

• DOCUMENT D'ÉTUDES •

AOUT 2022
N° 261

Quelles sont les chances de retour à l'emploi après une formation ?

L'apport de la base ForCe pour l'analyse des trajectoires individuelles du chômage vers l'emploi

Mathieu Chabaud
Anne Bucher
Pauline Givord
Alexandra Louvet

DARES

Quelles sont les chances de retour à l’emploi après une formation ? L’apport de la base ForCE pour l’analyse des trajectoires individuelles du chômage vers l’emploi

Mathieu Chabaud, Anne Bucher, Pauline Givord, Alexandra Louvet

Résumé

La formation des demandeurs d’emploi est un levier potentiel pour réduire le chômage de masse si celui-ci s’explique en partie par une inadéquation entre les compétences des demandeurs d’emploi et celles recherchées par les employeurs. Cependant, il n’existe en France que peu d’études documentant l’effet de suivre une formation sur le retour à l’emploi. L’exploitation de la base ForCE (Formation, Chômage, Emploi), qui permet de reconstituer de façon exhaustive la trajectoire professionnelle de l’ensemble des demandeurs d’emploi entrés en formation depuis 2017, permet d’obtenir une estimation de cet effet. Cette estimation est obtenue en appariant plus d’1,2 million de demandeurs d’emploi entrant en première formation entre 2017 et 2019 avec des demandeurs d’emploi non-formés ayant des caractéristiques similaires. La méthode utilisée est une méthode d’appariement mixte, combinant appariement sur score de propension et appariement exact. L’étude conclut à un effet estimé positif et significatif de l’entrée en première formation sur la probabilité de retrouver un emploi : la probabilité de connaître un épisode d’emploi dans les 2 ans suivant l’entrée en formation est plus de 9 points de pourcentage supérieure à celle des demandeurs d’emploi non-formés. L’écart entre formés et non-formés dépend du type de formation suivie ainsi que du type de personne y participant : l’effet estimé est plus élevé chez les publics les plus éloignés de l’emploi, notamment les demandeurs d’emploi ayant une ancienneté d’inscription de plus d’un an.

Introduction

La question de la formation des demandeurs d’emploi est devenue un enjeu de politique publique. Plusieurs plans successifs sur la dernière décennie ont montré un investissement volontariste de l’État sur ce sujet, les derniers en date étant le plan “500 000 formations supplémentaires” en 2016, et surtout le Plan d’investissement dans les compétences, lancé en 2018 et doté de près de 15 milliards d’euros sur cinq ans. Les objectifs de ces politiques sont de réduire une inadéquation des compétences (*skill mismatch*) qui contribuerait à la coexistence d’un chômage de masse et des difficultés de recrutement dans certains secteurs. Le Plan d’investissement dans les compétences affiche ainsi comme objectifs d’une part de mieux répondre aux besoins en compétences de l’économie, dans une situation de tension sur les recrutements et d’évolutions anticipées liées à la transformation numérique et à la transition écologique et, d’autre part, de développer les compétences des demandeurs d’emploi faiblement qualifiés et des jeunes sans qualification¹.

1. Avec l’objectif initial de former 1 million de demandeurs d’emploi faiblement qualifiés et 1 million de jeunes sans qualification sur cinq ans.

Ces politiques de formation à destination des moins qualifiés semblent confortées par la littérature internationale. Dans une méta analyse récente, Card *et al.* (2017) suggèrent que la plupart des analyses mettent en avant des rendements positifs de la formation, notamment pour les personnes les plus éloignées de l’emploi. La formation permet un accroissement des compétences et augmente donc les perspectives d’employabilité des formés (Crépon *et al.*, 2012, Cavaco *et al.*, 2013). Les formations sont aussi l’occasion pour les demandeurs d’emploi d’effectuer des stages en entreprise et de nouer ainsi des contacts susceptibles de les aider dans leur recherche d’emploi future (Blache, 2015). Cependant, sur le court terme, la formation peut également retarder le retour à l’emploi, dans la mesure où les demandeurs d’emploi sont moins disponibles pendant la formation pour faire des démarches actives de recherche d’emploi (on parle d’effet *lock in*). Le coût d’opportunité de la formation est cependant plus faible pendant les périodes de récession, ce qui explique sans doute que la littérature observe des rendements moins forts pendant les phases de reprise (Card *et al.*, 2017). Un arbitrage existe également sur les types de formation : Osikominu (2012) observe ainsi que les formations courtes de préparation à l’emploi permettent un retour plus rapide à l’emploi que les formations plus longues visant à l’acquisition de compétences. Toutefois, sur le long terme ces dernières assurent un effet plus durable, avec des retours moins fréquents vers le chômage et des gains salariaux plus élevés.

Il existe à ce jour peu d’études sur l’effet de la formation en France, faute de données permettant de suivre les trajectoires d’emploi notamment. Cette étude se propose d’apporter des éléments descriptifs sur les devenir des demandeurs d’emploi ayant bénéficié d’une formation professionnelle. Elle tire avantage du dispositif statistique sur la Formation, le Chômage et l’Emploi (ForCE), mis en place par la Dares dans le cadre de l’évaluation du Plan d’investissement dans les compétences. Mis à jour trimestriellement, ForCE est issu de l’appariement du Fichier Historique des demandeurs d’emploi (FH) géré par Pôle emploi, de la Base régionalisée des stagiaires de la formation professionnelle (Brest), de la base de suivi des jeunes inscrits en Missions locales (I-Milo) et du dispositif Sismmo issu des fichiers des déclarations statistiques nominatives, tous les trois gérés par la Dares². Le dispositif ForCE permet donc de suivre les trajectoires individuelles entre chômage, formation et emploi. Son exhaustivité permet de faire des analyses à des niveaux très fins, détaillées notamment au niveau des formations et des caractéristiques des personnes formées.

Cette étude utilise le dispositif ForCE pour estimer les probabilités de retour à l’emploi suite à une formation, en comparaison avec des demandeurs d’emploi de caractéristiques identiques. L’une des difficultés centrales pour évaluer la capacité des formations à améliorer les chances de retour à l’emploi tient à l’existence d’effets de sélection : le choix de suivre une formation est en général lié à des caractéristiques propres qui ont un effet en soi sur les chances de retour à l’emploi. Ces effets de sélection peuvent biaiser les estimations dans un sens *a priori* inconnu. Ainsi, si les formations sont réservées aux personnes les plus éloignées de l’emploi, l’impact de la formation risque d’être sous-estimé. A l’inverse, si ce sont les personnes les plus motivées par un projet professionnel qui suivent une formation, elles ont sans doute des chances d’insertion élevées avec ou sans formation : l’effet de la formation sera

2. L’appariement est réalisé par le Centre d’accès sécurisé à distance (CASD), voir El Bouhairi *et al.* (2022).

donc probablement sur-estimé. Recourir à des évaluations randomisées est le moyen le plus certain de s'affranchir de ces effets de sélection : il s'agit d'utiliser une affectation aléatoire des personnes qui pourront alternativement bénéficier ou ne pas bénéficier de la formation, ce qui permet d'éviter ces effets d'auto-sélection. La mise en oeuvre de ces méthodes peut être cependant complexe : pour des raisons éthiques, il n'est évidemment pas question de priver des personnes de formation, ce qui demande de recourir à des mécanismes indirects, au risque de n'évaluer l'effet de la formation que pour des catégories très spécifiques de la population³. À défaut de disposer d'un mécanisme aléatoire d'allocation aux formations, des méthodes d'appariement permettent de réduire le biais de sélection lié aux caractéristiques observables. Elles consistent à comparer les personnes formées avec des personnes non formées ayant des caractéristiques observables identiques. Disposer d'information suffisante dans la base de données est essentiel pour contrôler au maximum ce biais : Biewen *et al.* (2012) montrent que pour corriger la sélection, il est important de prendre en compte la trajectoire antérieure sur le marché du travail. Le dispositif ForCE constitue un atout important pour mesurer les effets de la formation. Il contient des informations personnelles détaillées, en particulier un historique de la trajectoire précédente au chômage (sur dix ans). Elle permet par ailleurs d'avoir des informations plus précises sur les emplois occupés. Exhaustive, elle autorise des analyses sur des sous-populations ou selon les types de formation à un niveau très détaillé, sans perte de précision. En pratique, on utilise une méthode d'appariement mixte, combinant appariement sur score de propension et appariement exact. En effet, il s'agit d'arbitrer entre deux enjeux contradictoires. D'une part, on souhaite augmenter le nombre de dimensions sur lesquelles on apparie, pour rendre plus plausible l'hypothèse d'indépendance conditionnelle. De l'autre, multiplier les dimensions avec un échantillon de taille finie expose à un risque de ne pas pouvoir trouver dans les données des personnes formées et non formées parfaitement comparables dans toutes les dimensions étudiées. Pour résoudre ce dilemme, Rosenbaum et Rubin (1983) proposent de s'appuyer sur le score de propension, c'est-à-dire la probabilité de suivre une formation en fonction de ces observables. Ils montrent en effet que conditionnellement au score de propension, les caractéristiques observables dans les deux groupes sont en moyenne identiques : appariement sur le score de propension revient donc en théorie à comparer des personnes comparables en termes de caractéristiques observables. Néanmoins, en pratique, on peut souhaiter conserver un appariement exact sur certaines dimensions. Le score de propension étant estimé sur plusieurs dimensions, il est possible de comparer deux personnes avec des scores proches, mais dont, par exemple, l'ancienneté au chômage est très différente⁴. Cela peut biaiser les estimations car l'ancienneté au chômage est particulièrement importante pour le retour à l'emploi, et peut également conditionner l'accès à la formation. C'est pour cette raison qu'une méthode mixte a été privilégiée ici, en utilisant à la fois un appariement sur le score de propension et un appariement exact sur certaines dimensions. Compte-tenu de la taille importante de l'échantillon, il est très peu probable de ne pas trouver de doublons même en utilisant plusieurs dimensions à la fois.

3. Par exemple, des mécanismes d'encouragement, qui consistent à faire un effort particulier d'information auprès de certaines personnes tirées au hasard pour les inciter à suivre une formation. Cette méthode permet d'estimer l'effet causal de la formation, mais seulement pour les personnes qui réagissent à ces incitations (les *compliers*). Par ailleurs, si les incitations ont un effet faible, les estimations seront très imprécises, et l'interprétation des résultats sera difficile.

4. D'autant plus que le score n'est pas observé, mais seulement estimé.

1 Données

Le dispositif ForCE (Formation, Chômage et Emploi) a été mis en place par la Dares en 2020. Il s’agit d’un dispositif trimestriel de croisement de données administratives individuelles sur la formation, le chômage et l’emploi. Il permet de reconstruire les trajectoires professionnelles de toutes les personnes ayant été en contact avec le Service public de l’emploi (Pôle emploi ou Missions locales) ou ayant suivi une formation professionnelle prise en charge totalement ou partiellement par les pouvoirs publics. Le dispositif ForCE apparie⁵ des données issues :

- de la base régionalisée des stagiaires de la formation professionnelle (Brest), qui comprend l’ensemble des formations des demandeurs d’emploi ayant donné lieu à une rémunération ou bénéficiant d’un régime de protection sociale ;
- du Fichier Historique (FH) des demandeurs d’emploi de Pôle emploi. Le FH enregistre l’ensemble des épisodes d’inscription pour un demandeur d’emploi sur une période de 10 ans. Il contient des caractéristiques du demandeur d’emploi (sexe, date de naissance, nationalité, situation de famille, niveau de formation atteint, diplôme obtenu, commune de résidence, habitat en quartier prioritaire de la ville,...), et des caractéristiques de l’épisode de chômage et de l’emploi demandé (date et motif d’inscription, date et motif d’annulation, code ROME, qualification, type de contrat, mobilité,...) ;
- de la base des Mouvements de Main d’œuvre (MMO), issue de la Déclaration sociale nominative (DSN) des entreprises. Cette base recense l’ensemble des contrats salariés du secteur privé ;
- de la base de suivi des jeunes inscrits en Mission locale (I-Milo).

ForCE permet donc de suivre, pour une même personne, ses transitions entre chômage, formation et emploi. Plusieurs conventions ont été posées afin de mesurer des dimensions pertinentes dans ces concepts, à partir des enregistrements administratifs fournis par ces données.

Concernant l’emploi, ne sont considérés comme des retours à l’emploi que les contrats d’au moins 31 jours. Par convention⁶, les emplois durables sont définis comme les emplois en contrat à durée indéterminée, ou avec une durée d’au moins six mois. De même, il est convenu d’agrèger les épisodes de chômage consécutifs s’ils sont éloignés de moins de 31 jours, ce qui est cohérent avec la règle retenue pour les emplois (par ailleurs, certaines désinscriptions sont parfois simplement administratives, par exemple du fait d’une non-présentation à un rendez-vous).

Concernant la formation, le choix a été fait de se concentrer sur la première formation suivie par un demandeur d’emploi. L’objectif de cette étude est en effet d’estimer les chances de retour à l’emploi des personnes ayant suivi une formation. Un même demandeur d’emploi peut suivre plusieurs formations successives : par exemple, pour certaines personnes éloignées de l’emploi, une formation de remise à niveau peut être nécessaire avant l’entrée dans une formation qualifiante. Se concentrer sur la première formation évite d’attribuer à tort le

5. L’appariement, réalisé par le Centre d’accès sécurisé à distance (CASD), se fait sur des identifiants non signifiants entre la DSN et le FH, et sur la base du nom/prénom/date de naissance pour Brest et I-Milo, voir El Bouhairi *et al.* (2022).

6. Cette convention est reprise de la convention Etat/Pôle emploi/Unédic.

bénéfice d'une formation à une autre qui aurait été suivie antérieurement. Le détail des formations est disponible dans la base Brest uniquement pour celles suivies à partir de 2017. À défaut de disposer de détails sur ces formations, le FH contient les éventuelles périodes de formations antérieures. En pratique, l'étude se concentre sur les "primo formés", c'est-à-dire les personnes qui s'inscrivent en formation à partir de 2017 (tel que mesuré dans Brest) et qui n'avaient pas suivi de formation depuis 2014 (les formations antérieures à 2014 sont peu susceptibles d'avoir encore un effet sur l'insertion dans l'emploi après 2017). À noter que les données Brest ne renseignent que sur les dates théoriques de la formation - en particulier, il n'est pas possible d'identifier les personnes qui abandonnent avant la fin. Ces personnes sont cependant très minoritaires : d'après une enquête menée par la Dares auprès des sortants de formation, environ neuf sur dix d'entre eux déclarent avoir suivi la formation jusqu'au bout (Voir Fosse et Al, 2021).

La période d'étude est restreinte de janvier 2017 à septembre 2019, ce qui permet d'avoir au moins un an de recul sur le retour à l'emploi (les données d'emploi au moment de la réalisation de cette étude n'étaient disponibles que jusqu'à septembre 2020). Le traitement considéré est le suivi d'une formation d'au moins 30 heures sur cette période⁷.

2 Statistiques descriptives

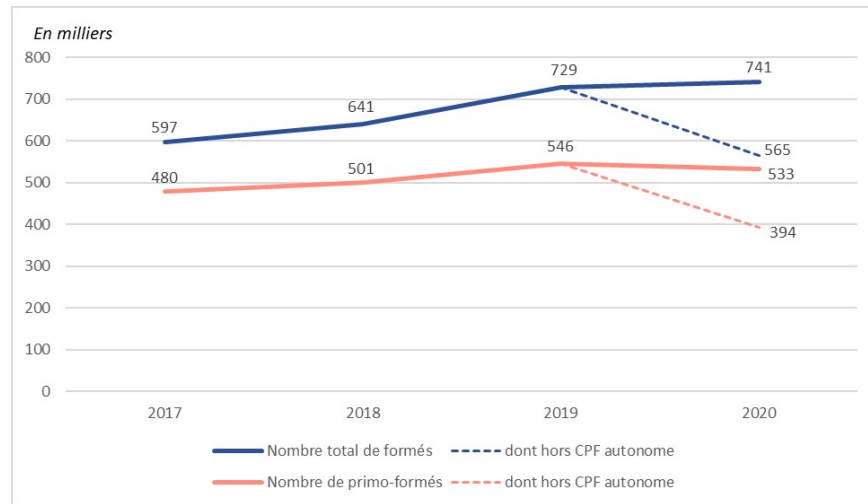
Entre 2017 et 2020, environ 2,5 millions d'inscrits à Pôle emploi sont entrés en formation. Pour 71% d'entre eux, il s'agissait de leur première formation depuis 2014. Le nombre de demandeurs d'emploi formés chaque année progresse continûment entre 2017 et 2020, passant de 597 000 à 741 000 (figure 1). Le nombre de primo-formés suit une dynamique similaire entre 2017 et 2019 mais reflue légèrement en 2020. L'introduction du Compte personnel de formation (CPF) autonome en 2020, qui permet aux demandeurs d'emploi de choisir plus facilement leur formation en la finançant directement par leur CPF crédité en euros et non plus en heures, peut néanmoins partiellement expliquer le nombre élevé d'entrées en formation en 2020, malgré la crise du Covid-19.

Ces évolutions montrent un effort conséquent en faveur de la formation des demandeurs d'emploi. Elles peuvent cependant refléter des évolutions sur le nombre de demandeurs d'emploi. Le taux d'accès à la formation, défini comme la probabilité d'accéder à la formation, illustre de manière plus précise les conséquences pour les demandeurs d'emploi de l'investissement en faveur de la formation. En pratique, pour chaque cohorte trimestrielle d'entrées au chômage, on calcule la proportion de demandeurs d'emploi accédant à la formation à des horizons temporels successifs.

