

Performance d’insertion des centres de formation d’apprentis en France : une analyse économétrique à partir des données InserJeunes

Hugo Spring-Ragain

CEDS Paris

hugo.springragain@edu.ceds.fr

Résumé¹

Cet article évalue les déterminants de la performance d’insertion professionnelle des centres de formation d’apprentis (CFA) en France, en exploitant les données de panel du dispositif InserJeunes. Nous examinons dans quelle mesure les écarts de performance sont associés aux arbitrages en faveur de la poursuite d’études, en analysant l’accès à l’emploi à court et moyen termes (taux d’emploi à 6 mois, valeur ajoutée d’emploi à 6 mois, et taux d’emploi à 12 mois). Notre stratégie empirique repose principalement sur des modèles à effets fixes (établissement et année) visant à purger les estimations de l’hétérogénéité inobservée invariante des CFA. Cette démarche est complétée par une distinction explicite entre CFA historiques et nouveaux entrants, par des spécifications en modèle fractionnaire pour le traitement des variables bornées, ainsi que par une approche en variables instrumentales générées en interne (Lewbel, 2012) pour pallier les biais d’endogénéité. Les résultats démontrent que l’association négative entre poursuite d’études et insertion, visible dans les spécifications transversales naïves, s’évanouit totalement une fois contrôlée l’hétérogénéité fixe des établissements. Les données ne font apparaître aucune relation intra-CFA robuste entre les variations de la poursuite d’études et les dynamiques d’insertion. *In fine*, ces conclusions soulignent que les écarts de placement observés reflètent fondamentalement des différences structurelles persistantes entre établissements, plutôt qu’un arbitrage mécanique et généralisable lié aux comportements post-formation.

Mots-clés : apprentissage, insertion professionnelle, CFA, InserJeunes, économétrie, France

Codes JEL : C23, I25, J24, J64

¹ Pour citer cet article : Spring-Ragain, H. (2026). Performance d’insertion des centres de formation d’apprentis en France : une analyse économétrique à partir des données InserJeunes. *Modélisation & Analyse Economique*, Vol.1 n°2.

Abstract

This article evaluates the determinants of the labor market integration performance of apprenticeship training centers (CFA) in France, using panel data from the InserJeunes system. We examine the extent to which performance gaps are associated with trade-offs in favor of further education, by analyzing short- and medium-term employment access (employment rate at 6 months, employment value-added at 6 months, and employment rate at 12 months). Our empirical strategy relies primarily on fixed-effects models (establishment and year) aimed at purging estimates of the time-invariant unobserved heterogeneity of the CFAs. This approach is complemented by an explicit distinction between incumbent CFAs and new entrants, fractional model specifications to handle bounded variables, as well as an internally generated instrumental variables approach (Lewbel, 2012) to mitigate endogeneity biases. The results demonstrate that the negative association between further education and labor market integration, visible in naive cross-sectional specifications, completely vanishes once the fixed heterogeneity of establishments is controlled for. The data reveal no robust intra-CFA relationship between variations in further education and integration dynamics. Ultimately, these findings emphasize that the observed placement gaps fundamentally reflect persistent structural differences between establishments, rather than a mechanical and generalizable trade-off driven by post-training behaviors.

Keywords: apprenticeship, labour market integration, CFA, InserJeunes, econometrics, France

JEL Classification: C23, I25, J24, J64

1. Introduction

L'expansion massive de l'apprentissage s'est imposée comme un axe structurel des politiques d'éducation et d'emploi en France, une dynamique particulièrement accélérée par la loi pour la « liberté de choisir son avenir professionnel » de 2018. Si les enquêtes empiriques de référence, telles que celles du CEREQ (Arrighi et Gasquet, 2010 ; CEREQ, 2021), s'accordent sur le rendement moyen favorable de l'alternance pour la transition école-emploi, ces performances macroéconomiques occultent une très forte hétérogénéité microéconomique et spatiale. L'insertion des sortants n'est ni uniforme ni homogène, et les trajectoires varient substantiellement selon les établissements (les Centres de Formation d'Apprentis, ou CFA). Cette variance soulève une question évaluative centrale : dans quelle mesure les écarts d'insertion observés entre les CFA reflètent-ils de véritables différentiels d'efficacité institutionnelle, plutôt que des biais de composition ou des comportements individuels post-formation ?

