée en formation. Nos résultats, qui portent sur une sous-population relativement proche de l'emploi, diffèrent tant qualitativement que quantitativement de ceux rapportés par Chabaud *et al.* (2022). D'un point de vue qualitatif, l'effet d'enfermement apparaît plus prononcé et plus durable sur notre population d'intérêt qu'il ne l'est pour l'ensemble des inscrits à Pôle emploi. D'un point de vue quantitatif, le différentiel redressé formé/non formé à 24 mois pour notre population d'intérêt est environ 20 % plus faible que le résultat que Chabaud *et al.* (2022) obtiennent sur l'ensemble de la population des inscrits à Pôle emploi. Ces différences peuvent s'expliquer par le fait que les individus de notre échantillon sont par construction plus proches de l'emploi ainsi que par le type de formations retenues. Ainsi, une entrée en formation est davantage susceptible de réduire significativement les opportunités de retour à l'emploi pour cette sous-population, pour laquelle de telles opportunités sont nombreuses, même en l'absence de formation.

Le tableau 4 montre par ailleurs que si l'effet initial d'enfermement est plus prononcé pour les formations certifiantes qui sont très souvent des formations longues, le différentiel formé/non formé à 24 mois de ce type de formations est plus élevé que le différentiel mesuré pour les formations prises dans leur ensemble. Cette conclusion demeure globalement inchangée lorsque l'on s'intéresse au retour à un emploi stable (CDI ou CDD de plus de 6 mois) : cependant le différentiel formé/non formé à 24 mois suite à une formation certifiante n'est pas significativement différent de celui pour l'ensemble des formations. L'effet d'enfermement est par ailleurs plus durable, encore significatif 12 mois après la formation.

Le tableau 5 résume nos principaux résultats sur la relation entre la formation et les trajectoires professionnelles des demandeurs d'emploi pour différents types de formation (toutes formations confondues et formations certifiantes) 24 mois après l'entrée en formation. Comme précédemment, nous présentons l'ensemble de ces résultats sur l'échantillon des demandeurs d'emploi pour lesquels nous observons un contrat stable au cours de l'année écoulée. Dans ce tableau, la variable dépendante de retour à l'emploi à 24 mois est décomposée en fonction de la distance entre l'emploi retrouvé et l'emploi d'origine. On distingue ainsi un

Tableau 4 – Les différences de retour à l'emploi observées entre les personnes formées et les personnes n'ayant pas recours à la formation

	(1) 3 mois	(2) 6 mois	(3) 12 mois	(4) 24 mois
Tout type d'emploi				
Toute formation	-0,100 (0,002)	-0,089 (0,002)	0,007 (0,002)	0,069 (0,002)
Formation certifiante	-0,119 (0,002)	-0,119 (0,003)	0,018 (0,003)	0,086 (0,003)
Emploi stable				
Toute formation	-0,063 (0,001)	0,06 (0,002)	-0,011 (0,002)	0,033 (0,002)
Formation certifiante	-0,078 (0,002)	-0,082 (0,002)	-0,013 (0,003)	0,040 (0,003)

Note : ce tableau présente les résultats de régressions séparées de 4 variables dépendantes différentes (en colonne, correspondant au retour à l'emploi pour différents horizons temporels) sur 2 variables explicatives (en ligne, correspondant aux différents types de formation considérés). La partie haute du tableau considère le retour à tout type d'emploi à différents horizons temporels, tandis que la partie basse répète les analyses en ne considérant comme variable dépendante que le retour à l'emploi stable (CDI, CDD de plus de 6 mois). Les erreurs standard clusterisées au niveau métier × zone d'emploi sont présentées entre parenthèses.