Parmi les personnes entrées au chômage en janvier 2017, environ 8,5% ont suivi au moins une formation dans les 12 mois suivants (figure 2). Ce chiffre s'élève à près de 11% pour celles entrées au chômage au premier semestre 2019, mais cette tendance à la hausse a été ensuite interrompue par la crise du Covid-19. La figure 2 montre également d'importants effets saisonniers dans les taux d'accès à la formation, notamment pour les horizons à 6 mois. En effet,

7. Les formations de moins de 30 heures semblent trop courtes pour avoir un effet propre en termes d'acquisition de connaissances et de compétences professionnelles.

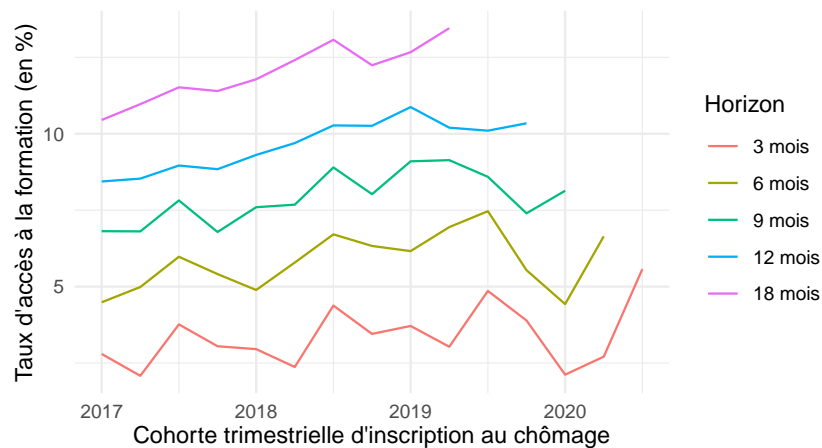
FIGURE 1 – Nombre de demandeurs d’emploi formés par année



Champ : Ensemble des inscrits à Pôle emploi entre janvier 2017 et décembre 2020, France entière.
Source : Dares, dispositif ForCE.

les entrées en formation ne se font pas toujours en continu sur l’année. Certaines formations longues (notamment les formations certifiantes) sont calées sur le calendrier scolaire : sur la période, 16% des entrées en formation en moyenne se font sur le mois de septembre. Les demandeurs d’emploi entrant au chômage au 3^e trimestre d’une année civile ont donc plus de chances d’entrer en formation dans les mois suivants. À l’inverse, ceux entrant au chômage au 2^e trimestre ont des taux d’accès à la formation plus faibles car peu de formations commencent durant l’été. Ces effets saisonniers se dissipent néanmoins lorsque l’horizon considéré est supérieur à un an.

FIGURE 2 – Taux d’accès à la formation



Champ : Ensemble des inscrits à Pôle emploi entre janvier 2017 et septembre 2020, France entière.
Source : Dares, dispositif ForCE.

Le détail des taux d’accès selon les types de formation et les caractéristiques des publics est

disponible en annexe A.

Une entrée en formation recouvre des réalités très hétérogènes selon l’objectif, le volume horaire et l’étalement de la formation dans le temps. En moyenne sur 2017/2019, 33,4 % des formations sont certifiantes (tableau 1). Près de 16 % des formations ont vocation à délivrer des savoirs de base, parfois pour une remise à niveau en vue d’une autre formation. Elles seront désignées par la suite comme “pré-qualifiantes”. 12 % concernent les formations d’adaptation au poste de travail (Action de formation préalable au recrutement (AFPR)/Préparation opérationnelle à l’emploi (POE)), qui sont préalables à un recrutement et incluent donc des périodes en entreprise. Les autres formations se partagent entre les formations professionnalisantes (13,9 %), de perfectionnement et élargissement des compétences (8,4 %) et d’aide à la création d’entreprise (4,6 %). Certaines n’ont pas d’objectif spécifique renseigné (12 %). Le volume horaire moyen d’une formation est de 467 heures. Plus de 20 % d’entre elles durent plus de 700 heures et 30 % durent moins de 120 heures. À volume horaire donné, une formation peut être plus ou moins étalée dans le temps. Dans 61 % des cas, le demandeur d’emploi n’est plus en formation trois mois après son début (tableau 1), et seulement une petite minorité (3,6 %) l’est encore douze mois après.

TABLE 1 – Répartition des formations selon leur type

	<i>Part</i>
Objectif	
Certifiante	33,4 %
Adaptation de poste (AFPR/POE)	12%
Pré-qualifiante	15,7 %
Autres	38,9 %
Volume horaire	
Plus de 700h	21,2 %
Entre 400h et 700h	12,9 %
Entre 120h et 400h	35,7 %
Moins de 120h	30,2 %
Etalement (en mois)	
Plus de 12 mois	3,6 %
Entre 3 et 12 mois	35,7 %
Entre 1 et 3 mois	32,4 %
Moins d’1 mois	28,3 %

Champ : Ensemble des demandeurs d’emploi primo-formés entre 2017-2019, France entière.

Source : Dares, dispositif ForCE.

Profil des primo-formés

Les primo-formés inscrits à Pôle Emploi sont plus jeunes, plus diplômés et inscrits au chômage depuis moins longtemps que la moyenne des demandeurs d’emploi (tableau 2). Ainsi, 28 % des primo-formés ont moins de 26 ans (contre 18 % des inscrits⁸) et 53 % ont un niveau de

8. Cette proportion correspond à une moyenne annuelle, estimée en stock. On considère l’ensemble des demandeurs d’emploi, à l’exclusion de ceux inscrits en catégorie E, qui sont en emploi et non tenus de

diplôme inférieur au bac (contre 57 % des inscrits). Alors que sur le stock des demandeurs d'emploi, plus de la moitié est inscrite depuis plus d'un an sur la période, ces chômeurs de longue durée ne représentent qu'un tiers des primo-formés. Les hommes sont sur-représentés parmi les primo-formés (51,5 %), alors qu'ils représentent 48 % des inscrits au chômage. Les proportions de demandeurs d'emploi vivant dans un quartier prioritaire de la ville (QPV, anciennement ZUS), soit 12 %, ou de ceux souffrant de handicap (8 %) sont, en revanche, proches de celles de l'ensemble des demandeurs d'emploi.

L'étude du profil des primo-formés suggère qu'il est d'emblée exclu de comparer naïvement les taux de retour à l'emploi des formés avec ceux des non-formés pour estimer l'effet de la formation tant ces deux groupes diffèrent. Pour estimer l'effet de la formation sur le retour à l'emploi, il faut donc recourir à des techniques d'appariement (*matching*) afin de rendre comparable le groupe traité et le groupe non-traité.

TABLE 2 – Répartition des demandeurs d'emploi selon leur profil

	<i>Ensemble des demandeurs d'emploi</i>	<i>Primo-formés</i>
Diplôme*		
Supérieur à bac + 2	14,7 %	14 %
Bac / Bac + 2	28,5 %	33 %
Inférieur au bac	56,8 %	53 %
Age		
Plus de 50 ans	22 %	11,4 %
Entre 26 et 50 ans	60,2 %	60,9 %
Moins de 26 ans	17,8 %	27,7 %
Sexe		
Hommes	48,1 %	51,5 %
Femmes	51,9 %	48,5 %
Ancienneté au chômage		
Plus d'un an	53,9 %	33 %
Un an ou moins	46,1 %	67 %
En QPV/ZUS	12,6 %	12 %
En situation de handicap	8,3 %	8,2 %

* Les valeurs manquantes sont exclues ici.

Note : Les personnes en situation de handicap correspondent aux demandeurs d'emploi bénéficiant de l'obligation d'emploi.

Champ : Ensemble des demandeurs d'emploi inscrits à Pôle emploi entre 2017 et 2019 (sauf catégorie E), France entière.

Source : Dares, dispositif ForCE.

Lecture : Les femmes représentent 48,5% des personnes entrées en formation pour la première fois entre 2017 et 2019 et 51,9% des demandeurs d'emploi inscrits au cours de cette période.

rechercher un emploi (par exemple : bénéficiaires de contrats aidés, créateurs d'entreprise). Pour chaque mois, on considère la structure selon les caractéristiques individuelles des demandeurs d'emploi qui sont inscrits au moins un jour du mois considéré à Pôle emploi. On prend ensuite la moyenne de ces structures mensuelles sur l'année.

3 Stratégie empirique

Variable de traitement et variables d'intérêt

L'objectif de l'étude est de comparer les chances de retour à l'emploi d'une personne inscrite au chômage, selon qu'elle décide de suivre une formation ou pas à une date donnée.

Plus précisément, les variables d'intérêt sont des indicatrices de retour à l'emploi sur l'intervalle entre le mois d'entrée en formation et 1, 3, 6, 9, 12, 18, 24 mois⁹. On définit deux variables d'intérêt :

- le retour à l'emploi salarié dans le secteur privé¹⁰ d'au moins 31 jours, sans distinction sur le type de contrat.
- le retour à l'emploi durable, défini comme un contrat salarié du secteur privé en CDI ou dont la durée observée est supérieure à 180 jours.

La variable de traitement est le fait d'entrer en formation pour la première fois un mois donné.

- $D_{i,t} = 1 \Leftrightarrow$ Le demandeur d'emploi i entre en première formation le mois t .
- $D_{i,t} = 0 \Leftrightarrow$ Le demandeur d'emploi i n'entre pas en formation le mois t (ni en première formation ni en aucune autre formation de plus de 30 heures) et n'a pas suivi de formation depuis au moins 2014.

À noter que les données d'emploi disponibles pour la réalisation de cette étude s'arrêtent au deuxième trimestre 2020. Pour disposer d'un recul suffisant, on ne considère donc que les entrées en formation de janvier 2017 à septembre 2019. Par ailleurs, le traitement est défini par le fait de suivre la formation *et* la date de celle-ci (Sianesi, 2004) : la question est bien de savoir si, pour un demandeur d'emploi, il est intéressant d'entrer en formation ce mois précis. En pratique, cela signifie qu'on conserve dans le groupe de contrôle les personnes qui entrent en formation *après* le mois considéré. Typiquement, exclure des personnes qui entrent en formation après la date considérée revient à exclure des personnes qui sont encore au chômage à cette date, en sur-sélectionnant donc dans le groupe de contrôle les personnes qui ont des chances élevées de retour à l'emploi à court terme. Ces effets de sélection dynamique, qui conduisent à conditionner par le futur, peuvent conduire à sous-estimer l'impact de la formation (Biewen *et al.*, 2012). En revanche, on exclut les personnes qui ont suivi des formations avant la date considérée.

En pratique, le groupe de contrôle est de taille beaucoup plus large que le groupe de traitement. Du fait de contraintes opérationnelles liées aux capacités de calcul du logiciel R utilisé dans cette étude, on restreint le groupe de contrôle à un échantillon aléatoire de 20% des non-formés, obtenu par tirage stratifié selon le numéro d'agence Pôle emploi, l'âge et le sexe. On vérifie que l'échantillon tiré est comparable à l'ensemble des non-traités sur les autres variables (notamment en termes d'ancienneté au chômage et de niveau de diplôme).

Hypothèses d'identification

Pour estimer l'effet du traitement, on s'appuie sur le cadre théorique développé par Rosen-

9. Pour certaines cohortes de 2017 et de début 2018, il est aussi possible de calculer ces variables à horizon 30 et 36 mois.

10. hors salariés des particuliers employeurs et une partie de l'emploi salarié agricole.

baum et Rubin (1983). En particulier, nous estimons un modèle d'appariement (*matching*) mêlant *matching sur score de propension* et *matching exact* sur plus proches voisins.

Le principe des méthodes d'appariement est de réduire les biais de sélection qui existent dans les données observationnelles (non expérimentales), dans la mesure où les personnes qui se forment ont des caractéristiques individuelles différentes de celles qui ne se forment pas. Les méthodes d'appariement consistent à restreindre le groupe de contrôle aux personnes dont les caractéristiques observables moyennes sont les plus proches possibles des personnes formées. Cette méthode ne permet donc que de réduire les biais de sélection liés à des caractéristiques observables, alors que d'autres déterminants (par exemple l'existence d'un projet de reconversion, une motivation particulière, etc.) peuvent également influencer à la fois sur le projet de formation et les chances de retour à l'emploi. Les méthodes d'appariement sont cependant en principe plus robustes que des méthodes simples de régression linéaire (notamment lorsque les caractéristiques observables sont très différentes entre les groupes), car elles ne reposent pas sur une extrapolation linéaire de la relation entre la variable d'intérêt et les caractéristiques observables (Imbens et Wooldridge, 2009). Les méthodes d'appariement reposent sur une intuition simple : il s'agit, pour chaque demandeur d'emploi entrant en formation, de trouver un demandeur d'emploi non formé ayant les mêmes caractéristiques observables. Sous l'hypothèse d'indépendance conditionnelle (qui suppose que toute la sélection se fait sur les variables observables), et de support commun (qui suppose que pour chaque valeur des observables, il est possible de trouver des formés et des non formés), ces méthodes permettent d'identifier l'effet moyen du traitement sur les traités.

Formellement, on note respectivement $Y(0)_{i,t}$ le résultat potentiel de la variable d'intérêt du demandeur d'emploi i s'il n'est pas traité au mois t et $Y(1)_{i,t}$ s'il est traité. Par exemple, lorsque la variable de résultat étudiée est le retour à l'emploi 12 mois après l'entrée en formation, $Y(0)_{i,t}$ correspond à la probabilité pour un demandeur d'emploi i de retrouver un emploi entre t et $t + 12$ s'il n'entre pas en formation à la date t , et $Y(1)_{i,t}$ cette même probabilité s'il entre en formation en t .

Pour pouvoir estimer l'effet de la formation sur les personnes formées (*Average Treatment Effect on the Treated*, ci-après dénoté ATT) avec une méthode de matching, il faut que le résultat potentiel sans traitement $Y(0)_{i,t}$ soit indépendant de la variable de traitement conditionnellement aux autres variables observables, notées X . L'hypothèse d'indépendance conditionnelle s'écrit :

$$\forall(i, t), Y(0)_{i,t} D_{i,t} \mid X_{i,t} \quad (1)$$

Cependant, les méthodes d'appariement peuvent être rapidement confrontées à un problème de dimension, dès lors que les caractéristiques observables sont nombreuses (et/ou continues). Il peut s'avérer impossible d'apparier des personnes traitées sur l'ensemble des dimensions considérées. Cette limite est en principe dépassée par l'utilisation du score de propension, comme démontré par Rosenbaum et Rubin (1983). Le score de propension désigne la probabilité de recevoir le traitement (ici d'être formé le mois considéré), conditionnellement à l'ensemble des caractéristiques observables. Pour une même valeur du score de propension, les caractéristiques observables sont en effet en moyenne identiques dans les groupes traités et

non traités. Formellement, le score de propension est une fonction équilibrante des covariables observables X , c'est-à-dire :

$$XD \mid p(X) \tag{2}$$

où $p(X) = P(D = 1 \mid X)$. On a alors sous hypothèse d'indépendance conditionnelle, $Y(0)D \mid p(X)$.

Il est donc en théorie équivalent d'apparier sur l'ensemble de ces dimensions et sur le score de propension, ce qui ramène à un problème à une dimension. Cependant, la mise en œuvre de l'appariement sur le score de propension est tributaire de deux questions importantes. En premier lieu, l'appariement ne peut se faire que sur le support commun de respectivement la distribution du score de propension dans le groupe de traitement et de celle de ce score dans le groupe de contrôle : c'est-à-dire pour les personnes formées pour lesquelles il existe des personnes non formées ayant la même probabilité d'être formées (même si elles ne l'ont pas été...). Cela impose *a minima* que la probabilité d'être traité ne dépend pas de manière déterministe des variables observables ($p(D|X) \notin \{0, 1\}$). Plus généralement, l'appariement ne pourra se faire que sur ce support commun, et il faut s'assurer que celui-ci est suffisamment large pour permettre une généralisation des résultats obtenus.

Par ailleurs, l'effet du traitement n'est identifié que sous une troisième hypothèse, la Stable-Unit-Treatment-Value Assumption (SUTVA). Cette hypothèse postule que la formation suivie par une personne n'a d'impact que sur cette personne. Elle exclut donc les possibilités d'externalités ou d'équilibre général - en particulier, cette hypothèse peut être remise en cause si la formation permet essentiellement de "remonter la file d'attente" de sortie du chômage, en assurant un bénéfice à certains mais au détriment des autres demandeurs d'emploi. De tels mécanismes ont été mis en avant sur certains dispositifs, notamment par Crépon *et al.* (2013). Cependant, si l'on ne peut exclure de tels effets indirects de la formation, la population des formés reste faible par rapport à l'ensemble des demandeurs d'emploi, et il est plausible de supposer que l'ampleur des effets de substitution est relativement faible.

Sous les hypothèses d'indépendance conditionnelle aux observables et de support commun, il est possible de calculer l'effet moyen du traitement sur les traités (ATT) :

$$\begin{aligned} ATT &= E[Y(1) - Y(0) \mid D = 1] \\ &= E[Y(1) \mid D = 1] - E_X[E_Y(Y(0) \mid D = 1, p(X))] \quad \text{par la loi des espérances itérées} \\ &= E[Y(1) \mid D = 1] - E_X[E_Y(Y(0) \mid D = 0, p(X))] \quad \text{par hypothèse d'indépendance conditionnelle} \end{aligned}$$

Pour estimer $E[Y(1) \mid D = 1]$, il suffit de calculer la moyenne empirique de la variable d'intérêt sur la population traitée. Pour estimer $E_X[E_Y(Y(0) \mid D = 0, p(X))]$, on calcule la moyenne empirique de la variable d'intérêt sur les demandeurs d'emploi non formés à la date t ayant le même score de propension que les formés.