La littérature théorique et empirique a largement documenté la difficulté d'isoler la contribution propre des établissements éducatifs aux trajectoires des apprenants (Hanushek, 1986), ce qui a conduit au développement des modèles de valeur ajoutée (Chetty, Friedman et Rockoff, 2014). Dans le champ spécifique de la formation professionnelle, la littérature européenne récente (Kemper et al., 2026 ; McGuinness et al., 2025) souligne que les rendements de l'apprentissage dépendent intrinsèquement des configurations institutionnelles et de l'intégration en entreprise (Acemoglu et Pischke, 1999 ; Bentolila, Cabrales et Jansen, 2023). Toutefois, ces travaux comparatifs comportent un angle mort significatif : ils peinent à capturer l'hétérogénéité microéconomique intra-nationale et tendent à omettre le rôle perturbateur des comportements de poursuite d'études dans l'évaluation de l'efficacité locale. Un établissement incitant ses apprentis à s'engager dans une formation supérieure affichera mécaniquement un taux d'emploi à court terme plus faible. L'utilisation conventionnelle des taux bruts d'insertion confond ainsi la qualité intrinsèque du CFA avec des trajectoires d'accumulation de capital humain de plus long terme, produisant une mesure fallacieuse de la performance.

Le présent article comble cette lacune en analysant les déterminants des écarts de performance d'insertion entre les CFA en France, à partir du nouveau dispositif de données InserJeunes. Notre démarche s'attache à décomposer ces écarts en testant l'hypothèse d'un effet d'éviction de court terme induit par la poursuite d'études, tout en contrôlant la variance liée aux bassins d'emploi régionaux et aux chocs macroéconomiques annuels. En mobilisant une approche économétrique sur données de panel (CFA-année) couvrant la période de cumul 2018-2019 à 2023-2024, nous comparons le taux d'emploi à 6 mois, la valeur ajoutée d'emploi à 6 mois (qui purge les effets de composition du public) et le taux d'emploi à 12 mois. Afin de neutraliser l'hétérogénéité structurelle inobservée des établissements, l'estimation centrale repose sur des modèles à effets fixes (CFA et année). Pour répondre aux défis d'identification, notre stratégie est complétée par une distinction entre CFA historiques et nouveaux entrants, par des modèles fractionnaires adaptés aux variables bornées, et par une approche en variables instrumentales générées en interne (Lewbel, 2012) pour traiter l'endogénéité de la poursuite d'études.

Les résultats de nos estimations invitent à une lecture prudente et remettent en cause l'interprétation naïve des taux bruts. Si les spécifications transversales simples suggèrent une association négative entre le taux de poursuite d'études et l'insertion à court terme, cette relation s'évanouit totalement une fois l'hétérogénéité fixe des établissements contrôlée. Nous ne trouvons aucune relation robuste entre la variation de la poursuite d'études et la variation de l'insertion au sein d'un même CFA, y compris en distinguant les établissements historiques des nouveaux. Les tests de robustesse (modèles fractionnaires et instruments de Lewbel) confirment l'absence de biais substantiel de l'estimateur OLS *within*. *In fine*, l'article démontre que les écarts d'insertion observés traduisent des différences structurelles persistantes entre établissements (hétérogénéité invariante), plutôt qu'un mécanisme causal généralisable liant la dynamique des poursuites d'études aux performances de placement.

Le reste de l'article est structuré comme suit. La section 2 présente la revue de la littérature théorique et empirique sur l'évaluation des dispositifs de formation professionnelle. La section 3 détaille la modélisation empirique : elle expose la base de données InserJeunes et la stratégie économétrique, avant de présenter les résultats principaux et d'éprouver leur solidité par divers tests de robustesse (modèles fractionnaires et approche par variables instrumentales). Enfin, la section 4 conclut l'analyse.

2. Revue de littérature

2.1 Ancrage théorique : évaluation de la performance, valeur ajoutée et transition professionnelle

L'évaluation de la performance des établissements de formation constitue un défi méthodologique et théorique central en économie de l'éducation. Les travaux fondateurs portant sur les fonctions de production éducative ont mis en évidence la difficulté persistante à isoler la contribution propre des institutions scolaires aux trajectoires des apprenants, indépendamment des caractéristiques sociodémographiques et académiques initiales de leur public (Hanushek, 1986). Pour surmonter ces effets de composition ou de sélection endogène, les approches empiriques récentes se sont structurées autour de la notion de valeur ajoutée (Chetty, Friedman et Rockoff, 2014). Ce cadre d'analyse vise à extraire l'effet causal et intrinsèque de l'établissement (ou de l'enseignant) sur les résultats individuels, en neutralisant l'hétérogénéité observable des cohortes accueillies. Cette littérature souligne que les indicateurs bruts (tels que les taux de réussite ou d'emploi) constituent des signaux biaisés de l'efficacité institutionnelle s'ils ne sont pas purgés de ces effets de sélection. C'est d'ailleurs cette logique de contrôle de l'hétérogénéité qui sous-tend la construction institutionnelle de l'indicateur de valeur ajoutée d'emploi au sein du dispositif InserJeunes.