retour à l'emploi dans le métier d'origine ($d=1$), dans un métier très proche du métier d'origine ($d \in [2 ; 5]$), dans un métier proche du métier d'origine ($d \in [6 ; 20]$), et finalement dans un métier dont les compétences sont éloignées de celles du métier d'origine ($d > 20$). En procédant ainsi on constate que la formation professionnelle réduit la probabilité de retour à l'emploi dans le métier d'origine des demandeurs d'emploi, est presque neutre pour les métiers proches, et augmente de manière substantielle la probabilité de retour à l'emploi dans les métiers dont les compétences sont éloignées de celles du métier d'origine. Ces conclusions sont encore plus prononcées lorsque l'on restreint les formations considérées aux seules formations certifiantes. Ainsi, le recours à la formation en général et le recours à des formations certifiantes en particulier semblent bien coïncider avec une réallocation de la main-d'œuvre vers des métiers que l'on peut considérer comme éloignés en termes de compétences des métiers exercés

avant la formation. Ce résultat est frappant dans la mesure où il implique que la totalité de la relation entre le retour à l'emploi et la formation professionnelle passe par des transitions professionnelles dirigées vers des métiers éloignés du métier d'origine des demandeurs d'emploi.

Le surcroît d'emploi observé chez les personnes ayant recours à la formation professionnelle est-il associé à une redirection plus fréquente vers les marchés du travail en forte tension ? Pour répondre à cette question, le tableau 6 décompose la relation entre la formation et le retour à l'emploi selon que l'emploi retrouvé se situe dans un marché en plus faible ou en plus forte tension que le marché de l'emploi d'origine. Les données de tension utilisées pour réaliser cette décomposition sont obtenues en régressant le retour à l'emploi des demandeurs d'emploi n'ayant pas recours à la formation sur des effets fixes marché (zone d'emploi × FAP) en contrôlant pour les caractéristiques individuelles

Tableau 5 – Les différences de retour à l'emploi après 24 mois entre personnes formées et non formées selon la distance entre le métier d'origine et le métier d'arrivée

	(1) Métier d'origine ($d = 1$)	(2) Métier très proche ($d \in [2 ; 5]$)	(3) Métier proche ($d \in [6 ; 20]$)	(4) Métier éloigné ($d > 20$)
Toute formation	-0,044 (0,001)	0,005 (0,001)	0,006 (0,001)	0,080 (0,002)
Formation certifiante	-0,052 (0,002)	0,009 (0,002)	0,005 (0,002)	0,095 (0,003)

Note : ce tableau présente les résultats de régressions séparées de 4 variables dépendantes différentes (en colonne) sur 2 variables explicatives (en ligne, correspondant aux différents types de formation considérés). À titre d'exemple, les coefficients de la colonne (2), Métier très proche ($d \in [2 ; 5]$), correspondent au différentiel formé/non formé de retour à l'emploi (à 24 mois) dans l'un des 4 métiers les plus proches du métier précédemment exercé, selon notre mesure de distance inter-métiers. Les erreurs standard clusterisées au niveau métier × zone d'emploi sont présentées entre parenthèses.

Tableau 6 – Les différences de retour à l’emploi après 24 mois entre personnes formées et non formées selon le niveau relatif de tension du métier d’arrivée

	(1) plus faible tension que le métier d’origine	(2) plus forte tension que le métier d’origine
Toute formation	0,046 (0,002)	0,045 (0,002)
Formation certifiante	0,050 (0,003)	0,059 (0,003)

Note : ce tableau reprend les analyses du tableau 4 en distinguant, dans la construction des variables dépendantes, le retour à l’emploi dans un marché en plus faible (ou plus forte) tension que le métier précédemment exercé par le demandeur d’emploi. Les erreurs standard clusterisées au niveau métier × zone d’emploi sont présentées entre parenthèses.