En pratique, deux problèmes se posent néanmoins : d'une part, le score de propension est inconnu et, d'autre part, il s'agit d'une fonction continue pour laquelle la probabilité d'un appariement exact est nulle. On estime ici le score de propension à l'aide d'un modèle logit. L'appariement se fait ensuite de manière approchée, chaque traité étant apparié avec des

non-traités ayant le score de propension le plus proche. Ici, il a été choisi d'apparier sur trois plus proches voisins. Ce choix est le résultat d'un arbitrage entre robustesse des estimations (plus le nombre de voisins est élevé moins l'estimation est dépendante d'un résultat particulier) et précision (plus on augmente le nombre de voisins moins on sélectionne des voisins proches de l'unité traitée). Des tests de robustesse ont été menés (voir section 5) pour vérifier la sensibilité des résultats à ce choix.

Soit I_1 la population des traités, N_1 le nombre de traités, i_0 l'ensemble des non-traités appariés au traité i et w_j le poids d'un non-traité j . Dans notre cas, on apparie chaque traité avec trois non-traités donc l'ensemble i_0 est composé de trois individus et les poids correspondant w_j valent alors $\frac{1}{3}$.

L'effet du traitement estimé est :

$$\widehat{ATT} = \frac{1}{N_1} \sum_{i \in I_1} (Y_i - \sum_{j \in i_0} w_j Y_j)$$

L'hypothèse d'indépendance conditionnelle implique qu'il n'existe pas de variables inobservables susceptibles d'influer à la fois sur la probabilité d'être traité et sur la variable d'intérêt. C'est évidemment une hypothèse très restrictive. Les données disponibles ne captent pas toutes les dimensions pouvant influencer à la fois sur la décision d'entrer en formation et les chances de retour à l'emploi (par exemple la motivation pour un parcours professionnel, certaines aptitudes particulières, ou encore les revenus passés et futurs). Cependant, les données disponibles dans le dispositif ForCE permettent de retracer une grande partie de la trajectoire professionnelle et des caractéristiques socio-démographiques, qui peuvent avoir un impact important pour réduire le biais de sélection (Biewen *et al.*, 2012). Le contexte local, notamment en termes d'offre de formation et de dynamisme du marché du travail, est également à prendre en compte. Les variables considérées sont les suivantes, pour l'essentiel issues du Fichier Historique des demandeurs d'emploi (FH) :

- caractéristiques socio-démographiques : âge, sexe, niveau de diplôme, qualification, situation familiale, nombre d'enfants, nationalité, situation de handicap, habitat dans un quartier prioritaire de la ville, habitat en zone de revitalisation rurale ;
- trajectoires antérieures : durée depuis l'inscription à Pôle emploi pour l'épisode considéré, temps passé inscrit à Pôle emploi avant l'épisode considéré, nombre d'inscriptions à Pôle emploi dans les 10 dernières années, nombre d'heures travaillées en activité réduite depuis le début de l'épisode considéré, motif d'inscription à Pôle emploi, catégorie d'inscription, ROME de l'emploi recherché, expérience dans le ROME de l'emploi recherché ;
- proxy de l'état de santé et de motivation : temps passé en maladie depuis le début de l'épisode considéré, temps passé en maladie inscrit à Pôle emploi avant l'épisode considéré, type de contrat recherché, mobilité déclarée entre le logement et l'emploi recherché¹¹, nombre d'absences à des contrôles ou entretiens de Pôle emploi ;

11. La distance déclarée pouvant être en termes de temps ou de distance, nous créons une variable indicatrice

- effets fixes correspondant au numéro d'agence Pôle emploi, pour capter des spécificités géographiques.

Un nouveau modèle est estimé pour chaque mois entre janvier 2017 et septembre 2019, ce qui permet de capter implicitement des effets spécifiques à la période étudiée (évolution de la politique de formation dans le temps, évolution de la conjoncture économique, etc.).

Cependant, deux personnes avec un score de propension proche peuvent avoir des valeurs très différentes sur certaines variables, dont certaines peuvent être déterminantes pour le retour à l'emploi comme l'ancienneté d'inscription (Biewen *et al.*, 2012). Afin de rendre plus robuste l'estimation de l'effet du traitement et d'éviter des appariements discutables, nous exploitons la grande quantité d'unités contrôles disponibles pour ajouter des contraintes. Ainsi, nous combinons *matching sur score de propension* et *matching exact* sur la catégorie d'âge¹² et la zone d'emploi Insee¹³. De plus, nous imposons une limite de différence d'ancienneté d'inscription de 30 jours maximum et de différence de temps total d'inscription avant l'épisode considéré de 150 jours maximum.

La section 5 propose plusieurs tests de robustesse des résultats obtenus, en utilisant d'autres méthodes de matching ou des spécifications alternatives.

Estimation de la précision des estimateurs

Le calcul des écarts types s'appuie sur les travaux d'Abadie et Spiess (2021) consacrés aux méthodes d'inférence dans un cadre de *matching*. Le choix de la méthode a été dicté par des considérations pratiques liées à la dimension de l'échantillon, qui rend notamment impraticable l'utilisation de méthodes de *bootstrap*. Ces méthodes, couramment utilisées pour le calcul d'inférence des estimateurs de *matching*, consistent à ré-échantillonner par tirage avec remplacement, pour obtenir une distribution empirique des estimateurs. Cependant, la taille de l'échantillon (plus d'un million de personnes traitées) et les temps associés à une seule estimation rendent peu opérationnelle cette méthode.

Comme détaillé par Abadie et Spiess (2021), il est possible d'estimer des écarts-types convergents de l'effet estimé de la formation à partir d'une régression linéaire "locale" (clusterisée au niveau de chaque appariement), qui est mise en œuvre sur les estimateurs. On vérifie comme test de robustesse que les écarts-types obtenus sont proches de ceux calculés en utilisant des méthodes de *bootstrap* (voir section 5.1).

"longue distance" égale à 1 si le temps de trajet déclaré est supérieur à une heure ou si la distance déclarée est supérieure à 40 km.

12. 3 catégories d'âge sont considérées : moins de 26 ans, entre 26 et 50 ans et plus de 50 ans.

13. Une zone d'emploi est un espace géographique à l'intérieur duquel la plupart des actifs résident et travaillent, et dans lequel les établissements peuvent trouver l'essentiel de la main d'œuvre nécessaire pour occuper les emplois offerts. Il existe 306 zones d'emploi en 2020, ce qui en fait une unité géographique précise et pertinente pour l'étude du marché du travail.

4 Résultats

Les résultats détaillés de l'estimation du score de propension sont disponibles en annexe B1. Les variables les plus corrélées avec l'entrée en formation sont l'âge, l'histoire professionnelle du demandeur d'emploi (ancienneté d'inscription, nombre d'épisodes connus, expérience professionnelle dans le secteur d'emploi recherché,...), le type d'emploi recherché (type de contrat et secteur) et le niveau de diplôme.

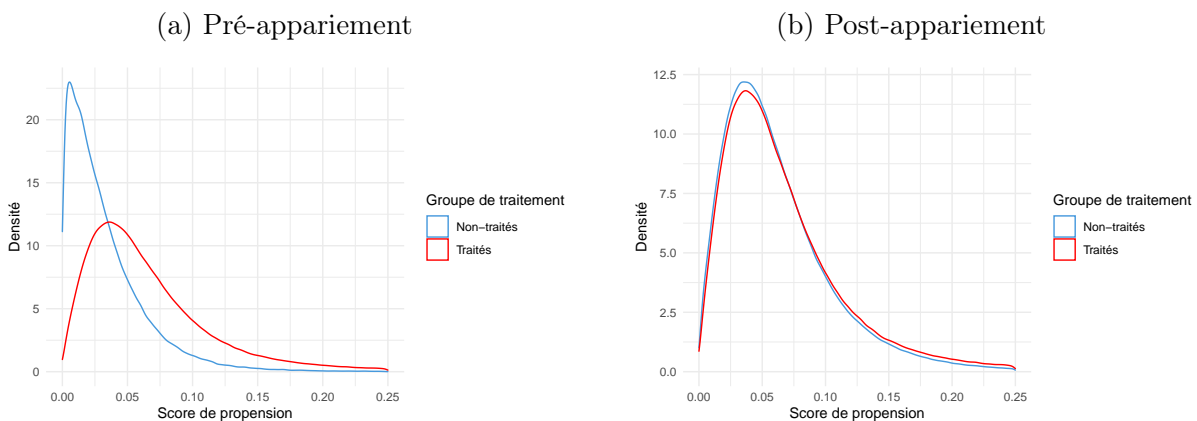
4.1 Evaluation de la qualité du modèle

Deux dimensions sont importantes à vérifier pour évaluer la qualité du modèle. D'une part, la taille du support commun, c'est-à-dire le nombre d'observations du groupe de traitement qui peuvent être appariées avec le groupe de contrôle. Un support commun trop réduit signifie que la généralisation des résultats obtenus sera questionnable. D'autre part, l'équilibre des covariables, c'est-à-dire le fait que l'appariement a bien permis de rendre la composition des deux groupes identiques en termes de caractéristiques observables.

Support commun

Comme on peut le voir sur la figure 3, les deux distributions du score de propension pour respectivement les personnes formées et les non formées se recouvrent très largement. Ce large support commun signifie qu'il est possible d'apparier chaque personnes formées avec des non formées, et donc assure que les résultats obtenus seront représentatifs de l'ensemble de la population des formés. De fait, la grande taille du groupe de contrôle comparée au groupe traité permet l'appariement des formés même dans les queues de distribution, comme on peut le voir dans la distribution post-appariement. Par ailleurs, si les deux distributions sont assez différentes avant appariement, la restriction aux plus proches voisins parmi les non formés permet une quasi superposition des distributions.

FIGURE 3 – Support du score de propension selon le groupe de traitement



Malgré les contraintes assez importantes imposées pour l'appariement (notamment un appariement exact sur la zone d'emploi et presque exact sur l'ancienneté d'inscription), 97,5%

des traités ont pu être appariés avec au moins un individu du groupe de contrôle (tableau 3).

TABLE 3 – Effectifs des groupes de traitement avant et après appariement

	Groupe traité	Groupe de contrôle
Nombre avant l'appariement	1 246 595	9 333 538
Nombre après l'appariement	1 202 913	2 760 026
Proportion après l'appariement	96,5 %	29,6 %

Equilibre des covariables

Trois statistiques sont calculées pour évaluer la qualité de l'appariement (voir Stuart, 2010 et Ho *et al.*, 2011) :

- Différence de moyenne standardisée (*Standardized Mean Difference*) : différence entre les moyennes du groupe traité et du groupe de contrôle, divisée par l'écart-type de la variable dans le groupe traité. Pour éviter qu'un biais de sélection persiste après l'appariement, cette statistique doit être idéalement inférieure (en valeur absolue) à 0,1 et ne doit pas excéder 0,25 (Stuart, 2010).
- Rapport de variance (*Variance ratio*) : rapport des variances du groupe traité et du groupe de contrôle. Le rapport de variances doit être idéalement proche de 1 pour que les deux groupes soient les plus proches possibles. Si les ratios de variance sont compris entre 0,5 et 2, l'appariement peut être considéré comme satisfaisant (Stuart, 2010).
- Statistiques de Kolmogorov-Smirnov : différence maximale entre les fonctions de répartition empiriques des deux groupes de traitement. Cette statistique est un bon complément à la différence de moyenne standardisée car elle permet d'évaluer la similitude des deux groupes sur l'ensemble de la distribution et non une simple moyenne. Si l'objectif est que cette valeur soit la plus faible possible, il n'existe pas de seuils faisant référence dans la littérature concernant cette statistique¹⁴. Il est cependant possible de comparer les valeurs de la statistique avant et après l'appariement.

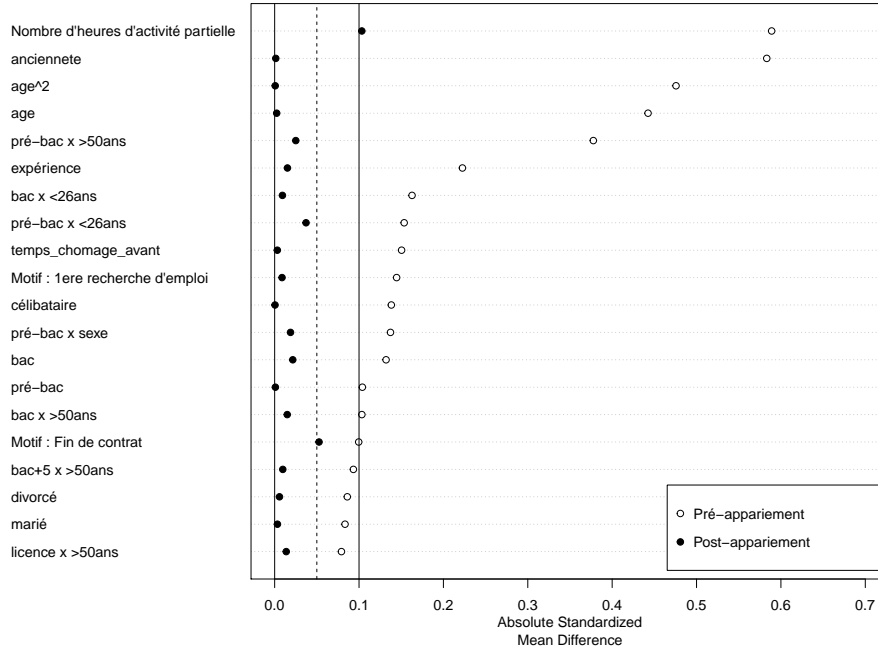
La figure 4 illustre l'amélioration de l'équilibre entre les deux groupes permise par l'appariement sur score de propension couplé à des contraintes supplémentaires (voir section 3), en utilisant les différences de moyenne. Seules sont représentées les vingt variables qui présentaient les plus fortes hétérogénéités dans les deux groupes avant appariement. Pour chacune de ces variables, l'appariement améliore nettement l'équilibre entre les deux groupes. Le détail des autres indicateurs est fourni en annexes (tableau B.2). Pour la plupart des variables continues, les ratio de variance sont très proches de 1¹⁵. Les statistiques de Kolmogorov-Smirnov

14. En effet, cette statistique est généralement utilisée dans un test d'hypothèse d'égalité de deux distributions. Cependant, ces tests ne sont pas pertinents ici car ils dépendent de la taille de l'échantillon alors que l'équilibre entre les groupes de traitement n'est pas lié à la taille de l'échantillon.

15. Les valeurs les plus grandes observées sont pour le score de propension, avec une valeur après appariement de 1,7 nettement plus faible que la valeur observée avant appariement.

après appariement sont toutes inférieures à 0,1 soit des valeurs divisées par au moins 2 par rapport à leurs valeurs sur les échantillons avant appariement.

FIGURE 4 – Équilibre des covariables avant et après l’appariement (différences de moyennes standardisées)



Note : Les 20 variables ayant la différence de moyenne standardisée la plus élevée avant l’appariement sont représentées.

Les seuils 0,05 et 0,10 sont représentés par des lignes verticales (respectivement en pointillés et continue).

4.2 Résultats

Nous indiquons d’abord les taux de retour à l’emploi des primo-formés et des non-formés avant l’appariement. La différence entre ces taux est appelée “différence brute”. Elle ne peut être interprétée comme l’effet moyen de la formation compte tenu de la différence de composition entre les deux groupes. Nous indiquons ensuite, afin d’illustrer l’importance de prendre en compte les différences d’ancienneté d’inscription entre traités et non-traités, la "différence ajustée", qui est calculée en équilibrant les deux groupes de traitement sur la seule variable de l’ancienneté d’inscription. Pour réaliser cet équilibre, nous effectuons un *matching exact* sur des tranches de 30 jours d’ancienneté d’inscription puis nous estimons un modèle de régression linéaire pondérée par le poids que représente chaque tranche d’ancienneté dans le groupe traité. Cette différence ajustée permet de réduire le biais de sélection en comparant des primo-formés avec des non-formés ayant la même ancienneté d’inscription. Cependant, elle ne peut toujours pas être interprétée comme l’effet moyen du traitement car elle néglige l’effet des autres covariables. Nous indiquons donc enfin notre estimation de l’effet moyen du traitement sur les traités (ATT), calculée selon la méthode présentée dans la section 3.

TABLE 4 – Taux de retour à l’emploi et estimation de l’effet de la première formation sur le retour à l’emploi

	Taux de retour à l’emploi des primo-formés (1)	Taux de retour à l’emploi des non-formés (2)	Différence brute (1) – (2) (en points)	Différence ajustée (en points)	ATT (en points)
Tout type d’emploi					
6 mois	29,5 %	25,5 %	4	2,3	0,6 (0,27)**
12 mois	48,3 %	37,4 %	10,9	8,3	5,7 (0,29)***
18 mois	60,1 %	45 %	15,1	12,1	8,8 (0,29)***
24 mois	66,9 %	50,3 %	16,6	13,2	9,2 (0,29)***
Emploi durable					
6 mois	20,8 %	13,9 %	6,9	5,2	4 (0,24)***
12 mois	34,5 %	23,2 %	11,3	8,8	6,8 (0,28)***
18 mois	45,4 %	30,5 %	14,9	11,9	9,1 (0,29)***
24 mois	52,6 %	36,2 %	16,4	13	9,5 (0,30)***

Note : * p<0,1 ; ** p<0.05 ; *** p<0.01.