Parallèlement, la littérature relative à l'économie de la formation professionnelle et de la transition école-emploi démontre que l'insertion sur le marché du travail dans un cadre de concurrence imparfaite dépend non seulement du capital humain accumulé, mais également des structures institutionnelles et des comportements de signalement (Acemoglu et Pischke, 1999 ; Ryan, 2001). Dans ce contexte, les comportements post-formation, et tout particulièrement les arbitrages en faveur d'une poursuite d'études, modifient drastiquement l'interprétation des performances d'insertion. Un centre de formation d'apprentis (CFA) dont une fraction importante des sortants opte pour une formation complémentaire de niveau

supérieur affichera mécaniquement un taux d'emploi à court terme plus faible. Toutefois, ce phénomène ne traduit pas nécessairement une rentabilité dégradée de la formation initiale ni une inefficacité de l'établissement, mais signale plutôt des trajectoires de maximisation du capital humain à plus long terme.

Enfin, la modélisation théorique de l'insertion professionnelle requiert l'intégration stricte des effets de contexte spatiaux et temporels (Ryan, 2001). Les disparités interrégionales d'absorption du marché du travail, la structure locale de la demande de travail, ainsi que les chocs macroéconomiques conjoncturels influencent significativement et de manière exogène les probabilités d'accès à l'emploi des apprentis (CEREQ, 2021). Une mesure rigoureuse de la performance des CFA impose par conséquent d'isoler le rôle de ces contraintes environnementales et temporelles, une logique au cœur des modèles d'évaluation en valeur ajoutée (Chetty, Friedman et Rockoff, 2014 ; DEPP, 2024), pour éviter d'imputer aux établissements des différentiels d'insertion qui relèvent en réalité d'externalités géographiques ou macroéconomiques.

2.2 Littérature empirique : hétérogénéité institutionnelle et angle mort microéconomique

Dans une perspective comparative, la littérature empirique européenne confirme que les effets de la formation professionnelle sur l'insertion dépendent intrinsèquement des configurations institutionnelles et des modalités concrètes de mise en œuvre des dispositifs. Une méta-analyse récente des évaluations causales portant sur la formation professionnelle formelle (Kemper et al., 2026) met en évidence un effet moyen positif sur la probabilité d'accès à l'emploi, mais révèle une variance nettement plus prononcée concernant les rendements salariaux. Ces résultats invitent à interpréter les rendements de la voie professionnelle non pas comme une constante, mais comme une variable fortement conditionnée par les contextes macroéconomiques nationaux et les groupes de contrôle retenus.

Dans le même esprit, des évaluations récentes portant sur la dualisation de la formation professionnelle, à l'instar du cas espagnol (Bentolila, Cabrales et Jansen, 2023), démontrent que l'implantation de dispositifs hybrides se rapprochant du modèle de l'apprentissage tend à améliorer l'insertion initiale sur le marché du travail. Ces travaux laissent toutefois ouverte la question de la validité externe et de la généralisation de ces effets à d'autres architectures de formation. De surcroît, les travaux comparatifs à l'échelle européenne (McGuinness et al., 2025 ; Michaelis et Findeisen, 2024) soulignent que les trajectoires des diplômés de la voie professionnelle divergent structurellement selon le degré d'intégration de l'apprentissage en situation de travail, l'élasticité de la demande locale de travail et l'organisation institutionnelle des filières.

Cependant, si cette littérature macro-institutionnelle documente largement l'impact des architectures nationales sur les rendements moyens de l'apprentissage (Ryan, 2001 ; Kemper et al., 2026), elle comporte un angle mort significatif : elle peine souvent à capturer l'hétérogénéité microéconomique intra-nationale des établissements et tend à omettre le rôle des comportements post-formation dans l'évaluation de l'efficacité locale (Hanushek, 1986 ; Chetty, Friedman et Rockoff, 2014). La majorité des études empiriques interprètent

les écarts d'emploi comme de pures différences d'efficacité des centres de formation, sans isoler le biais mécanique induit par les stratégies individuelles de prolongation de la scolarité.