des demandeurs d’emploi présents sur chaque marché. Cette manière de procéder nous permet de nous appuyer sur un indicateur de tension (taux de retour à l’emploi par marché) qui évite les problèmes de mesure habituels liés au caractère non observable de l’effort de recherche des demandeurs d’emploi et des entreprises¹⁴. Le tableau 6 montre qu’aussi bien dans le cas des formations certifiantes que pour l’ensemble des formations, le différentiel formé/non formé à 24 mois sur le retour à l’emploi ne semble pas être porté par des redirections plus nombreuses vers des marchés de l’emploi en plus forte tension que le marché d’origine de chaque demandeur d’emploi. Dans l’ensemble, l’entrée en formation (certifiante ou non) semble augmenter de manière comparable la probabilité d’une transition vers un métier en plus forte ou en plus faible tension. Dans la perspective d’un usage du système de formation visant à réduire les déséquilibres d’offre et de demande de travail entre différents marchés du travail, ce résultat peut apparaître comme relativement décevant. Il suggère qu’une plus grande attention portée au ciblage de l’offre de formation permettrait de rediriger plus encore la main-d’œuvre vers les marchés les plus en tension, quand à ce jour les transitions générées par le passage en formation apparaissent comme relativement neutres de ce point de vue.

* *
*

La formation professionnelle permet-elle de réduire les déséquilibres structurels du marché

du travail ? Cette étude vise à apporter des éléments de réponse à cette question en comparant les trajectoires professionnelles réalisées par les demandeurs d’emploi ayant ou non recours à la formation professionnelle. À cette fin, nous avons construit une mesure originale de distance en compétences entre métiers en nous appuyant sur les textes des offres d’emploi postées à Pôle emploi. Nous avons utilisé cette mesure de distance pour étudier les transitions professionnelles des demandeurs d’emploi ayant ou non recours à la formation. Sur un échantillon de demandeurs d’emploi relativement proches de l’emploi, nos résultats montrent que, par rapport aux demandeurs d’emploi n’ayant pas recours à la formation professionnelle, les demandeurs d’emploi formés accomplissent des transitions professionnelles plus lointaines dans l’espace de compétences. Nos résultats reposent sur une hypothèse – forte dans ce contexte – d’indépendance conditionnelle du recours à la formation et doivent donc être interprétés avec prudence. D’un point de vue purement réallocatif, le surcroît de retour à l’emploi des personnes accédant à la formation professionnelle ne semble pas être tiré par des redirections plus nombreuses vers des métiers en forte tension. Ce résultat laisse présumer qu’un ciblage plus systématique de l’offre de formation professionnelle sur les compétences demandées par les métiers en tension permettrait d’accroître l’effet réallocatif de la formation professionnelle. □

14. La Dares utilise une mesure comparable (taux d’écoulement de la main-d’œuvre par marché) lorsqu’elle construit un indicateur synthétique de tension sur le marché du travail.

Lien vers l’Annexe en ligne :

www.insee.fr/fr/statistiques/fichier/8678968/ES547_Frick-et-al_Annexe-en-ligne.pdf