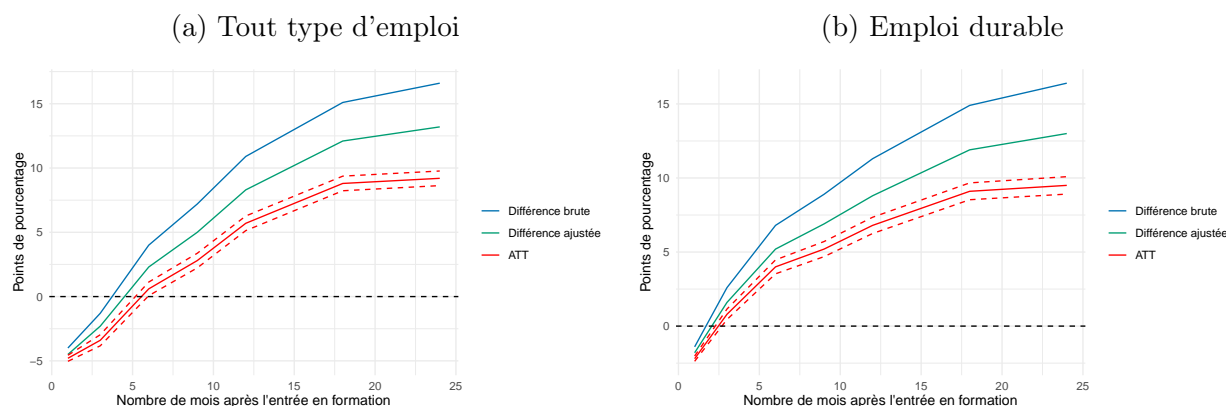
L’emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L’emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Moyenne pondérée des estimations mensuelles obtenues entre janvier 2017 et septembre 2019 (horizons 6 et 12 mois), janvier 2017 et mars 2019 (horizon 18 mois), et janvier 2017 et septembre 2018 (horizon 24 mois). *Lecture* : 48,3 % des primo-formés ont retrouvé un emploi dans les 12 mois après leur entrée en formation, soit 5,7 points de plus que des non-formés ayant des caractéristiques observables similaires.

Champ : Ensemble des demandeurs d’emploi inscrits à Pôle emploi entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

FIGURE 5 – Estimation de l’effet de la première formation sur le retour à l’emploi



Champ : Ensemble des demandeurs d’emploi inscrits entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

Note : Moyenne pondérée des estimations mensuelles obtenues entre janvier 2017 et septembre 2019.

La figure 5 montre l’effet estimé de la formation, à partir de l’entrée en formation. Sur les six premiers mois après l’entrée en formation, un demandeur d’emploi qui entre en formation a des chances de retour à l’emploi plus faibles que des demandeurs d’emploi avec des caractéristiques observables comparables. Cet effet négatif, appelé effet *lock in*, s’explique par le fait que la formation réduit le temps de recherche d’emploi. Six mois après l’entrée en formation, les chances de retour à l’emploi sont identiques pour les personnes formées (autour de 30%) et non formées. Elles deviennent ensuite plus élevées en faveur des personnes formées. Au bout de 2 ans, la probabilité de retour à l’emploi des formés est supérieure de plus de 9 points de pourcentage par rapport aux non formés comparables, un ordre de grandeur cohérent

avec ceux trouvés dans la littérature existante (voir par exemple Card *et al.*, 2017). Cet effet est par ailleurs persistant sur l’horizon temporel observé (jusqu’à 36 mois après l’entrée en formation).

Par ailleurs, la formation semble permettre un retour plus rapide vers des emplois de meilleure qualité (lorsque l’on se concentre sur l’emploi durable). Dès deux mois après l’entrée en formation, les personnes formées ont une probabilité identique de retrouver un emploi durable que des non formés comparables, et l’effet est un peu plus élevé ensuite.

4.3 Analyses par type de formation

Les écarts entre formés et non-formés diffèrent selon les objectifs, le volume horaire et l’étalement dans le temps de la formation. Pour prendre en compte l’hétérogénéité des effets de la formation selon ces dimensions, il est possible d’estimer le modèle en redéfinissant le traitement comme l’entrée dans un certain type de formation (le groupe contrôle reste composé uniquement de demandeurs d’emploi n’entrant pas en formation). Nous estimons donc pour chaque type de formation un score de propension différent et appliquons ensuite la même méthode que celle décrite dans la section 3.

Du fait de l’effet *lock-in*, les formations longues (définies comme plus de 700 heures) et certifiantes (dont respectivement 15 % et 7 % sont étalées sur plus d’un an, voir tableau 5) donnent des écarts positifs plus tard. Pour isoler ce qui relève de cet effet spécifique lié à l’étalement, on distingue les formations selon qu’elles s’étalent sur plus ou moins d’un an. Afin d’avoir un recul plus large sur le retour à l’emploi, seules les formations ayant débuté en 2017 sont considérées dans cette sous-section, ce qui permet de calculer des taux de retour à l’emploi à des horizons supérieurs à deux ans.

TABLE 5 – Durée et étalement moyens selon le type de formation

	<i>Durée moyenne</i> (en heures)	<i>Étalement moyen</i> (en mois)	<i>Part de formations étalées</i> sur plus d’un an
Formations longues	1 355	10,5	15,2 %
Formations certifiantes	734	6	7,3 %
Formations d’adaptation de poste	244	1,7	0 %
Formations pré-qualifiantes	376	3,2	1,3 %

Note : Les formations longues désignent les formations de plus de 700 heures.

Une formation longue peut avoir tout type d’objectif. En revanche, les 3 types d’objectifs indiqués ici sont mutuellement exclusifs.

Champ : Premières formations de plus de 30 heures, 2017, France entière.

Source : Dares, dispositif ForCE.

Lorsqu’on se restreint aux formations certifiantes étalées sur moins d’un an, l’écart estimé entre formés et non formés 12 mois après l’entrée en formation est identique à celui observé sur l’ensemble des formations. Au bout de 18 mois, il est supérieur pour les formations certifiantes (2,8 points de pourcentage de plus que l’écart moyen observé sur l’ensemble des

formations). Cet avantage relatif apporté par les formations certifiantes met cependant plus de temps à apparaître lorsque l'on prend aussi en compte les 7% de celles-ci qui s'étalent sur plus d'un an : au bout de 18 mois, l'écart est identique à celui observé sur l'ensemble des formations. Deux ans après l'entrée en formation, l'écart est supérieur (0,5 points de plus que l'écart moyen). Si les formations certifiantes sont en général plus longues que la moyenne des formations, ce n'est cependant pas le volume horaire qui garantit une insertion plus rapide. Lorsqu'on considère l'ensemble des formations longues (de plus de 700 heures), l'écart entre formés et non formés est plus faible qu'en moyenne, même 36 mois après le début de la formation, et même en ne considérant que les formations étalées sur moins d'un an (tableau 6).

Les formations pour lesquelles l'écart entre formés et non-formés est le plus élevé sont les formations d'adaptation de poste. L'écart est important très rapidement car ce sont des formations courtes (244 heures en moyenne en 2017). L'ampleur de l'écart diminue avec le temps, mais reste à un niveau élevé. Ce résultat est à interpréter avec prudence, dans la mesure où par construction, ces formations sont adossées à des offres d'emploi et doivent, sauf imprévu, déboucher sur une embauche. Les méthodes de *matching* utilisées ici, si elles permettent de contrôler des dimensions observables, ne peuvent rendre compte de dimensions non directement mesurables (motivation, savoirs-être professionnels identifiables en entretien) qui expliquent à la fois le recrutement dans ce type de formation et la reprise d'emploi.

Finalement, les écarts entre formés et non-formés pour les formations pré-qualifiantes sont nettement inférieurs à la moyenne, et ce particulièrement pour l'emploi durable. Ces formations n'ont en général pas vocation à préparer directement à l'insertion sur le marché du travail, mais plutôt à offrir une remise à niveau en vue de formations complémentaires. Parmi les personnes ayant suivi une première formation pré-qualifiante, 38 % ont suivi une autre formation par la suite (cette proportion est de 24 % lorsque l'on considère l'ensemble des primo-formés, tout type de formation confondu).

TABLE 6 – Effet de la formation selon le type de formation

	ATT formations longues		ATT formations certifiantes		ATT adaptation	ATT pré-	ATT toutes
	Ensemble	<1an	Ensemble	<1an	poste	qualifiantes	formations
Tout type d'emploi							
12 mois	-9,6*** (0,58)	-6*** (0,65)	3,9*** (0,51)	6,7*** (0,53)	40,9*** (0,64)	-0,2 (0,75)	6,8*** (0,30)
18 mois	0,6 (0,60)	5,3*** (0,65)	9,3*** (0,50)	12,2*** (0,52)	34,7*** (0,60)	2,7*** (0,76)	9,4*** (0,29)
24 mois	3,6*** (0,57)	7,6*** (0,62)	10,4*** (0,47)	12,9*** (0,49)	30,3*** (0,56)	4,5*** (0,75)	9,9*** (0,28)
36 mois	6,5*** (0,53)	8,8*** (0,58)	11,2*** (0,44)	12,7*** (0,46)	25,1*** (0,52)	6*** (0,72)	10*** (0,26)
Emploi durable							
12 mois	-7,6*** (0,51)	-4,7*** (0,58)	2,6*** (0,47)	4,7*** (0,50)	48,7*** (0,70)	0,40 (0,67)	7,6*** (0,28)
18 mois	-0,50 (0,58)	3,8*** (0,65)	7,1*** (0,50)	9,6*** (0,53)	43,4*** (0,68)	2,4*** (0,73)	9,5*** (0,29)
24 mois	2,1*** (0,60)	6,4*** (0,66)	8,5*** (0,51)	11,1*** (0,53)	38,9*** (0,65)	3,7*** (0,75)	10*** (0,30)
36 mois	4,7*** (0,59)	8,1*** (0,64)	9,8*** (0,50)	11,8*** (0,52)	33,2*** (0,62)	4,9*** (0,76)	10,1*** (0,29)

Note : *p<0,1 ; ** p<0.05 ; *** p<0.01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Les formations longues recouvrent toutes les formations dont le volume horaire est supérieur à 700 heures.

Lecture : La probabilité d'avoir trouvé un emploi dans les 18 mois après l'entrée en formation certifiante étalée sur moins d'un an est en moyenne de 12,2 points de pourcentage supérieure à celle d'un demandeur d'emploi non-formé ayant un profil similaire. Cet effet est statistiquement significatif au seuil de 1%.

Champ : Ensemble des inscrits à Pôle emploi entre janvier et décembre 2017, France entière.

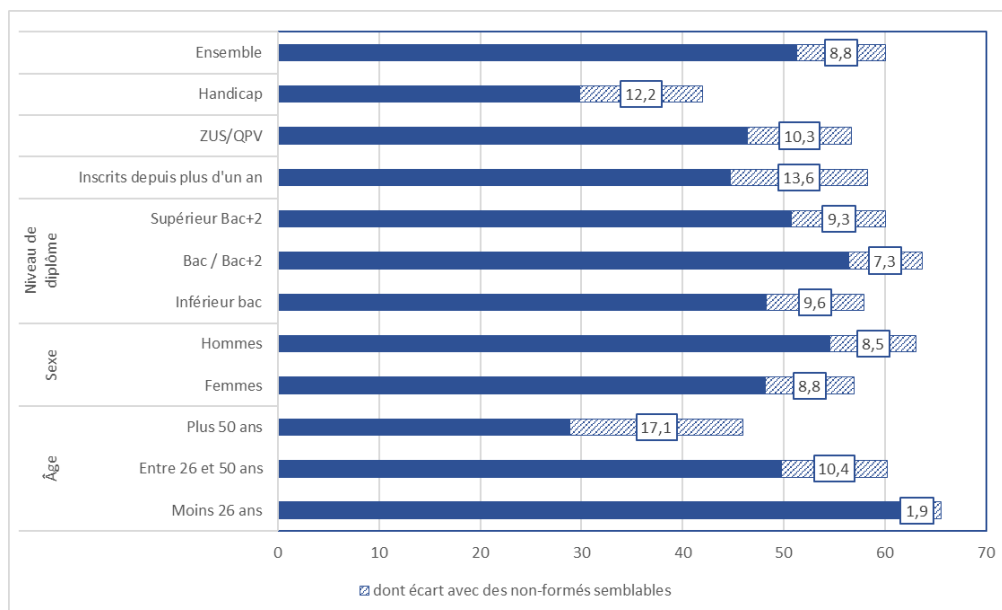
Source : Dares, dispositif ForCE.

4.4 Variations selon les caractéristiques des formés

La littérature a déjà montré que les effets de la formation pouvaient être très différents selon les types de public formé. Pour explorer cette dimension, nous estimons notre modèle en ajoutant une contrainte d'appariement exact sur une variable socio-démographique (sexe, niveau de diplôme, habitat en QPV ou ZUS, handicap)¹⁶. Par exemple, pour estimer l'effet de la formation spécifiquement sur le public des femmes, nous imposons qu'une femme primo-formée ne soit appariée qu'avec des femmes non-formées.

16. Pour l'âge, il n'est pas nécessaire d'ajouter une contrainte car notre modèle impose déjà un appariement exact sur la tranche d'âge (voir Section 3).

FIGURE 6 – Taux de retour à l’emploi des primo-formés et estimation de l’effet de la formation selon le type de public (horizon 18 mois)



Note : Les personnes en situation de handicap regroupent l’ensemble des demandeurs d’emploi bénéficiant de l’obligation d’emploi.

Lecture : La probabilité d’avoir trouvé un emploi dans les 18 mois suivant l’entrée en formation est en moyenne de 58,3 % pour les primo-formés inscrits à Pôle emploi depuis plus d’un an, soit 13,6 points de plus que celle d’un demandeur d’emploi non-formé inscrit depuis plus d’un an et ayant par ailleurs un profil similaire.

Champ : Ensemble des inscrits à Pôle emploi entre janvier 2017 et mars 2019, France entière.

Source : Dares, dispositif ForCE.

Les estimations montrent que l’ampleur des écarts entre formés et non formés varient selon les caractéristiques observables. Ainsi, 18 mois après l’entrée en formation, le taux de retour à l’emploi des personnes avec un diplôme inférieur au bac est de 9,6 points de pourcentage supérieur à celui de non formées de caractéristiques comparables (figure 6). Cet écart est de 7,3 pour les personnes ayant le bac ou un niveau bac+2, et de 9,3 points de pourcentage pour celles de niveau de diplôme supérieur.

Le taux de retour à l’emploi des demandeurs d’emploi primo-formés de plus de 50 ans est supérieur de 17,1 points par rapport à des demandeurs d’emploi de la même tranche d’âge et avec des caractéristiques observables proches. Ces résultats sont à interpréter avec prudence. Les plus de 50 ans ont en moyenne des taux de retour à l’emploi relativement faibles (seulement 30% au bout de 18 mois), et entrent moins souvent en formation : il est possible que les seniors qui choisissent d’entrer en formation soient plus sélectionnés sur des caractéristiques (motivation par exemple) qui ne sont pas observées dans les bases administratives utilisées ici. A l’inverse, les effets estimés pour les moins de 26 ans sont très faibles (1,9 point de pourcentage). Ce constat relativement décevant des programmes de formation pour les jeunes est déjà observé dans la littérature internationale (Card *et al.*, 2017; Barnow et Smith, 2015). Il est à pondérer par le fait que les moins de 26 ans ont un taux de retour à l’emploi très élevé

par rapport à leur aînés : sans formation, le taux de retour à l'emploi au bout de 18 mois est de 60%.

Les effets estimés sont plus élevés pour les demandeurs d'emploi de longue durée (ici inscrits depuis plus d'un an, pour lesquels l'écart entre formés et non formés à 18 mois est de 13,6 points de pourcentage) et les personnes souffrant de handicap (12,2 points de pourcentage). Ce résultat peut être rapproché des conclusions de la littérature internationale, qui observe des rendements de la formation plus élevés pour les personnes les plus éloignées de l'emploi. L'écart entre formés et non-formés pour les femmes est proche de celui estimé pour les hommes (0,3 point d'écart). Le détail des résultats par type de public est disponible en annexe C.

Cependant, comme déjà souligné, les différences selon les types de public ne signifient pas nécessairement que les formations sont effectivement plus efficaces pour ces publics. Ces écarts peuvent refléter le fait que les processus de sélection varient avec les caractéristiques de la population. Cette plus ou moins grande sélectivité peut induire des biais qui ne seront peut être pas entièrement corrigés par les méthodes d'appariement. En outre, les formations suivies peuvent être également différentes selon les publics (voir tableau D.1). Par exemple, les formations longues sont sur-représentées parmi les formations suivies par les jeunes (les formations de plus de 700 heures représentent 29,3 % des formations suivies par les moins de 26 ans, et seulement 10,6 % des formations suivies par les plus de 50 ans), ce qui peut expliquer en partie que les rendements apparents des formations suivies par les jeunes soient plus faibles¹⁷.

4.5 Effet sur la création d'entreprise

L'appariement avec la DSN permet de repérer les retours à l'emploi salarié privé, mais pas les sorties du chômage vers la fonction publique ou la création d'entreprise. Omettre ces sorties peut conduire à sous-estimer l'effet de la formation sur le retour à l'emploi. Il est cependant possible d'approcher au moins les sorties vers la création d'entreprise. En effet, la base du Fichier Historique (FH) des demandeurs d'emploi permet de repérer partiellement les chômeurs commençant une activité d'auto-entrepreneur par le motif d'annulation de leur demande d'emploi.

L'effet de la formation a été estimé en considérant cette fois comme variable d'intérêt la "création d'entreprise" à partir de cette auto-déclaration. En outre, les estimations sont menées en considérant l'ensemble des formations, mais également en singularisant les formations d'aide à la création d'entreprise, qui représentent 4,6 % des entrées en formation sur la période 2017-2019. Le détails des résultats de ces estimations sont disponibles en annexe E.