C'est précisément cette lacune analytique que la modélisation proposée dans la section suivante s'attache à combler. En s'appuyant sur le cas français à travers le dispositif InserJeunes (DEPP, 2024), notre démarche empirique examine comment les écarts d'insertion entre établissements sont spécifiquement associés au taux de poursuite d'études. Cette stratégie permet de dépasser les simples comparaisons descriptives en neutralisant rigoureusement, par une approche économétrique en panel (Angrist et Pischke, 2009 ; Wooldridge, 2010), les hétérogénéités inobservées et fixes des Centres de Formation d'Apprentis (CFA), les chocs temporels, ainsi que les effets de composition induits par la transformation structurelle récente du secteur, observée depuis la réforme de 2018 (DARES, 2023).

3. Modélisation et résultats

3.1 Source des données et unité d'analyse

L'analyse empirique repose sur les données du dispositif InserJeunes appliquées aux centres de formation d'apprentis (CFA) en France. Cette base fournit, pour plusieurs millésimes récents, des indicateurs harmonisés relatifs à l'insertion professionnelle des sortants de l'apprentissage. Les observations sont organisées au niveau des établissements de formation et des années disponibles, ce qui permet de construire un panel de type CFA-année.

L'unité d'analyse retenue dans cette étude est donc le couple (i, t) , où i désigne un CFA et t une année d'observation. Cette structure de données permet d'étudier à la fois les différences de performance entre établissements et leur évolution dans le temps.

3.2 Variables d'intérêt

Le tableau 1 présente les principales variables mobilisées dans l'analyse.

Table 1 - Présentation des variables retenues

Variable	Définition	Rôle
Taux emploi 6 mois	Part des sortants du CFA en emploi six mois après la sortie de formation	Variable dépendante principale
VA emploi 6 mois	Écart entre le taux d'emploi observé à 6 mois et le taux d'emploi attendu	Variable dépendante alternative
Taux emploi 12 mois	Part des sortants du CFA en emploi douze mois après la sortie de formation	Variable de robustesse
Taux poursuite études	Part des sortants poursuivant une formation après la sortie du CFA	Variable explicative principale
Région	Région d'implantation du CFA	Contrôle / effets fixes
Année	Millésime d'observation	Contrôle / effets fixes
UAI	Identifiant unique de l'établissement	Identifiant panel

Trois variables dépendantes sont étudiées dans l'article : le taux d'emploi à 6 mois, la valeur ajoutée d'emploi à 6 mois et le taux d'emploi à 12 mois. La variable explicative principale est le taux de poursuite d'études, qui mesure la part des sortants poursuivant une formation après la sortie du CFA.

Précision sur la valeur ajoutée d'emploi à 6 mois. Dans le dispositif InserJeunes, la valeur ajoutée d'emploi à 6 mois ne correspond pas à un taux brut d'insertion, mais à un écart entre le taux d'emploi observé et un taux d'emploi attendu, calculé pour des établissements comparables. Elle vise ainsi à apprécier la performance relative d'un établissement en neutralisant, autant que possible, l'influence de facteurs extérieurs à l'établissement, tels que le profil des jeunes, le type et le niveau de formation, la spécialité préparée ou encore les conditions locales du marché du travail.

Formellement, pour un établissement i à la date t , on peut écrire :

$$VA_{it}^{6m} = E_{it}^{obs,6m} - E_{it}^{att,6m} \quad (1)$$

où $E_{it}^{obs,6m}$ désigne le taux d'emploi salarié observé à 6 mois et $E_{it}^{att,6m}$ le taux d'emploi attendu à 6 mois. Le taux attendu est obtenu à partir d'un modèle statistique mobilisant notamment le sexe, l'âge, la certification, la spécialité de formation, la catégorie socioprofessionnelle des parents, les résultats à l'examen, certaines caractéristiques de l'établissement et le taux de chômage de la zone d'emploi de résidence. Pour les apprentis, sont également pris en compte la situation avant l'apprentissage, la NAF de l'établissement d'accueil et la reconnaissance éventuelle de la qualité de travailleur handicapé (RQTH).

Une valeur ajoutée positive signifie donc que le taux d'emploi observé est supérieur au taux attendu pour des établissements comparables ; une valeur négative indique l'inverse. Cet indicateur a une portée relative : il complète le taux d'emploi brut mais ne constitue pas, à lui seul, un classement absolu des établissements.