BIBLIOGRAPHIE

- Baley, I., Figueiredo, A. & Ulbricht, R. (2022).** Mismatch Cycles. *Journal of Political Economy*, 130(6). <https://doi.org/10.1086/720461>
- Bana, S. H., Brynjolfsson, E., Rock, D. & Steffen, S. (2019).** Job2Vec: Job Title Benchmarking with Collective Multi-View Representation Learning. In Proceedings of the 28th ACM International Conference on Information and Knowledge Management (CIKM '19). Association for Computing Machinery, New York, NY, USA, 2763–2771. <https://doi.org/10.1145/3357384.3357825>
- Barnichon, R. & Figura, A. (2015).** Labor Market Heterogeneity and the Aggregate Matching Function. *American Economic Journal: Macroeconomics*, 7(4), 222–249. <https://doi.org/10.1257/mac.20140116>
- Basco, S., Liégey, M., Mestieri, M. & Smagghue, G. (2024).** The effect of import competition across occupations. *Journal of International Economics*, 153, 104001. <https://doi.org/10.1016/j.inteco.2024.104001>
- Becker, G. S. (1964).** *Human capital: A theoretical and empirical analysis with special reference to education*. New York & London: Columbia University Press for the National Bureau of Economic Research. <https://www.nber.org/books-and-chapters/human-capital-theoretical-and-empirical-analysis-special-reference-education-first-edition>
- Card, D., Kluve, J. & Weber, A. (2018).** What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, 16(3), 894–931. <https://doi.org/10.1093/jeea/jvx028>
- Chabaud, M., Bucher, A., Givord, P. & Louvet, A. (2022).** Quelles sont les chances de retour à l'emploi après une formation ? DARES, *Document d'études* N° 261. <https://dares.travail-emploi.gouv.fr/publication/quelles-sont-les-chances-de-retour-lemploi-apres-une-formation>
- Chernozhukov, V., Chetverikov, D., Demirer, M., Duflo, E., Hansen, C., Newey, W. & Robins, J. (2018).** Double/debiased machine learning for treatment and structural parameters. *The Econometrics Journal*, 21(1), C1–C68. <https://doi.org/10.1111/ectj.12097>
- Dawson, N., Williams, M. A. & Rizoiu, M. A. (2021).** Skill-driven recommendations for job transition pathways. *PLoS ONE*, 16(8), e0254722. <https://doi.org/10.1371/journal.pone.0254722>
- Deming, D. & Kahn, L. B. (2018).** Skill Requirements across Firms and Labor Markets: Evidence from Job Postings for Professionals. *Journal of Labor Economics*, 36(S1), S337–S369. <https://doi.org/10.1086/694106>
- Eckardt, D. (2022).** Are Chemists Good Bankers? Returns to the Match between Training and Occupation. *Working Paper*. https://conference.iza.org/conference_files/TAM_2022/eckardt_d29661.pdf
- Fontaine, F. & Rathelot, R. (2022).** Le marché du travail français à l'épreuve de la crise sanitaire. *Notes du Conseil d'analyse économique*, 71(2), 1–12. <https://www.cairn.info/revue-notes-du-conseil-d-analyse-economique-2022-2-page-1.htm>
- Gathmann, C. & Schönberg, U. (2010).** How General Is Human Capital? A Task-Based Approach. *Journal of Labor Economics*, 28(1), 1–49. <https://doi.org/10.1086/649786>
- Gentzkow, M., Kelly, B. & Taddy, M. (2019).** Text as Data. *Journal of Economic Literature*, 57(3), 535–574. <https://doi.org/10.1257/jel.20181020>
- Guvenen, F., Kuruscu, B., Tanaka, S. & Wiczer, D. (2020).** Multidimensional Skill Mismatch. *American Economic Journal: Macroeconomics*, 12(1), 210–244. <https://doi.org/10.1257/mac.20160241>
- Hyman, B. G. (2018).** Can Displaced Labor Be Retrained? Evidence from Quasi-Random Assignment to Trade Adjustment Assistance. In: *Proceedings. Annual Conference on Taxation and Minutes of the Annual Meeting of the National Tax Association*, 111, 1–70.
- LaLonde, R. J. (1986).** Evaluating the Econometric Evaluations of Training Programs with Experimental Data. *The American Economic Review*, 76(4), 604–620. <http://www.jstor.org/stable/1806062>
- Le, H., Vial, L., Frej, J., Segonne, V., Coavoux, M., Lecouteux, B., Allauzen, A., ... & Schwab, D. (2020).** FlauBERT: Unsupervised Language Model Pre-training for French. In: *Proceedings of the 12th Language Resources and Evaluation Conference (LREC 2020)*, Marseille. European Language Resources Association, 2479–2485. <https://aclanthology.org/2020.lrec-1.302>
- Lindenlaub, I. & Postel-Vinay, F. (2021).** The Worker-Job Surplus. NBER, *Working Paper* 28402. <http://www.nber.org/papers/w28402>
- Marinescu, I. & Rathelot, R. (2018).** Mismatch Unemployment and the Geography of Job Search. *American Economic Journal: Macroeconomics*, 10(3), 42–70. <https://doi.org/10.1257/mac.20160312>
- Rosenbaum, P. R. & Rubin, D. B. (1983).** The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>