Les estimations suggèrent que les demandeurs d'emploi primo-formés ont en moyenne une probabilité de 2,5 points de pourcentage supérieure à celle d'un non-formé au profil similaire

17. A l'inverse, les différences des rendements apparents des formations peuvent s'expliquer par le fait qu'elles ne ciblent pas les mêmes publics : ainsi, comme discuté plus haut, une sélection à l'emploi est effectuée pour les formations de préparation à l'embauche.

de créer une entreprise dans les 18 mois suivant sa formation (voir table E.1). Cela s’explique essentiellement par l’effet des formations d’aide à la création d’entreprise, les autres formations n’ont quasiment aucun effet sur la probabilité de créer une entreprise. Le choix de suivre ce type de formation est effectivement lié à un projet personnel, qu’il est donc naturel de voir aboutir. A noter cependant que seulement 38,2 % des primo-formés entrés dans ce type de formation ont déclaré avoir créé une entreprise dans les 18 mois suivants. C’est 34 points de plus que des demandeurs d’emploi comparables, ce qui suggère que cette formation est suivie dans la plupart des cas par des personnes souhaitant devenir entrepreneur. Néanmoins, la majorité de ceux qui la suivent ne crée finalement pas d’entreprise ensuite.

5 Sensibilité des résultats à la méthode choisie

Pour s’assurer de la robustesse des résultats présentés jusqu’ici, il est nécessaire d’évaluer leur sensibilité aux spécifications du modèle adopté. Plusieurs tests de robustesse sont proposés. En premier lieu, on compare la méthode d’estimation de la précision obtenue avec les résultats obtenus par *bootstrap*. En second lieu, on compare les résultats obtenus en conservant une méthode d’appariement sur les plus proches voisins, mais en modifiant certains des paramètres utilisés : le nombre de “plus proches voisins” du groupe de contrôle appariés avec chaque traité, le choix de ces voisins avec ou sans remplacement, le choix de la spécification pour l’estimation du score de propension (probit plutôt que logit), le fait d’ajouter des variables de contrôle dans l’estimation linéaire de l’effet du traitement après appariement et la distance utilisée (distance de Mahalanobis, *prognostic score*). Enfin, les résultats sont comparés en utilisant des méthodes alternatives d’appariement (en appariant sur les strates de score de propension plutôt que sur les plus proches voisins). On présente en annexe un test complémentaire en ajoutant, dans les variables de conditionnement, l’historique d’emploi mais qui n’est disponible que pour les cohortes les plus récentes.

5.1 Estimation de la précision par *bootstrap*

Les estimateurs de la précision sont obtenus en utilisant des régressions linéaires locales, clusterisées au niveau des appariements (la personne formée et les plus proches voisins retenus). Une méthode alternative est d’estimer les écarts-types par *bootstrap* par bloc en rééchantillonnant chacun de ces appariements (plutôt que les personnes comme dans un *bootstrap* identique), voir Abadie et Spiess (2021). L’une et l’autre de ces méthodes sont cependant en principe adaptées lorsqu’on utilise un *matching* exact, mais pas pour une méthode de *matching sur score de propension* car l’aléa issu de l’étape d’estimation du score de propension n’est pas intégré au calcul des écarts-types, ce qui peut conduire à estimer des écarts-types biaisés (Abadie et Imbens, 2016). Une méthode alternative consiste à utiliser une méthode de *bootstrap* en rééchantillonnant en amont du calcul du score de propension. Compte tenu des contraintes computationnelles, il est très long de mettre en place ces deux méthodes sur l’ensemble de l’échantillon. Comme analyse de robustesse, on a cependant estimé des écarts-types par *bootstrap* en se concentrant sur une petite partie de l’échantillon (les personnes entrées en formation au mois de janvier 2017). En pratique, les trois méthodes (cluster au niveau de chaque appariement, *bootstrap* par blocs et *bootstrap* complet) donnent des résultats très

proches (voir table F.1).

5.2 Variation des paramètres pour l'appariement sur plus proches voisins

L'appariement sur plus proche voisin est très coûteux en temps d'estimation : pour cette raison, estimer plusieurs modifications de spécification s'avère très complexe sur l'ensemble de l'échantillon. Sur cette sous-section, les estimations sont menées uniquement sur une partie de l'échantillon (les personnes entrées en formation au mois de janvier 2017). Aussi, les rendements estimés de la formation ne sont pas directement comparables avec ceux obtenus sur l'ensemble de l'échantillon : non seulement les volumes, mais également les types de formations suivies varient au cours de l'année. La synthèse des résultats est présentée dans le tableau 7.

Nombre de voisins optimaux

Le choix d'apparier chaque traité avec trois non-traités résulte d'un arbitrage entre précision des résultats et faisabilité, compte tenu de la taille de l'échantillon (voir discussion en section 3). Cependant, ce nombre est assez faible. Des méthodes de *machine learning* peuvent être utilisées pour déterminer le nombre optimal de "plus proches voisins" nécessaires pour prédire au mieux nos variables d'intérêt. En pratique, il s'agit d'entraîner un modèle de k plus proches voisins (*K Nearest Neighbour (KNN)*) sur le groupe des non-traités¹⁸ pour prédire la variable d'intérêt en l'absence de traitement (le Y_0). Le nombre optimal est déterminé par validation croisée (10 découpages successifs de l'échantillon en échantillon d'entraînement et en échantillon de test, le tout répété 3 fois). D'après cette méthode, le nombre optimal de voisins pour prédire le retour à l'emploi tout type, à horizon 18 mois, à partir de nos variables observables, serait de 23.

Le tableau 7 montre les résultats obtenus en appariant chaque traité avec 23 non-traités, plutôt qu'avec les 3 plus proches voisins retenus dans la méthode principale. Les estimateurs apparaissent proches, mais un peu plus élevés en utilisant un nombre plus important de voisins. Ainsi, le taux de retour à l'emploi au bout de 18 mois est de 9,1 points de pourcentage supérieur pour les personnes entrées en formation en janvier 2017 par rapport à des personnes qui n'auraient pas été formées, en appariant sur les 3 plus proches voisins. En utilisant les 23 plus proches voisins pour l'appariement, l'effet estimé à 18 mois serait de 0,5 point de pourcentage supérieur. Les écarts sont un peu plus élevés lorsqu'on regarde un horizon de 24 mois (on passe d'un effet estimé de 9,8 points de pourcentage à 10,6 points de pourcentage). Cependant, d'après les estimations obtenues pour les écarts-types, et malgré la taille des échantillons concernés, il n'est pas possible de dire que ces deux estimateurs sont significativement différents au seuil de 5%¹⁹.

Appariement sur plus proches voisins avec remplacement

Jusqu'ici, l'appariement a été réalisé sans remplacement, c'est-à-dire qu'une personne du

18. Ces estimations sont faites en utilisant le package "caret" du logiciel R.

19. Dit autrement, les intervalles de confiance à 95% des deux estimateurs se recouvrent.

groupe contrôle ne pouvait être appariée qu'avec une seule personne du groupe traité. Cette méthode est la plus recommandée car cela facilite le calcul d'écart-types valides (voir Abadie et Imbens, 2008). Pour procéder à un *matching* sans remplacement, il faut que le groupe de contrôle soit largement plus grand que le groupe traité, ce qui est bien le cas ici (voir tableau 3). Cependant, il existe un inconvénient au *matching* sans remplacement : l'ordre d'appariement a une influence sur l'estimation, car la première personne traitée peut être appariée avec n'importe quelle personne du groupe de contrôle tandis que les suivantes sont limitées par les appariements précédents. En pratique, pour limiter ce problème, l'appariement est effectué en ordre décroissant du score de propension, pour que ceux ayant le score de propension le plus élevé (qui sont aussi ceux les moins susceptibles d'avoir des proches voisins non-traités) soient appariés les premiers. Les résultats du modèle avec appariement avec remplacement sont présentés dans le tableau 7. Les estimations obtenues avec ou sans remplacement sont très proches, les différences étant inférieures à 0,1 point de pourcentage. Les écart-types des estimateurs avec remplacement sont cependant un peu plus élevés que ceux de la spécification principale, car une même unité contrôle peut être utilisée plusieurs fois, si bien que moins d'information est mobilisée pour réaliser l'estimation.

TABLE 7 – Sensibilité des estimations de l'ATT à différentes spécifications du modèle (janvier 2017).

	spécif. principale	IC à 95%	23 voisins	Remplacement	Covariables	Probit	Mahalanobis	Prognostic score
Tout type d'emploi								
6 mois	-0,46* (0,241)	[-0,93; 0,02]	-0,45** (0,222)	-0,50** (0,250)	0,30 (0,237)	-0,54** (0,241)	-0,63*** (0,239)	-0,47** (0,238)
12 mois	5,10*** (0,265)	[4,59; 5,62]	5,44*** (0,245)	5,06*** (0,273)	5,95*** (0,260)	5*** (0,265)	5,08*** (0,261)	5,31*** (0,260)
18 mois	9,05*** (0,260)	[8,54; 9,56]	9,64*** (0,240)	8,99*** (0,268)	9,91*** (0,255)	8,98*** (0,260)	9,25*** (0,257)	9,29*** (0,256)
24 mois	9,80*** (0,250)	[9,31; 10,29]	10,61*** (0,230)	9,77*** (0,259)	10,62*** (0,245)	9,78*** (0,250)	10,17*** (0,248)	10,16*** (0,246)
Emploi durable								
6 mois	2,65*** (0,209)	[2,24; 3,06]	2,90*** (0,193)	2,57*** (0,216)	2,93*** (0,206)	2,64*** (0,208)	2,77*** (0,206)	2,84*** (0,206)
12 mois	5,99*** (0,249)	[5,50; 6,48]	6,50*** (0,231)	5,88*** (0,258)	6,39*** (0,245)	5,90*** (0,249)	6,08*** (0,246)	6,29*** (0,245)
18 mois	8,89*** (0,261)	[8,38; 9,41]	9,58*** (0,242)	8,82*** (0,270)	9,38*** (0,257)	8,83*** (0,261)	9,12*** (0,260)	9,23*** (0,258)
24 mois	9,70*** (0,263)	[9,18; 10,22]	10,51*** (0,243)	9,67*** (0,272)	10,22*** (0,259)	9,62*** (0,263)	9,87*** (0,261)	10,10*** (0,260)

Note : *p<0,1 ; **p<0.05 ; ***p<0.01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

L'ATT "avec covariables" est le résultat de la régression post-appariement de la variable d'intérêt sur la variable de traitement et sur l'ensemble des covariables disponibles et des termes d'interaction entre variable de traitement et covariables.

Champ : Ensemble des inscrits à Pôle emploi en janvier 2017, France entière.

Source : Dares, dispositif ForCE.

Avec variables d'interaction dans l'estimation de la régression post-appariement

L'appariement permet en théorie de recréer des groupes de traitement similaires à ceux qui auraient existé dans le cas d'une expérimentation aléatoire. L'effet du traitement peut donc être estimé en régressant simplement la variable d'intérêt sur la variable de traitement. Cependant, si pour certaines variables l'équilibre n'est pas vérifié entre les deux groupes (dans

la mesure où l'appariement principal se fait sur le score de propension), cette estimation peut être biaisée. Pour corriger ces écarts, il est possible d'ajouter d'autres variables dans la régression linéaire post-appariement et notamment de les interagir avec la variable de traitement. Afin que le coefficient associé à la variable de traitement estime l'effet de la formation, il faut au préalable centrer l'ensemble des covariables en leur retranchant leur moyenne dans la population traitée²⁰. Ainsi, le coefficient associé à la variable de traitement s'interprète comme l'effet de la formation une fois que toutes les autres variables sont égales à leur moyenne dans le groupe traité, c'est-à-dire une fois qu'on a contrôlé pour le biais de sélection sur les variables observables.

Les résultats de l'estimation de l'effet du traitement après ajout de l'ensemble des covariables observables (voir section 3) et de leur interaction avec la variable de traitement dans la régression post-appariement sont présentés dans le tableau 7. Les estimations obtenues sont un peu supérieures à celles obtenues avec le modèle de base, mais l'écart est inférieur à 1 point de pourcentage. Là encore, ces écarts ne permettent pas de conclure à des valeurs statistiquement différentes en utilisant les seuils usuels.

Estimation du score de propension par probit

Le score de propension a été estimé par un modèle logit, c'est-à-dire en supposant que les termes d'erreur de la variable latente correspondant au fait de se former suivent une loi logistique. Ce choix est classique, mais une alternative serait d'utiliser un modèle probit, c'est-à-dire en supposant que les termes d'erreur suivent une loi gaussienne. Les résultats pour ces deux types de spécification sont très proches, avec des écarts inférieurs à 0,1 point de pourcentage (voir tableau 7).

Distances alternatives au score de propension

L'estimation sur les plus proches voisins se fait en utilisant une distance euclidienne sur le score de propension. D'autres distances ont été proposées dans la littérature pour équilibrer les deux groupes, et notamment :

- la distance de Mahalanobis : elle correspond à la moyenne pondérée des écarts de l'ensemble des covariables incluses dans le modèle. La pondération permet de tenir compte de la dispersion relative de chaque covariable, afin de normaliser les écarts. Formellement, en notant X^k la valeur de la variable k et σ^k son écart-type dans la population, la distance de Mahalanobis entre deux individus i et j , notée $d^M(i, j)$, s'écrit :

$$d^M(i, j) = \sqrt{\sum_k \left(\frac{X_i^k - X_j^k}{\sigma^k} \right)^2}$$

Comme le score de propension, cette distance permet de réduire à une seule dimension l'ensemble des distances aux variables. Cependant, elle est plus susceptible de réaliser l'équilibre entre les deux groupes de traitement que le score de propension car elle

20. Si l'on ne centre pas les covariables, le coefficient associé à la variable de traitement s'interpréterait comme l'effet du traitement une fois que l'ensemble des covariables est égal à 0, ce qui n'a pas beaucoup de sens.

reflète directement les écarts des covariables (King et Nielsen, 2019). A noter que la distance de Mahalanobis est moins pertinente lorsque de nombreuses variables catégorielles sont incluses dans le modèle. En effet, avec les variables catégorielles, l'écart entre deux valeurs est discontinu, ce qui rend l'expression ci-dessus peu interprétable car elle agrège des écarts de nature différente. Les estimateurs obtenus avec cette distance (en utilisant là aussi les 3 plus proches voisins) sont également très proches de ceux obtenus dans la spécification principale. L'écart maximum est observé pour le taux de retour à l'emploi - non nécessairement durable - à 24 mois (10,2 points de pourcentage avec la distance de Mahalanobis contre 9,8 points de pourcentage avec le score de propension), là encore une différence non significative. Pour tous les autres indicateurs, ces écarts sont inférieurs à 0,2 point de pourcentage. De fait, la spécification principale est une méthode d'appariement mixte, avec un appariement sur le score de propension mais également avec plusieurs contraintes d'appariement exact ou quasi-exact sur certaines variables (voir section 3). Comme indiqué par les tests d'équilibre (voir section 4.1), l'appariement permet de réaliser un bon équilibre sur l'ensemble des covariables.

- le *prognostic score* : le *prognostic score* attribue à chacun une probabilité censée prédire le résultat de la variable d'intérêt en fonction des variables observables (voir Hansen, 2008 pour le détail des propriétés associées au *prognostic score*). L'appariement sur le *prognostic score* permettrait d'assurer un meilleur équilibre entre les deux groupes de traitement pour les variables les plus explicatives de la variable d'intérêt (voir Stuart *et al.*, 2013), par rapport à l'appariement sur le score de propension. En effet, cette méthode apparie des personnes non pas sur leur probabilité d'être traitées mais sur leur probabilité d'atteindre une certaine valeur de la variable d'intérêt (dans notre cas la probabilité de retour à l'emploi à un certain horizon). Les appariements se font donc en priorité à partir des variables fortement explicatives de la variable d'intérêt, c'est-à-dire entre personnes qui auraient eu vraisemblablement les mêmes résultats sans le traitement. Dans le cas de l'appariement sur le score de propension, certaines variables ayant un fort pouvoir explicatif de la participation au traitement mais un faible pouvoir explicatif de la variable d'intérêt peuvent influencer l'équilibre plus que nécessaire. En pratique, il s'agit donc dans une première étape d'estimer le *prognostic score* en fonction des variables observables selon un modèle binaire. Comme dans les méthodes de machine learning citées plus haut, il s'agit de prédire la valeur de la variable d'intérêt en l'absence de traitement, et l'estimation se fait donc uniquement sur le groupe de contrôle. Une fois les coefficients du modèle estimés (par un modèle logit par exemple), un *prognostic score* est calculé pour chaque personne (traitée et non-traitée) en fonction de leurs caractéristiques observables. Enfin, un appariement similaire à l'appariement sur score de propension est réalisé, mais cette fois en appariant sur le *prognostic score*.

Il est donc nécessaire d'estimer un *prognostic score* pour chaque variable d'intérêt. Pour simplifier, il n'a été estimé ici que le retour à l'emploi tout type à horizon 18 mois et le retour à l'emploi durable à horizon 18 mois. Les effets sur les autres variables d'intérêt sont calculés en appariant selon le *prognostic score* correspondant au type d'emploi concerné. Là encore, les estimations obtenues avec le *prognostic score* sont très proches de celles obtenues par le score de propension (tableau 7) : ainsi,

l'estimation de l'effet de la formation sur le taux de retour à l'emploi à 18 mois est de 9,3 points de pourcentage en utilisant le *prognostic score*, ce qui est supérieur de 0,2 point de pourcentage (différence non significative) par rapport à la valeur estimée par le score de propension. A noter que comme suggéré par Hansen (2008), les estimateurs obtenus en appariant sur le *prognostic score* peuvent être biaisés, notamment parce que l'estimation du score n'est réalisée que sur le groupe non-traité.