3.3 Spécifications économétriques

L'objectif de l'analyse économétrique est d'identifier les facteurs associés aux écarts de performance d'insertion entre CFA en France. Compte tenu des limites de la base mobilisée, la démarche ne vise pas à établir un effet causal de la poursuite d'études, mais à documenter des associations conditionnelles entre cette variable et différents indicateurs d'insertion, après prise en compte d'un ensemble de contrôles fixes.

Le point de départ consiste à estimer une spécification avec effets fixes régionaux et annuels :

$$Y_{it} = \alpha + \beta_1 Poursuite_{it} + \gamma_r + \delta_t + \varepsilon_{it} \quad (2)$$

où :

- Y_{it} désigne l'indicateur d'insertion observé pour le CFA i à l'année t ;
- $Poursuite_{it}$ représente le taux de poursuite d'études ;

- γ_r capture les effets fixes de région ;
- δ_t représente les effets fixes d'année ;
- ε_{it} est le terme d'erreur.

Dans cette première famille de modèles, Y_{it} correspond successivement au taux d'emploi à 6 mois, à la valeur ajoutée d'emploi à 6 mois, puis au taux d'emploi à 12 mois.

Toutefois, cette spécification ne neutralise pas les caractéristiques inobservées propres à chaque établissement. Pour cette raison, l'estimation centrale de l'article repose sur une seconde spécification intégrant des effets fixes CFA et année :

$$Y_{it} = \alpha + \beta_1 Poursuite_{it} + \mu_i + \delta_t + \varepsilon_{it} \quad (3)$$

où μ_i représente les effets fixes propres à chaque CFA. Cette formulation exploite la variation intra-établissement dans le temps et permet de neutraliser les caractéristiques fixes non observées des établissements, telles que leur spécialisation dominante, leur taille structurelle ou certaines composantes persistantes de leur public.

Afin de tenir compte plus explicitement du changement de composition du panel au cours de la période étudiée, une spécification avec interaction entre poursuite d'études et statut d'ancienneté de l'établissement est également estimée :

$$Y_{it} = \alpha + \beta_1 Poursuite_{it} + \beta_2 (Poursuite_{it} \times Nouveau_i) + \mu_i + \delta_t + \varepsilon_{it} \quad (4)$$

où $Nouveau_i$ désigne un indicateur valant 1 pour les nouveaux CFA et 0 pour les CFA historiques. Le coefficient β_1 mesure alors l'association estimée pour les CFA historiques, tandis que $\beta_1 + \beta_2$ correspond à l'association estimée pour les nouveaux CFA.

3.4 Lecture descriptive de la relation étudiée

Avant même l'estimation économétrique, l'examen graphique de la relation entre poursuite d'études et emploi à 6 mois permet d'illustrer la logique de l'hypothèse testée.

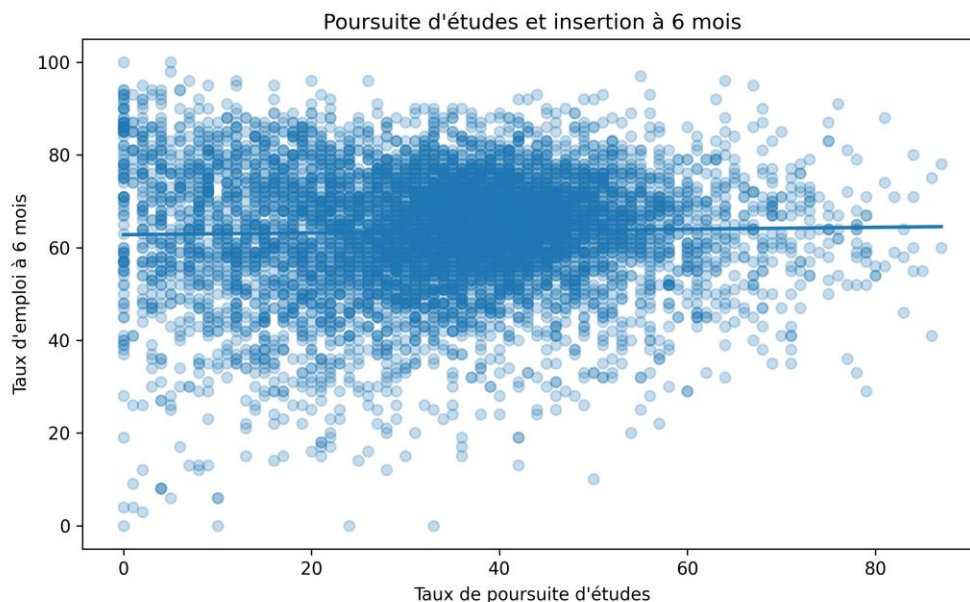


Figure 1 - Poursuite d'études et insertion à 6 mois

La figure 1 met en évidence une relation descriptive globalement peu marquée entre taux de poursuite d'études et taux d'emploi à 6 mois. La forte dispersion du nuage de points confirme que la relation brute entre ces deux variables est insuffisante pour tirer une conclusion solide, ce qui justifie le recours à une modélisation contrôlant les effets fixes territoriaux, temporels et, surtout, les effets fixes établissement.