- Şahin, A., Song, J., Topa, G. & Violante, G. L. (2014).** Mismatch Unemployment. *American Economic Review*, 104(11), 3529–3564. <https://doi.org/10.1257/aer.104.11.3529>
- Shaw, K. L. (1984).** A formulation of the earnings function using the concept of occupational investment. *Journal of Human Resources*, 19(3), 319–340. <https://doi.org/10.2307/145876>
- Traiberman, S. (2019).** Occupations and import competition: Evidence from Denmark. *American Economic Review*, 109(12), 4260–4301. <https://doi.org/10.1257/aer.20161925>
- Weston, J., Bengio, S. & Usunier, N. (2011).** WSABIE: Scaling up to large vocabulary image annotation. In: *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence (IJCAI-11)*, 2764–2770. <https://dl.acm.org/doi/10.5555/2283696.2283856>
-

ANNEXE**AUTRES APPROCHES POUR L'ÉVALUATION DU RÉSEAU DE NEURONES****Performance sur les tâches prétextes**

Durant l'entraînement du réseau de neurones, nous surveillons l'évolution de la fonction de pénalité générale définie précédemment, ainsi que la capacité du réseau de neurones à prédire correctement le code ROME d'une offre d'emploi donnée (i.e. sa précision).

Ces deux métriques sont calculées sur l'ensemble d'entraînement et de test, et l'entraînement est arrêté pour la valeur des paramètres qui donne la plus haute précision sur le jeu de données d'entraînement – environ 80 % dans notre cas. Nous avons observé que le fait de ré-entraîner le modèle de langage de manière non supervisée sur notre corpus d'offres avant de réaliser l'entraînement de notre propre réseau de neurones augmente significativement, de l'ordre de 20 %, la précision que nous parvenons à atteindre. Cela suggère que la disponibilité de larges corpus d'offres d'emploi comme la base JOCAS peut s'avérer précieuse pour l'entraînement de modèles, même si le code ROME correspondant aux offres d'emplois est inconnu (il est aujourd'hui imputé par des techniques d'apprentissage statistique).

Nous avons observé une baisse de la pénalité générale durant l'entraînement. La décomposition selon chaque pénalité montre que, malgré la normalisation, c'est la tâche de prédiction du code ROME à partir d'une offre d'emploi qui joue le rôle le plus important dans l'entraînement. En effet, les autres pénalités décroissent assez rapidement, sans doute car les contraintes qu'elles imposent sont plus simples à satisfaire dans la géométrie de l'espace.

Validation qualitative de la mesure de distance inter-métiers

Pour l'analyse qualitative, nous nous sommes concentrés sur le métier le plus représenté parmi les demandeurs d'emploi dans chacun des 14 secteurs d'activité définis par la classification ROME. Pour chacun de ces 14 métiers types, nous avons déterminé les 5 métiers les plus proches selon notre mesure, ainsi que ceux parmi eux vers lesquels aucune mobilité n'est suggérée dans la classification ROME (voir Annexes en ligne S3 et S4). Cette deuxième liste permet d'illustrer la capacité du réseau de neurones à prédire des transitions professionnelles plausibles en dehors de celles qui lui ont été présentées durant l'entraînement. Les résultats suggèrent que, non seulement notre mesure de distance retrouve les mobilités suggérées par la classification ROME V3, mais également qu'elle parvient souvent à faire d'autres suggestions qui semblent qualitativement cohérentes. Cela conforte notre intuition initiale selon laquelle les suggestions du champ Mobilités de la classification ROME sont relativement étroites et ne permettent pas d'explorer l'ensemble des mobilités professionnelles pertinentes. On notera toutefois que notre mesure de distance est moins cohérente pour les métiers peu représentés dans notre corpus d'offres, comme Musique et chant (L1202). Dans l'Annexe en ligne S2, nous présentons également une méthode de visualisation de la représentation qui préserve la distance cosinus entre métiers au voisinage d'un métier donné tout en projetant les représentants dans un espace en deux dimensions.