5.3 Méthodes d'appariement alternatives aux plus proches voisins

La méthode d'appariement sur les plus proches voisins a l'avantage d'être intuitive. Dans le cas présent, où les voisins sont nombreux, il est possible d'avoir des voisins "proches" et donc d'estimer sans biais le résultat. Cependant, l'appariement par les plus proches voisins a l'inconvénient de ne pas utiliser l'ensemble de l'information disponible dans l'échantillon puisque seule une petite proportion²¹ du groupe non-traité est prise en compte dans l'estimation. Si cela permet de réduire le biais de l'estimation en ne sélectionnant que des individus ayant un profil comparable aux traités, la perte d'information qui en résulte peut faire augmenter la variance des estimateurs et rendre les résultats moins robustes.

Une solution alternative est d'utiliser un appariement par strates²². Le principe consiste à découper l'ensemble de l'échantillon en strates, définies par des intervalles du score de propension. A l'intérieur de ces strates, les personnes traitées et non traitées ont des caractéristiques proches (du fait des propriétés équilibrantes du score de propension, voir section 3). L'effet du traitement est estimé dans chacune de ces strates : ces estimations sont ensuite agrégées en fonction du poids de chaque strate dans la population traitée. Cette méthode permet de tester la robustesse de notre modèle de base car le score de propension n'est utilisé que pour former des strates et non pour appairer des individus entre eux, rendant sa bonne estimation moins importante. Pour obtenir l'ATT sur l'ensemble des strates, deux méthodes sont possibles et donnent des résultats équivalents : agréger les estimateurs obtenus dans chaque strate, ou directement estimer une unique régression linéaire pondérée par le poids que représente la strate de l'individu dans la population des traités (voir Hong, 2010). L'un des avantages de cet estimateur par strate est d'être beaucoup moins coûteux en temps de calcul que l'estimateur par plus proches voisins. Cependant, cet estimateur a aussi plus de risque d'être biaisé : les personnes formées et non formées sont comparées sur un intervalle et non une valeur précise du score de propension. Il y a donc plus de risque de comparer des personnes avec des valeurs différentes sur certaines variables, d'autant plus qu'on n'impose plus, contrairement à la spécification principale, un appariement exact sur certaines dimensions.

Les résultats de cette estimation sur l'ensemble de la période avec un découpage en 20 strates sont présentés dans le tableau 8. Ils sont un peu plus élevés que les estimations obtenues avec l'appariement sur les plus proches voisins, les écarts pouvant aller jusqu'à un point de

21. Environ 30% dans notre cas (voir tableau 3).

22. Une méthode classique est d'utiliser des méthodes à noyau. Le principe des méthodes à noyau consiste à utiliser pour chaque traité l'ensemble des non-traités en les pondérant par leur distance (transformée par une fonction d'une certaine forme, un "noyau") avec le traité. Les contraintes opérationnelles liées à la trop grande taille du groupe de contrôle rendent cependant difficile l'application de cette méthode ici.

pourcentage. Là encore, ces écarts ne sont pas significatifs statistiquement - et par ailleurs, comme indiqué, ces écarts pourraient provenir du fait qu'il demeure au sein des strates des écarts sur certaines dimensions. Une solution pourrait être de faire des régressions au niveau de chacune des strates, pour corriger des différences de composition sur ces variables.

A noter que l'appariement par strates permet en théorie d'estimer un paramètre plus général que celui estimé ici : l'effet moyen du traitement sur l'ensemble de la population (*Average Treatment Effect* ou ATE), et non seulement l'effet moyen sur les traités. Dit autrement, il s'agit d'estimer l'effet de la formation non seulement sur les personnes qui ont été formées, mais également sur l'effet qu'aurait cette formation sur les personnes qui ne l'ont pas été²³. Cependant, cette extrapolation repose sur une hypothèse encore plus forte que l'estimateur de l'effet sur les traités, et les résultats doivent donc être pris avec beaucoup de précaution.

TABLE 8 – Estimations de l'ATT avec un appariement par strates (2017-2019)

	spécif. principale	IC à 95 %	Appariement par strates
Tout type d'emploi			
6 mois	0,57** (0,270)	[0, 04; 1, 10]	1,41*** (0,229)
12 mois	5,72*** (0,293)	[5, 15; 6, 30]	6,72*** (0,252)
18 mois	8,75*** (0,291)	[8, 18; 9, 32]	9,78*** (0,257)
24 mois	9,20*** (0,286)	[8, 64; 9, 76]	10,29*** (0,260)
Emploi durable			
6 mois	3,95*** (0,237)	[3, 49; 4, 42]	4,37*** (0,189)
12 mois	6,76*** (0,276)	[6, 22; 7, 31]	7,35*** (0,227)
18 mois	9,08*** (0,292)	[8, 51; 9, 65]	9,77*** (0,247)
24 mois	9,49*** (0,300)	[8, 90; 10, 07]	10,28*** (0,260)

Note : * p<0,1 ; ** p<0,05 ; *** p<0,01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Champ : Ensemble des inscrits à Pôle emploi entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

23. Formellement, il faut supposer que $\forall X_i, (Y(0)_i, Y(1)_i) D_i | X_i$.

6 Conclusion

Le dispositif ForCE permet de retracer les trajectoires professionnelles des personnes ayant été en contact avec le service public de l'emploi entre 2017 et 2019. Il permet donc un suivi inédit de ces personnes, et en particulier d'étudier le devenir de celles entrées en formation. En mobilisant des méthodes d'appariement, cette étude conclut à des écarts positifs dans les chances de retour à l'emploi pour les demandeurs d'emploi entrés en formation pour la première fois, par comparaison avec des personnes non formées, dès 6 mois après le début de la formation. Cet effet est de 8 points de pourcentage 12 mois après l'entrée en formation, pour atteindre 60% de taux de retour à l'emploi pour les demandeurs d'emploi primo-formés. Cet écart, suggérant un effet positif de la formation, semble persistant sur l'horizon considéré (36 mois), en se stabilisant autour de 9 points de pourcentage. Les emplois occupés semblent par ailleurs de meilleure qualité : lorsqu'on se concentre sur l'emploi durable (CDI ou emploi d'une durée d'au moins six mois), les écarts entre formés et non formés s'observent dès deux mois après l'entrée en formation, et sont même supérieurs.

En cohérence avec les résultats observés dans la littérature internationale (voir la méta-analyse de Card *et al.*, 2017), les écarts entre formés et non formés sont les plus élevés pour les personnes les plus éloignées de l'emploi, notamment les demandeurs d'emploi de longue durée (*i.e.* inscrits à Pôle emploi depuis plus d'un an) et les moins qualifiés (niveau de diplôme inférieur au baccalauréat), tandis qu'ils sont plus faibles pour les plus jeunes.

Cependant, ces résultats sont à interpréter avec prudence. Le choix de suivre une formation (ou de se voir prescrire une entrée en formation) est lié à des caractéristiques individuelles qui expliquent en elles-même le retour à l'emploi. En l'absence de méthodes expérimentales, l'estimation de l'effet de la formation est potentiellement biaisée par ces effets de sélection. Même si les méthodes d'appariement utilisées permettent de réduire une partie du biais de sélection, des facteurs individuels qui ne sont pas mesurés dans les données administratives disponibles peuvent affecter ces estimations. C'est particulièrement flagrant pour l'estimation des effets des formations préalables à l'emploi, qui sont en général liées à une pré-embauche. Celle-ci peut se faire à partir de caractéristiques observables en entretien (comme la motivation ou la maîtrise de savoirs-être professionnels) qui ne sont pas disponibles dans les bases administratives. À noter que le sens de ce biais de sélection n'est pas univoque : les prescripteurs peuvent également cibler en priorité pour les formations (et c'était en principe l'un des objectifs du Plan d'investissement dans les compétences) les demandeurs d'emploi les moins employables, ce qui aurait tendance à biaiser les estimations de l'effet de la formation à la baisse.

Plusieurs pistes d'approfondissement peuvent être envisagées. En termes opérationnels, la méthode d'appariement sur les plus proches voisins est très coûteuse en temps de calcul. Les tests de robustesses suggèrent qu'une méthode d'appariement par strate, qui ne demande que l'estimation du score de propension, puis un nombre limité de régressions linéaires, pourraient fournir des résultats très proches. Ces résultats sont cependant à confirmer, notamment en vérifiant que cette stabilité s'observe également en décomposant les estimations par public et par type de formations. Par ailleurs, il serait possible dans le futur d'enrichir la spécification. Si les données disponibles aujourd'hui ne fournissent pas un recul important sur l'histoire

professionnelle passée (les données sur les contrats de travail ne sont disponibles que depuis 2017, à l'inverse de l'historique d'inscription à Pôle emploi qui court sur dix ans), avec le temps il sera possible de disposer d'un recul plus important. Biewen *et al.* (2012) insistent sur l'importance de la prise en compte de ces variables pour la qualité des estimateurs. Disposer d'un plus long recul serait intéressant pour évaluer si les écarts entre formés et non formés se maintiennent sur le futur, en cohérence avec l'idée que les formations permettent un accroissement du capital humain. À ce titre, observer les salaires concernés permettrait d'estimer les effets des formations au moins sur les revenus des personnes concernées, voire sur leur productivité, afin d'apporter des éléments permettant d'évaluer de manière plus complète les conséquences du Plan d'investissement dans les compétences.

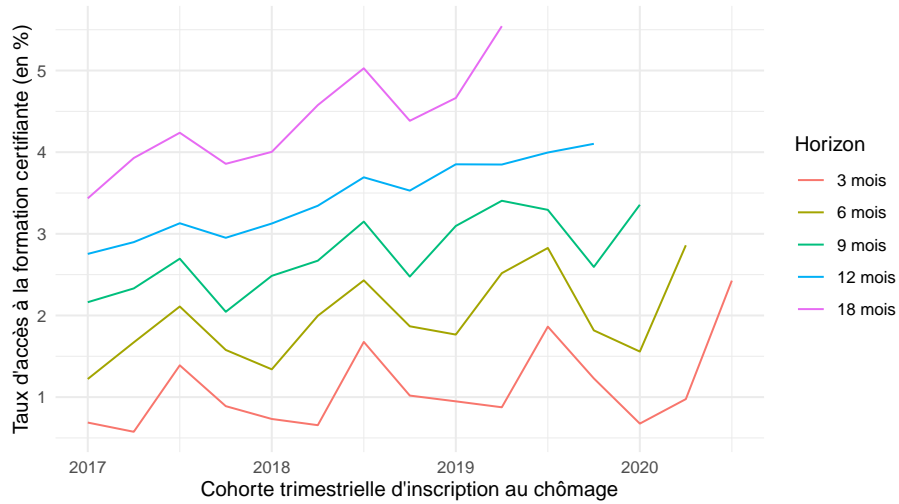
Références

- ABADIE, A. et IMBENS, G. W. (2008). On the failure of the bootstrap for matching estimators. *Econometrica*, 76(6):1537–1557.
- ABADIE, A. et IMBENS, G. W. (2016). Matching on the estimated propensity score. *Econometrica*, 84(2):781–807.
- ABADIE, A. et SPIESS, J. (2021). Robust post-matching inference. *Journal of the American Statistical Association*.
- BARNOW, B. S. et SMITH, J. (2015). Employment and training programs. NBER Working Papers 21659, National Bureau of Economic Research, Inc.
- BIEWEN, M., FITZENBERGER, B., OSIKOMINU, A. et PAUL, M. (2012). The effectiveness of public sponsored training revisited : The importance of data and methodological choices. *IZA Working Paper No. 1209*.
- BLACHE, G. (2015). La formation des demandeurs d’emploi : quels effets sur l’accès à l’emploi? *Études et Recherches, Pôle Emploi*, 3.
- CARD, D., KLUVE, J. et WEBER, A. (2017). What Works? A Meta Analysis of Recent Active Labor Market Program Evaluations. *Journal of the European Economic Association*, 16(3):894–931.
- CAVACO, S., FOUGÈRE, D. et POUGET, J. (2013). Estimating the effect of a retraining program on the re-employment rate of displaced workers. *Empirical Economics*, 44(1):261–287.
- CRÉPON, B., DUFLO, E., GURGAND, M., RATHELOT, R. et ZAMORA, P. (2013). Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment. *The Quarterly Journal of Economics*, 128(2):531–580.
- CRÉPON, B., FERRACCI, M. et FOUGÈRE, D. (2012). Training the unemployed in France : how does it affect unemployment duration and recurrence? *Annals of Economics and Statistics*, 107/108:175–199.
- EL BOUHAIRI, Y., GADOUCHE, K. et MARQUIER, R. (2022). Appariements sécurisés de données personnelles : un exemple pour la recherche. *Journées de méthodologie statistique de l’Insee*, (14ème édition).
- HANSEN, B. (2008). The prognostic analogue of the propensity score. *Biometrika*, 95(2):481–488.
- HO, D. E., IMAI, K., KING, G. et STUART, E. A. (2011). MatchIt : Nonparametric preprocessing for parametric causal inference. *Journal of Statistical Software*, 42(8):1–28.
- HONG, G. (2010). Marginal mean weighting through stratification : Adjustment for selection bias in multilevel data. *Journal of Educational and Behavioral Statistics*, 35(5):499–531.
- IMBENS, G. W. et WOOLDRIDGE, J. M. (2009). Recent developments in the econometrics of program evaluation. *Journal of Economic Literature*, 47(1):5–86.

- KING, G. et NIELSEN, R. (2019). Why propensity scores should not be used for matching. *Political Analysis*, 27(4):435–454.
- OSIKOMINU, A. (2012). Quick Job Entry or Long-Term Human Capital Development? The Dynamic Effects of Alternative Training Schemes. *The Review of Economic Studies*, 80(1): 313–342.
- ROSENBAUM, P. R. et RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–55.
- SIANESI, B. (2004). An evaluation of the Swedish system of active labor market programs in the 1990s. *Review of Economics and Statistics*, 86:133–155.
- STUART, E. (2010). Matching methods for causal inference : A review and a look forward. *Statistical science : a review journal of the Institute of Mathematical Statistics*, 25(1):1–21.
- STUART, E. A., LEE, B. K. et LEACY, F. P. (2013). Prognostic score-based balance measures can be a useful diagnostic for propensity score methods in comparative effectiveness research. *Journal of Clinical Epidemiology*, 66:84–90.

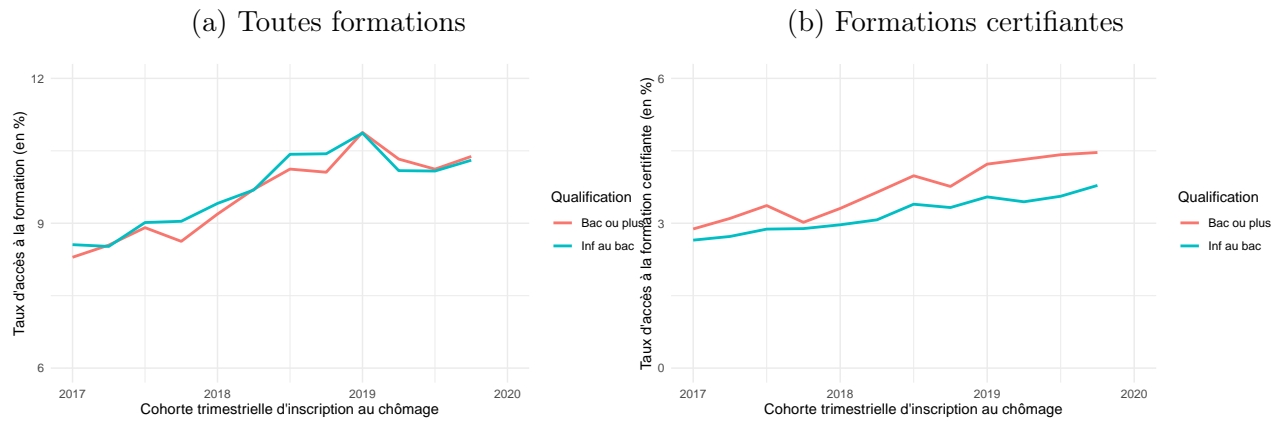
A Taux d'accès à la formation

FIGURE A.1 – Taux d'accès à la formation certifiante



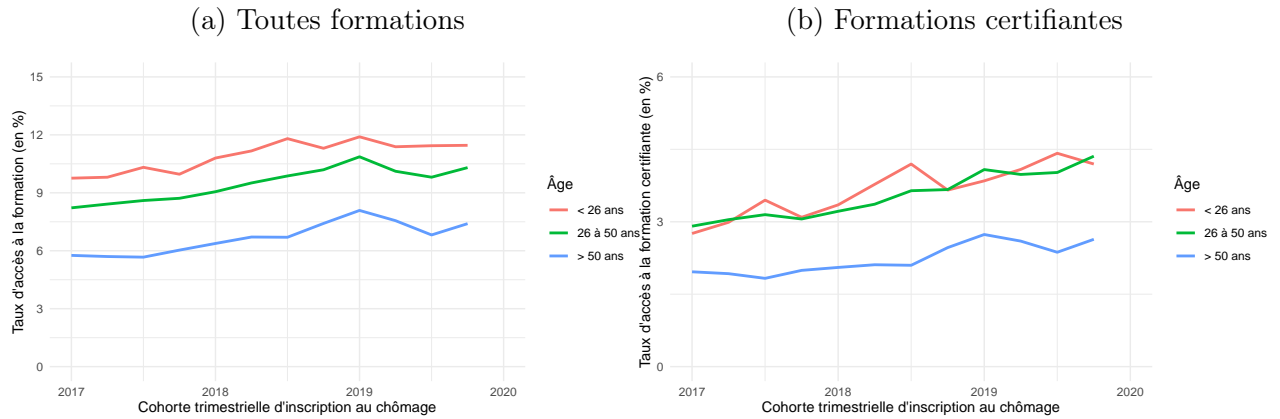
Champ : Ensemble des inscrits à Pôle emploi entre janvier 2017 et décembre 2020, France entière.
Source : Dares, dispositif ForCE.