3.5 Robustesses

Au-delà des estimations principales, cinq exercices de robustesse sont retenus.

Premièrement, la substitution du taux d'emploi à 12 mois au taux d'emploi à 6 mois permet d'évaluer si les associations observées à court terme se prolongent à un horizon plus long.

Deuxièmement, une réestimation sur des sous-échantillons distinguant CFA historiques et nouveaux CFA permet de vérifier que les résultats ne sont pas uniquement portés par la transformation progressive de la composition du panel.

Troisièmement, afin de tenir compte plus explicitement du changement de composition de l'échantillon au cours de la période étudiée, des modèles avec interaction entre taux de poursuite d'études et statut d'ancienneté de l'établissement sont estimés. Cette stratégie permet de tester si les associations observées diffèrent entre les CFA historiques et les établissements apparus plus récemment dans la base.

Quatrièmement, pour les seules variables de taux d'emploi à 6 mois et à 12 mois, une robustesse en modèle fractionnaire est mise en oeuvre. Les taux sont alors ramenés à l'intervalle [0,1] et la relation conditionnelle peut être écrite sous la forme suivante :

$$E(Y_{it} | X_{it}) = G(\alpha + \beta_1 Poursuite_{it} + \mu_i + \delta_t) \quad (5)$$

où $G(\cdot)$ désigne une fonction de lien logistique. Cette spécification permet de vérifier que les résultats ne dépendent pas uniquement de l'approximation linéaire retenue dans les modèles principaux.

Cinquièmement, afin d'examiner dans quelle mesure le biais d'endogénéité potentiel du taux de poursuite d'études affecte les estimations principales, une stratégie d'identification par variables instrumentales générées en interne est mise en, suivant. Cette méthode exploite l'hétéroscédasticité des résidus du premier stade pour construire des instruments valides sans recourir à une source de variation externe. Formellement, pour chaque variable exogène centrée $\tilde{z}_{it}$, l'instrument est défini comme :

$$q_{it} = \hat{\varepsilon}_{it} \cdot \tilde{z}_{it} \quad (6)$$

où $\hat{\varepsilon}_{it}$ désigne le résidu de la régression du taux de poursuite d'études sur les effets fixes année. Sous l'hypothèse d'hétéroscédasticité, q_{it} est corrélé avec la variable endogène mais non corrélé avec le terme d'erreur de l'équation principale. Les résultats de cette robustesse sont discutés dans la section 6.4.

3.6 Limites d'identification

L'interprétation des coefficients estimés doit demeurer prudente. En particulier, le taux de poursuite d'études ne peut être considéré comme exogène. Plusieurs mécanismes peuvent simultanément affecter la poursuite d'études et l'insertion professionnelle : composition académique et sociale du public accueilli, qualité ou spécialisation des formations, stratégies pédagogiques des CFA, conditions locales du marché du travail ou encore préférences des jeunes en matière de poursuite de parcours. Dans ces conditions, le coefficient associé à la poursuite d'études ne saurait recevoir une interprétation causale.

De plus, la variable de poursuite d'études agrège potentiellement des situations hétérogènes. Elle peut refléter aussi bien des poursuites choisies, associées à des trajectoires ambitieuses de montée en qualification, que des poursuites davantage contraintes, liées à une insertion plus difficile à court terme. Faute d'information plus fine sur la nature des poursuites observées, l'interprétation économique de cette variable reste partiellement ouverte.

Une exploration complémentaire a été menée à partir de la variable *Part contrats apprentissage interrompus*, seule variable de contrôle additionnelle directement exploitable dans la base au niveau du CFA et variant dans le temps. Toutefois, cette variable n'est renseignée que sur un sous-échantillon beaucoup plus restreint que celui mobilisé dans l'analyse principale. Dans les spécifications avec effets fixes établissement, ce sous-échantillon se révèle trop limité pour permettre une inférence économétrique satisfaisante, les modèles devenant quasi saturés. Pour cette raison, cette variable n'a pas été retenue dans les estimations centrales.