FIGURE A.2 – Taux d'accès à la formation (horizon 12 mois) selon le niveau de diplôme



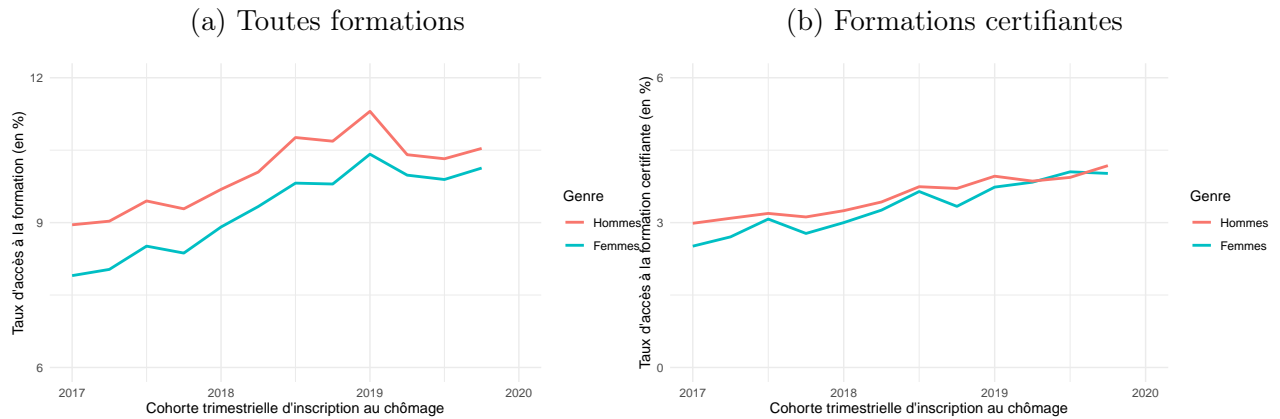
Champ : Ensemble des inscrits à Pôle emploi entre janvier 2017 et décembre 2020, France entière.
Source : Dares, dispositif ForCE.

FIGURE A.3 – Taux d'accès à la formation (horizon 12 mois) selon l'âge



Champ : Ensemble des inscrits à Pôle emploi entre janvier 2017 et décembre 2020, France entière.
Source : Dares, dispositif ForCE.

FIGURE A.4 – Taux d'accès à la formation (horizon 12 mois) selon le sexe



Champ : Ensemble des inscrits à Pôle emploi entre janvier 2017 et décembre 2020, France entière.
Source : Dares, dispositif ForCE.

B Évaluation du modèle

B.1 Résultats de l'estimation du score de propension (modèle logit avec effets fixes)

Nous présentons ci-dessous les résultats de l'estimation du modèle logit pour le calcul du score de propension. La variable dépendante est donc l'indicatrice d'entrée en première formation sur la période entre janvier 2017 et septembre 2019. Le modèle estimé est un modèle logit avec effets fixes pour l'agence Pôle emploi.

TABLE B.1 – Estimation du score de propension par un modèle logit

	<i>Coefficient logit</i>	<i>Effet marginal (en points)</i>
Age (en année)	0,0601 (0,0040)***	0,22
Age au carré	-0,0010 (0,0000)***	-0,004
Sexe (ref. = Hommes)	-0,0353 (0,0147)***	-0,01
Ancienneté (en jours)	-0,0003 (0,00001)***	-0,001
Durée cumulée d'inscription à Pôle emploi avant l'épisode (en jours)	-0,0002 (0,00001)***	-0,001
Nombre d'épisodes de chômage	-0,0552 (0,0058)***	-0,22
Nombre d'heures d'activité partielle	-0,0003 (0,00001)***	-0,001
Temps passé en maladie (en jours)	-0,0016 (0,0006)***	-0,01
Temps passé en maladie avant l'épisode	0,0004 (0,0003)	
QPV/ZUS	0,0039 (0,0192)	
ZRR	-0,1347 (0,0348)***	-0,48
Handicap	0,3496 (0,0233)***	1,39
Expérience (en années)	-0,0059 (0,0010)***	-0,03
Nombre d'absences	0,0293 (0,0057)***	0,11
Mobilité longue distance	0,2152 (0,0143)***	0,79
Diplôme (ref. = Bac)		
Pré-bac	-0,1358 (0,0145)***	-0,62
Bac + 2	-0,1196 (0,0186)***	-0,45
Licence	-0,2339 (0,0236)***	-0,82
Bac + 5	-0,4072 (0,0296)***	-1,55
Indicatrice d'obtention du diplôme	0,0928 (0,0121)***	0,34
Qualification (ref. = Cadre)		
Ouvrier	0,0690 (0,0333)**	0,26
Employé non qualifié	0,1928 (0,0321)***	0,77
Employé qualifié	0,1097 (0,0299)***	0,43
Technicien	0,0897 (0,0336)***	0,35
Inconnue	-0,2392 (0,0558)***	-0,76
Situation familiale (ref. = Célibataire)		
Marié	0,0003 (0,0146)	
Divorcé	0,0566 (0,0222)***	0,20
Veuf	-0,1347 (0,0754)	
Nombre d'enfants (ref. = 0)		
1 enfant	0,0010 (0,0167)	
2 enfants	0,0400 (0,0185)**	0,18
3 enfants	-0,0070 (0,0247)	
4 enfants ou plus	-0,0986 (0,0358)***	-0,34
Catégorie d'inscription (ref. = Catégorie 1)		
Catégorie 2	-0,4435 (0,0226)***	-1,64
Catégorie 3	0,5125 (0,2589)**	1,87
Catégorie 4	1,4113 (0,0451)***	5,54
Contrat recherché (ref. = CDD)		
CDI	0,5713 (0,2584)**	1,87
Saisonnier	-1,4284 (0,1001)***	-5,49
Motif d'inscription (ref. = Autre)		
Fin de contrat/activité	-0,1019 (0,0153)***	-0,39
Licenciement	0,1112 (0,0165)***	0,45
Première recherche d'emploi	0,1360 (0,0195)***	0,40
Nationalité (ref. = Afrique subsaharienne)		
France	-0,4460 (0,0350)***	-1,66
UE proche	-0,5176 (0,0485)***	-1,99
Autre Europe	-0,6398 (0,0595)***	-2,42
Maghreb	-0,4458 (0,0432)***	-1,71
Asie	-0,2765 (0,0659)***	-1,04
Autre	-0,1641 (0,0827)***	-0,56
ROME (ref. = Agriculture)		
Arts	0,3791 (0,0686)***	1,37
Banque et assurance	-0,0452 (0,0608)	
BTP	0,2043 (0,0366)***	0,65
Communication et médias	0,1567 (0,0507)***	0,49
Commerce et vente	0,1733 (0,0352)***	0,57
Industrie	0,1932 (0,0377)***	0,64
Loisirs et tourisme 3	0,1325 (0,0372)***	0,50
Maintenance et installation	0,3221 (0,0399)***	1,10
Santé	0,2492 (0,0448)***	1,33
Services à la personne	0,0006 (0,0356)	
Spectacle	-0,1612 (0,0634)**	-0,70
Support entreprise	0,2396 (0,0369)***	0,81
Transport et logistique	0,4382 (0,0358)***	1,51
Non-renseigné	0,4847 (0,1182)***	1,77

Note : *p<0,1; **p<0,05; ***p<0,01. Les effets marginaux ne sont reportés que pour les variables significatives.

Champ : Ensemble des demandeurs d'emploi inscrits à Pôle emploi entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

L'âge n'est pas linéaire mais quadratique, avec au début un effet positif de l'âge sur la probabilité d'entrer en formation puis la tendance s'inverse et l'effet devient négatif à partir d'une cinquantaine d'années. De même, la probabilité d'entrer en formation décroît à mesure que l'ancienneté d'inscription, le nombre d'épisodes d'inscription connus dans les 10 dernières années ou l'expérience dans le secteur d'emploi recherché augmentent. Le type d'emploi recherché est décrit par les variables "contrat recherché" et "catégorie d'inscription"²⁴. Sans surprise, les demandeurs d'emploi recherchant des emplois à temps partiel ou saisonnier ont une probabilité nettement plus faible d'entrer en formation (respectivement -5,5 points et -1,6 points), leur projet professionnel pouvant être jugé moins propice à un investissement en capital humain. Le fait que la catégorie 4 soit associée à une forte hausse de la probabilité d'entrer en formation ne peut en revanche pas être interprété en tant que tel car il peut exister des cas où les demandeurs d'emploi s'inscrivent à Pôle emploi (en catégorie 4) en même temps que leur entrée en formation, ce qui explique cette corrélation positive. On peut également remarquer que la probabilité d'entrer en formation est, toutes choses égales par ailleurs, maximale pour les titulaires d'un diplôme de niveau baccalauréat. Les chômeurs de niveau de diplôme inférieur ou supérieur au baccalauréat ont moins de chance d'entrer en formation que les chômeurs de ce niveau intermédiaire. Les chômeurs de nationalité d'un pays d'Afrique subsaharienne ou d'Asie ont, toutes choses égales par ailleurs, une probabilité plus élevée que les chômeurs français d'entrer en formation, ce qui s'explique par le nombre non-négligeable de formations pour apprendre le français (environ 13 500 sur la période). Enfin, les secteurs d'emploi recherché pour lesquels la probabilité d'entrer en formation est la plus forte sont les secteurs du transport et de la logistique, des arts, de la santé et de la maintenance et de l'installation.

24. Ces catégories ne recoupent pas exactement les catégories traditionnelles A, B, C, D et E : la catégorie 1 désigne les demandeurs d'emploi immédiatement disponibles à la recherche d'un CDI à temps plein ; la catégorie 2 désigne les demandeurs d'emploi immédiatement disponibles à la recherche d'un CDI à temps partiel ; la catégorie 3 désigne les demandeurs d'emploi immédiatement disponibles à la recherche d'un emploi à durée limitée (CDD, intérim, vacation) ; la catégorie 4 désigne les demandeurs d'emploi non-immédiatement disponibles car en maladie ou en formation.

B.2 Statistiques d'équilibre entre les groupes de traitement

TABLE B.2 – Statistiques d'équilibre avant et après appariement

	Pré-appariement			Post-appariement		
	Différence de moyennes standardisée	Ratio de variances	Stat de Kolmogorov-Smirnov	Différence de moyennes standardisée	Ratio de variances	Stat de Kolmogorov-Smirnov
Score de propension	0,604	2,498	0,323	0,174	1,764	0,057
Nombre d'heures d'activité partielle	-0,589	0,292	0,178	-0,103	0,787	0,084
Ancienneté d'inscription	-0,583	0,375	0,196	0,001	0,999	0,013
Age ²	-0,476	0,665	0,149	-0,001	0,940	0,018
Age	-0,443	0,802	0,149	0,002	0,981	0,018
Pré-bac × >50 ans	-0,378		0,092	-0,025		0,006
Expérience	-0,223	0,712	0,087	-0,015	0,956	0,019
Bac × <26 ans	0,163		0,053	0,009		0,009
Pré-bac × <26 ans	0,154		0,049	0,037		0,014
Temps d'inscription à Pôle emploi passé avant	-0,150	0,741	0,055	0,003	1,005	0,013
Motif d'inscription : 1ere recherche	0,145		0,050	0,009		0,006
Célibataire	0,138		0,069	0,0004		0,001
Pré-bac × sexe	-0,137		0,053	-0,019		0,007
Bac	0,132		0,060	0,022		0,011
Pré-bac	-0,104		0,051	-0,001		0,011
Bac × >50 ans	-0,103		0,014	0,015		0,002
Motif d'inscription : Fin de contrat	-0,100		0,045	-0,053		0,024
Bac + 5 × >50 ans	-0,093		0,007	0,010		0,001
Divorcé	-0,086		0,023	-0,006		0,002
Marié	-0,083		0,040	-0,003		0,002

Note : Seules les 20 variables ayant les différences de moyenne standardisées les plus élevées sont représentées. Les ratios de variance pour les variables binaires ne sont pas indiqués car ils n'ajoutent pas d'informations supplémentaires aux autres statistiques.

Champ : Ensemble des demandeurs d'emploi inscrits à Pôle emploi entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

C Résultats par type de public

TABLE C.1 – Estimation de l'effet de la formation selon le niveau de diplôme

	Inférieur au bac		Bac / Bac + 2		Supérieur à bac + 2	
	Taux de retour à l'emploi des primo-formés	ATT (en points)	Taux de retour à l'emploi des primo-formés	ATT (en points)	Taux de retour à l'emploi des primo-formés	ATT (en points)
Tout type d'emploi						
6 mois	28,4 %	2 (0,38)***	31 %	-1,7 (0,50)***	29,7 %	0,4 (0,77)
12 mois	46,7 %	7 (0,41)***	51,2 %	3,9 (0,53)***	47,7 %	5,4 (0,83)***
18 mois	57,9 %	9,6 (0,41)***	63,7 %	7,3 (0,52)***	60,1 %	9,3 (0,84)***
24 mois	64,9 %	10,4 (0,41)***	70,2 %	7,3 (0,50)***	66,3 %	9,5 (0,84)***
Emploi durable						
6 mois	18,8 %	4,7 (0,32)***	22,7 %	3,1 (0,45)***	23,6 %	3,1 (0,71)***
12 mois	31,8 %	7,9 (0,38)	37,5 %	5,6 (0,51)***	37,7 %	5,9 (0,80)***
18 mois	41,8 %	9,9 (0,41)***	49,6 %	8,2 (0,54)***	49,7 %	9,2 (0,85)***
24 mois	49,2 %	10,7 (0,42)***	56,7 %	7,9 (0,54)***	56,2 %	9,3 (0,87)***

Note : *p<0,1; **p<0,05; ***p<0,01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Champ : Ensemble des demandeurs d'emploi inscrits entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

TABLE C.2 – Estimation de l'effet de la formation selon l'âge

	< 26 ans		Entre 26 et 50 ans		> 50 ans	
	Taux de retour à l'emploi des primo-formés	ATT (en points)	Taux de retour à l'emploi des primo-formés	ATT (en points)	Taux de retour à l'emploi des primo-formés	ATT (en points)
Tout type d'emploi						
6 mois	32,7 %	-4,2 (0,53)***	29,2 %	1,7 (0,34)***	22,9 %	6,7 (0,76)***
12 mois	53,1 %	-0,7 (0,55)	48,2 %	7,4 (0,37)***	37 %	13 (0,87)***
18 mois	65,6 %	1,9 (0,53)***	60,2 %	10,4 (0,37)***	46 %	17,1 (0,91)***
24 mois	72,4 %	2 (0,51)***	67 %	10,8 (0,37)***	51,6 %	19 (0,95)***
Emploi durable						
6 mois	23,4 %	2,1 (0,47)***	20,6 %	4,4 (0,30)***	15,4 %	6,5 (0,64)***
12 mois	37,4 %	2,7 (0,54)	34,8 %	7,9 (0,35)***	25,8 %	11,2 (0,78)***
18 mois	49,1 %	4 (0,56)***	45,9 %	10,5 (0,37)***	33,5 %	14,4 (0,85)***
24 mois	56,3 %	3,5 (0,56)***	53,3 %	11,1 (0,38)***	39 %	16,3 (0,91)***

Note : *p<0,1 ; **p<0,05 ; ***p<0,01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Champ : Ensemble des demandeurs d'emploi inscrits entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

TABLE C.3 – Estimation de l'effet de la formation selon le sexe

	Hommes		Femmes	
	Taux de retour à l'emploi des primo-formés	ATT (en points)	Taux de retour à l'emploi des primo-formés	ATT (en points)
Tout type d'emploi				
6 mois	32,6 %	1,4 (0,40)***	26,3 %	-0,7 (0,38)*
12 mois	51,8 %	6,3 (0,43)***	44,7 %	4,8 (0,42)***
18 mois	63,1 %	8,5 (0,42)***	57 %	8,8 (0,43)***
24 mois	69,6 %	8,7 (0,41)***	64 %	9,5 (0,42)***
Emploi durable				
6 mois	22,3 %	5,1 (0,35)***	19,1 %	2,6 (0,34)***
12 mois	36,6 %	8 (0,41)***	32,3 %	5,3 (0,40)***
18 mois	47,1 %	9,7 (0,42)***	43,7 %	8,4 (0,42)***
24 mois	54,2 %	10 (0,43)***	50,8 %	8,9 (0,44)***

Note : *p<0,1 ; **p<0,05 ; ***p<0,01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Champ : Ensemble des demandeurs d'emploi inscrits entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

TABLE C.4 – Estimation de l'effet de la formation selon l'ancienneté d'inscription

	Ancienneté ≤ 1 an		Ancienneté > 1 an	
	Taux de retour à l'emploi des primo-formés	ATT (en points)	Taux de retour à l'emploi des primo-formés	ATT (en points)
Tout type d'emploi				
6 mois	29,8 %	-1 (0,33)***	28,8 %	4 (0,48)*
12 mois	48,8 %	3,7 (0,35)***	47,2 %	10,1 (0,52)***
18 mois	61 %	6,5 (0,35)***	58,3 %	13,6 (0,52)***
24 mois	68 %	6,8 (0,34)***	64,5 %	14,3 (0,52)***
Emploi durable				
6 mois	21,4 %	2,9 (0,29)***	19,4 %	6,3 (0,41)***
12 mois	35,4 %	5,3 (0,34)***	32,6 %	10 (0,48)***
18 mois	46,8 %	7,3 (0,36)***	42,7 %	12,8 (0,51)***
24 mois	54,2 %	7,5 (0,36)***	49,4 %	13,7 (0,53)***

Note : *p<0,1 ; **p<0,05 ; ***p<0,01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Champ : Ensemble des demandeurs d'emploi inscrits entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

TABLE C.5 – Estimation de l’effet de la formation selon l’habitat en quartier prioritaire de la ville (QPV) et selon le handicap

	En QPV		En situation de handicap	
	Taux de retour à l’emploi des primo-formés	ATT (en points)	Taux de retour à l’emploi des primo-formés	ATT (en points)
Tout type d’emploi				
6 mois	26,5 %	1,7 (0,79)**	17,6 %	2,8 (0,84)*
12 mois	45 %	7,2 (0,88)***	31,8 %	8,3 (1,01)***
18 mois	56,7 %	10,3 (0,88)***	42 %	12,2 (1,06)***
24 mois	64,3 %	11,4 (0,88)***	49,1 %	14,5 (1,09)***
Emploi durable				
6 mois	17,6 %	4,2 (0,67)***	11,9 %	3,4 (0,70)***
12 mois	30,6 %	7,6 (0,80)***	22,2 %	7,1 (0,90)***
18 mois	40,8 %	9,9 (0,87)***	30,7 %	10,1 (0,99)***
24 mois	48,4 %	11,2 (0,91)***	37,4 %	12,3 (1,05)***

Note : *p<0,1 ; **p<0,05 ; ***p<0,01.

L’emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L’emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Champ : Ensemble des demandeurs d’emploi inscrits entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

D Types de formation suivie selon les caractéristiques des primo-formés

TABLE D.1 – Répartition des types de formation selon le profil des primo-formés

	Part >12 mois	Part longues	Part certifiantes	Part adaptation de poste	Part pré-qualifiantes	Domaine principal
Ensemble	3,6 %	21,2 %	33,4 %	12 %	15,6 %	Échange et gestion (20 %)
Diplôme						
Supérieur à bac + 2	5,4 %	22,3 %	34,2 %	13 %	11,9 %	Échange et gestion (20,1 %)
Bac / Bac + 2	5,8 %	27,7 %	36,9 %	13,8 %	11,9 %	Échange et gestion (21,8 %)
Inférieur au bac	1,8 %	16,8 %	31,1 %	10,6 %	18,9 %	Échange et gestion (18,9 %)
Âge						
Plus de 50 ans	1,4 %	10,6 %	29,8 %	6,8 %	20,8 %	Général, langues et lettres (23,5 %)
Entre 26 et 50 ans	2,9 %	19,5 %	34,6 %	10,8 %	14,6 %	Échange et gestion (21,5 %)
Moins de 26 ans	6,1 %	29,3 %	32,3 %	16,8 %	15,8 %	Services aux personnes (20,7 %)
Sexe						
Hommes	2 %	17,3 %	33,6 %	12,8 %	13,4 %	Échange et gestion (22,3 %)
Femmes	5,4 %	25,3 %	33,2 %	11,2 %	18 %	Services aux personnes (23,4 %)
Inscrits depuis plus d’un an	2,6 %	19 %	33,7 %	10,9 %	15,3 %	Échange et gestion (20,5 %)
En QPV/ZUS	2,1 %	18,2 %	30,4 %	9,8 %	19,3 %	Général, langues et lettres (20,3 %)
En situation de handicap	2,4 %	16,1 %	24,2 %	5,2 %	27,8 %	Général, langues et lettres (31 %)

Champ : Ensemble des demandeurs d’emploi entrant en première formation entre 2017 et 2019.

Source : Dares, dispositif ForCE.

TABLE D.2 – Classement des domaines de formation selon le taux de retour à l’emploi à 18 mois

	<i>Part dans l’ensemble des formations</i>	<i>Taux de retour à l’emploi (18 mois)</i>
Production mécanique	3,1 %	79,1 %
Électricité, électronique	1,9 %	77,4 %
Sciences	0,3 %	74,8 %
Transformation	3,2 %	74,1 %
Fonction production	6,3 %	72,9 %
Services à la collectivité	4,7 %	71,5 %
Échange et gestion	20,4 %	70,1 %
Génie civil, construction, bois	2,9 %	66,9 %
Information, communication	8,2 %	65,8 %
Services aux personnes	15,4 %	57 %
Agriculture	1,8 %	55,2 %
Général, lettres et langues	16,4 %	47,2 %
Arts	1,9 %	47 %
Non renseigné	8,9 %	44,9 %
Sciences humaines, économie, droit	4,6 %	34,8 %

Note : En moyenne, le taux de retour à l’emploi des primo-formés à horizon 18 mois est de 60,1 %.

Champ : Ensemble des demandeurs d’emploi primo-formés entre 2017 et 2019, France entière.

Source : Dares, dispositif ForCE.

TABLE D.3 – Effet de la formation longue étalée sur moins d’un an selon le profil des formés

	ATT (en points)									
	Diplôme			Âge			Sexe		Ancienneté	
	Inf au bac	Bac / Bac + 2	Sup à bac + 2	<26 ans	26-50 ans	>50 ans	Hommes	Femmes	≤1 an	>1 an
Tout type d’emploi										
12 mois	-5,4*** (0,96)	-9,5*** (1,03)	-6,7*** (1,73)	-16,2*** (1,05)	-3,2*** (0,81)	5,6** (2,49)	-8,9*** (0,99)	-6*** (0,84)	-9,3*** (0,75)	-2,5** (1,15)
18 mois	4,9*** (1,01)	4,1*** (1,04)	7,4*** (1,83)	-4,8*** (1,08)	9,2*** (0,84)	16,7*** (2,74)	2,4** (1,01)	6,7*** (0,89)	2,6*** (0,77)	9,8*** (1,20)
24 mois	7,3*** (1,00)	5,4*** (0,98)	8,8*** (1,79)	-2,9*** (1,03)	11*** (0,82)	21*** (2,88)	4,9*** (0,98)	8,2*** (0,86)	4,3*** (0,74)	12,3*** (1,18)
Emploi durable										
12 mois	-3,4*** (0,84)	-8*** (0,93)	-6,9*** (1,61)	-12,4*** (0,94)	-2,8*** (0,73)	3,6 (2,10)	-6,2*** (0,88)	-5,7*** (0,75)	-7,5*** (0,67)	-1,8 (1,01)
18 mois	4*** (0,98)	2,8*** (1,06)	5,1*** (1,85)	-4,5*** (1,09)	7,4*** (0,84)	11,8*** (2,50)	2,2** (1,00)	4,6*** (0,88)	1,7** (0,78)	7,7*** (1,17)
24 mois	6,3*** (1,02)	4,5*** (1,06)	6,9*** (1,88)	-3,2*** (1,10)	9,7*** (0,86)	16,1*** (2,75)	4,4*** (1,04)	6,4*** (0,89)	3,5*** (0,79)	10,2*** (1,22)

Note : *p<0,1 ; **p<0,05 ; ***p<0,01.

L’emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L’emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Les formations longues recouvrent toutes les formations dont le volume horaire est supérieur à 700 heures.

Lecture : La probabilité d’avoir trouvé un emploi dans les 18 mois après l’entrée en formation longue étalée sur moins d’un an pour les titulaires d’un diplôme inférieur au baccalauréat est en moyenne de 4 points de pourcentage supérieure à celle d’un demandeur d’emploi non-formé titulaire d’un diplôme inférieur au baccalauréat et ayant par ailleurs un profil similaire. Cet effet est statistiquement significatif au seuil de 1%.

Champ : Ensemble des inscrits à Pôle emploi entre 2017 et 2019, France entière.

Source : Dares, dispositif ForCE.

TABLE D.4 – Effet de la formation certifiante étalée sur moins d’un an selon le profil des formés

	ATT (en points)									
	Diplôme			Âge			Sexe		Ancienneté	
	Inf au bac	Bac / Bac + 2	Sup à bac + 2	<26 ans	26-50 ans	>50 ans	Hommes	Femmes	≤1 an	>1 an
Tout type d’emploi										
12 mois	8,5*** (0,76)	1,4 (0,92)	0,1 (1,45)	-4,2*** (1,03)	6,5*** (0,65)	15,5*** (1,64)	6,2*** (0,76)	2,8*** (0,76)	2,3*** (0,64)	9,3*** (0,90)
18 mois	13,7*** (0,75)	9,4*** (0,90)	8,2*** (1,51)	3,1*** (1,00)	12,7*** (0,64)	21,7*** (1,71)	11,3*** (0,73)	11,1*** (0,77)	8,7*** (0,63)	16*** (0,89)
24 mois	14,5*** (0,73)	9,5*** (0,87)	8,5*** (1,49)	3,2*** (0,95)	13*** (0,62)	24,1*** (1,74)	11,8*** (0,70)	11,6*** (0,76)	9*** (0,61)	16,8*** (0,87)
Emploi durable										
12 mois	6,5*** (0,70)	0,7 (0,87)	-0,9 (1,37)	-3,5*** (0,98)	4,7*** (0,60)	11,4*** (1,46)	4,9*** (0,71)	1,5** (0,70)	1,6*** (0,60)	6,7*** (0,82)
18 mois	11,1*** (0,75)	7,7*** (0,94)	6,3*** (1,52)	2* (1,06)	10,4*** (0,65)	16,6*** (1,61)	9,3*** (0,75)	8,8*** (0,78)	7,1*** (0,65)	12,6*** (0,89)
24 mois	12,6*** (0,77)	8,4*** (0,94)	7,1*** (1,55)	2,3** (1,08)	11,6*** (0,66)	19,1*** (1,70)	10,5*** (0,77)	9,9*** (0,79)	8*** (0,66)	14*** (0,91)

Note : *p<0,1 ; **p<0,05 ; ***p<0,01.

L’emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L’emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Lecture : La probabilité d’avoir trouvé un emploi dans les 18 mois après l’entrée en formation certifiante étalée sur moins d’un an pour les titulaires d’un diplôme inférieur au baccalauréat est en moyenne de 13,7 points de pourcentage supérieure à celle d’un demandeur d’emploi non-formé titulaire d’un diplôme inférieur au baccalauréat et ayant par ailleurs un profil similaire. Cet effet est statistiquement significatif au seuil de 1%.

Champ : Ensemble des inscrits à Pôle emploi entre 2017 et 2019, France entière.

Source : Dares, dispositif ForCE.

E Effet sur la création d’entreprise

TABLE E.1 – Estimation de l’effet de la formation sur la création d’entreprise

	Toutes formations			Formations création d’entreprise	
	Taux de création d’entreprise des non-formés	Taux de création d’entreprise des primo-formés	ATT (en points)	Taux de création d’entreprise des primo-formés	ATT (en points)
6 mois	0,8 %	2,8 %	1,7 (0,09)***	29,9 %	27,9 (1,07)***
12 mois	1,3 %	3,8 %	2,1 (0,11)***	35,4 %	32,4 (1,12)***
18 mois	1,6 %	4,6 %	2,5 (0,12)***	38,2 %	34,4 (1,12)***
24 mois	1,9 %	4,9 %	2,5 (0,12)***	38,9 %	34,7 (1,14)***

Note : *p<0,1 ; **p<0,05 ; ***p<0,01.

On considère qu’il y a création d’entreprise lorsque le demandeur d’emploi annule son inscription à Pôle emploi pour motif de création d’entreprise et ne se réinscrit pas dans les 31 jours suivant cette annulation.

Lecture : La probabilité de créer une entreprise dans les 18 mois suivant l’entrée en formation d’aide à la création d’entreprise est en moyenne de 34,4 points de pourcentage supérieure à celle d’un demandeur d’emploi non formé ayant un profil similaire.

Champ : Ensemble des demandeurs d’emploi inscrits entre janvier 2017 et septembre 2019, France entière.

Source : Dares, dispositif ForCE.

F Test de robustesse

F.1 Comparaison des différentes méthodes de calcul des écarts-types

TABLE F.1 – Estimation des écarts-types selon différentes méthodes (janvier 2017)

	Cluster par match		Bootstrap par blocs		Bootstrap complet	
	ATT	IC à 95 %	ATT moyen	IC à 95 %	ATT moyen	IC à 95 %
Tout type d'emploi						
6 mois	-0,46* (0,241)	[-0,93; 0,02]	-0,44* (0,249)	[-0,92; 0,06]	-0,65*** (0,228)	[-1,07; -0,19]
12 mois	5,10*** (0,265)	[4,59; 5,62]	5,12*** (0,252)	[4,64; 5,60]	5,01*** (0,264)	[4,52; 5,56]
18 mois	9,05*** (0,260)	[8,54; 9,56]	9,06*** (0,261)	[8,56; 9,54]	8,95*** (0,265)	[8,39; 9,43]
24 mois	9,80*** (0,250)	[9,31; 10,29]	9,80*** (0,255)	[9,33; 10,29]	9,77*** (0,257)	[9,28; 10,22]
Emploi durable						
6 mois	2,65*** (0,209)	[2,24; 3,06]	2,66*** (0,222)	[2,21; 3,10]	2,53*** (0,205)	[2,14; 2,93]
12 mois	5,99*** (0,249)	[5,50; 6,48]	6,01*** (0,248)	[5,51; 6,50]	5,98*** (0,248)	[5,55; 6,49]
18 mois	8,89*** (0,261)	[8,38; 9,41]	8,90*** (0,270)	[8,35; 9,40]	8,85*** (0,255)	[8,33; 9,33]
24 mois	9,70*** (0,263)	[9,18; 10,22]	9,70*** (0,266)	[9,19; 10,20]	9,69*** (0,264)	[9,21; 10,18]

Note : *p<0,1 ; **p<0,05 ; ***p<0,01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Chaque type de bootstrap a été effectué avec 500 répliquions. L'ATT moyen correspond à la moyenne des 500 ATT. L'écart-type correspond à l'écart-type des 500 ATT. L'intervalle de confiance est de type "percentile", c'est-à-dire que les bornes correspondent respectivement au percentile 2,5 % et 97,5 %.

Champ : Ensemble des inscrits à Pôle emploi en janvier 2017, France entière.

Source : Dares, dispositif ForCE.

F.2 Spécification alternative : ajout de variables d'histoire d'emploi

L'historique d'emploi peut être un déterminant important pour l'insertion future mais aussi l'entrée en formation. Le dispositif ForCE ne contient cependant des informations sur les contrats de travail que depuis janvier 2017. Dans les estimations principales, l'historique passé est pris en compte par l'historique d'inscription capté sur 10 ans, mais qui ne fournit qu'une partie (en creux) de l'expérience passé en emploi. Pour estimer la sensibilité à ce choix, on peut cependant se concentrer sur 2018, ce qui permet de disposer de l'historique des contrats de travail en cours sur 2017. La base MMO recense tous les contrats de travail incluant au moins un jour de travail dans l'année considérée, et ne se restreint pas aux contrats de travail commencés dans l'année.

En pratique, quatre variables sont construites pour synthétiser l'histoire professionnelle dans les mois précédant la formation : il s'agit d'indicatrices correspondant au fait d'avoir été en emploi (soit tout type, soit durable) dans les mois précédant l'entrée en formation (soit avec une fin de contrat avant le mois d'entrée en formation, soit pendant ou après le mois d'entrée en formation). Outre les contraintes décrites dans la section 3, un appariement exact est imposé sur ces variables. Les résultats de cette estimation sont présentés dans le tableau F.2. Les estimations avec cette nouvelle contrainte d'appariement sont un peu supérieures aux

estimations de la procédure principale, de l'ordre de 0,5 à 0,6 point de pourcentage, sans être statistiquement différentes des estimations principales.

Il sera intéressant d'utiliser ces informations avec un recul suffisant pour pouvoir estimer de manière plus précise et détaillée l'historique de l'emploi. Une limite notamment est que les épisodes d'emploi sont "censurés à gauche" : on observe ceux qui sont encore actifs en 2017, au risque de sur-représenter les contrats longs (qui seront encore actifs en 2017 alors qu'ils ont commencé avant), au détriment des personnes ayant enchaîné des contrats courts. Une analyse approfondie des biais que cela pourrait introduire devrait être menée.

TABLE F.2 – Estimation par appariement, avec des variables d'histoire d'emploi pour les entrées en formation de 2018

	ATT standard	IC à 95 %	ATT avec appariement sur l'histoire d'emploi
Tout type d'emploi			
6 mois	0,34 (0,268)	[-0, 18; 0, 87]	0,76*** (0,271)
12 mois	5,68*** (0,292)	[5, 11; 6, 26]	6,34*** (0,294)
18 mois	8,09*** (0,288)	[7, 53; 8, 66]	8,73*** (0,290)
24 mois	8,25*** (0,293)	[7, 68; 8, 83]	8,79*** (0,296)
Emploi durable			
6 mois	3,91*** (0,236)	[3, 45; 4, 37]	3,94*** (0,240)
12 mois	6,84*** (0,277)	[6, 29; 7, 38]	7,15*** (0,282)
18 mois	8,65*** (0,290)	[8, 08; 9, 22]	9,03*** (0,294)
24 mois	8,76*** (0,306)	[8, 16; 9, 36]	9,16*** (0,311)

Note : * p<0,1 ; ** p<0.05 ; *** p<0.01.

L'emploi tout type recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 1 mois. L'emploi durable recouvre tout contrat salarié du privé en CDI ou dont la durée est supérieure à 6 mois.

Champ : Ensemble des inscrits à Pôle emploi entre janvier et décembre 2018, France entière.

Source : Dares, dispositif ForCE.