Une tentative d'estimation en panel dynamique GMM a également été explorée afin d'atténuer l'endogénéité potentielle du taux de poursuite d'études, en traitant cette variable comme potentiellement endogène et en introduisant une dynamique de l'indicateur d'insertion. Toutefois, ces estimations n'ont pas été retenues dans l'analyse principale. En effet, les diagnostics usuels de validité des instruments et de corrélation sérielle en

différences se sont révélés défavorables dans les spécifications considérées, ce qui ne permet pas d'interpréter les coefficients obtenus de manière fiable.

Cette difficulté tient vraisemblablement à plusieurs caractéristiques de la base mobilisée : le faible nombre de périodes disponibles, le caractère déséquilibré du panel, ainsi que la structure même des millésimes observés, qui reposent sur des cumuls glissants et induisent une forte persistance mécanique entre périodes successives. Dès lors, si ces explorations confirment l'importance de la question d'endogénéité et du choix des contrôles disponibles, elles ne permettent pas, en l'état, d'apporter une réponse économétrique pleinement satisfaisante.

Une stratégie complémentaire d'identification par variables instrumentales générées en interne a été mise en suivant, afin d'évaluer directement l'ampleur du biais d'endogénéité potentiel. Les résultats de cette exploration sont présentés en section 6.4.

En conséquence, les résultats présentés dans la suite doivent être lus comme des *associations conditionnelles* et non comme des effets propres identifiés de la poursuite d'études sur les performances d'insertion.

3.7 Description de la base de données

La base de données mobilisée dans cette étude provient du dispositif InserJeunes appliqué aux centres de formation d'apprentis (CFA) en France. Elle couvre plusieurs millésimes récents et fournit des indicateurs harmonisés sur l'insertion professionnelle des sortants de l'apprentissage, observés au niveau des établissements.

Au total, le fichier brut comprend 11 314 observations et 3 166 CFA distincts. Après nettoyage et sélection des variables pertinentes pour l'analyse, l'échantillon principal retenu comprend 7 833 observations correspondant à 2 340 CFA distincts sur six millésimes, allant de *cumul 2018-2019* à *cumul 2023-2024*. La structure des données permet ainsi de constituer un panel CFA-année.

L'échantillon principal est construit à partir des observations pour lesquelles les variables nécessaires à l'estimation des modèles sont renseignées, en particulier le taux d'emploi à 6 mois, la valeur ajoutée d'emploi à 6 mois, le taux d'emploi à 12 mois, le taux de poursuite d'études, la région et l'année. Certaines variables présentent toutefois une couverture plus large que d'autres. En particulier, le taux de poursuite d'études est renseigné pour 8 883 observations, tandis que l'échantillon principal des estimations est restreint à 7 833 observations en raison de la disponibilité conjointe des indicateurs d'emploi et de valeur ajoutée.

3.8 Statistiques descriptives des variables

Avant d'estimer les modèles économétriques, il convient de présenter les principales caractéristiques de l'échantillon et la distribution des variables retenues. Cette étape permet de mieux apprécier les ordres de grandeur des indicateurs d'insertion, leur dispersion entre établissements, ainsi que la qualité de la couverture statistique de la base.

Le tableau 2 présente les statistiques descriptives des variables quantitatives utilisées dans l'analyse.

Table 2 - Statistiques descriptives des principales variables

Variable	Obs.	Moyenne	Écart-type	Min.	Max.
Taux emploi 6 mois	7 833	63.47	12.61	0.00	100.00
VA emploi 6 mois	7 833	-0.06	6.65	-12.00	12.00
Taux emploi 12 mois	7 833	69.03	12.15	0.00	100.00
Taux poursuite études	8 883	35.45	16.97	0.00	100.00

Afin de visualiser plus directement les grandes tendances de l'échantillon, plusieurs graphiques descriptifs sont présentés ci-dessous. Ils permettent de rendre compte à la fois de l'évolution moyenne des indicateurs dans le temps, de leur dispersion entre établissements, ainsi que de certaines hétérogénéités territoriales.

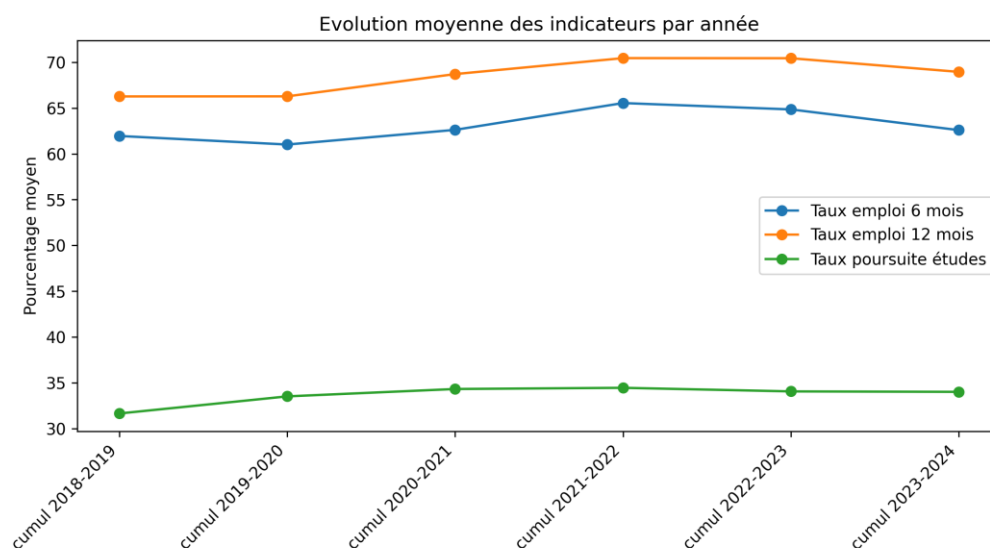


Figure 2 - Évolution moyenne des indicateurs par année

La figure 2 met en évidence une progression modérée des taux moyens d'emploi à 6 mois et à 12 mois entre les premiers millésimes et les années les plus récentes. Le taux moyen de poursuite d'études suit également une légère tendance haussière avant de se stabiliser en fin de période.

L'examen descriptif de la base met en évidence plusieurs enseignements. D'abord, le taux d'emploi à 6 mois s'établit en moyenne à 63.47 %, ce qui confirme qu'une part importante des sortants de CFA accède rapidement à l'emploi, tout en laissant apparaître une dispersion notable entre établissements. Ensuite, le taux d'emploi à 12 mois atteint en moyenne 69.03 %, ce qui suggère une amélioration progressive de l'insertion à mesure que l'horizon temporel s'allonge.

Par ailleurs, la valeur ajoutée d'emploi à 6 mois présente une moyenne très proche de zéro (-0.06), ce qui est cohérent avec sa définition comme écart entre performance observée et performance attendue. Enfin, le taux de poursuite d'études s'élève en moyenne à 35.45 %, avec une forte hétérogénéité entre établissements, ce qui justifie pleinement son introduction comme variable explicative centrale dans l'analyse économétrique.

Cette première lecture des données peut être complétée par l'examen de la distribution des indicateurs selon les années. Les figures suivantes montrent que les écarts entre établissements demeurent importants pour l'ensemble des variables, malgré une relative stabilité des médianes au fil du temps.

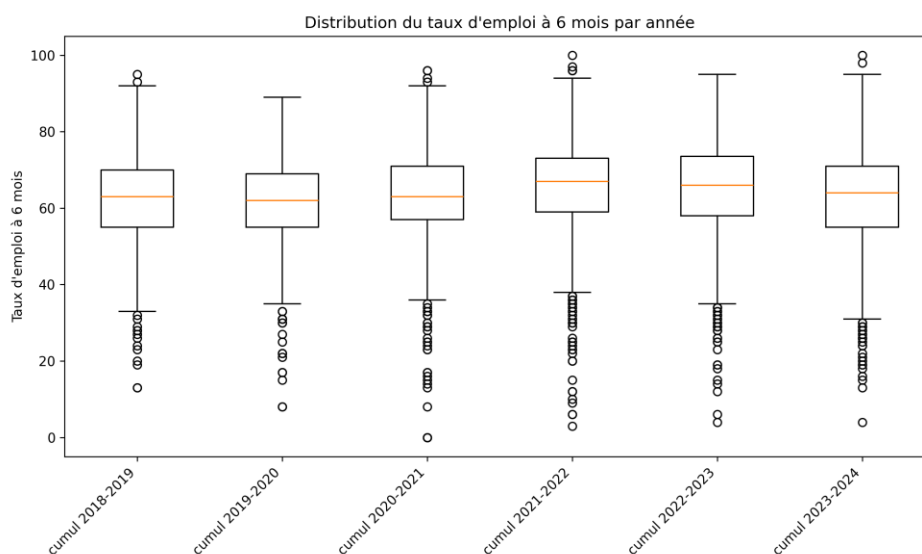


Figure 3 - Distribution du taux d'emploi à 6 mois par année

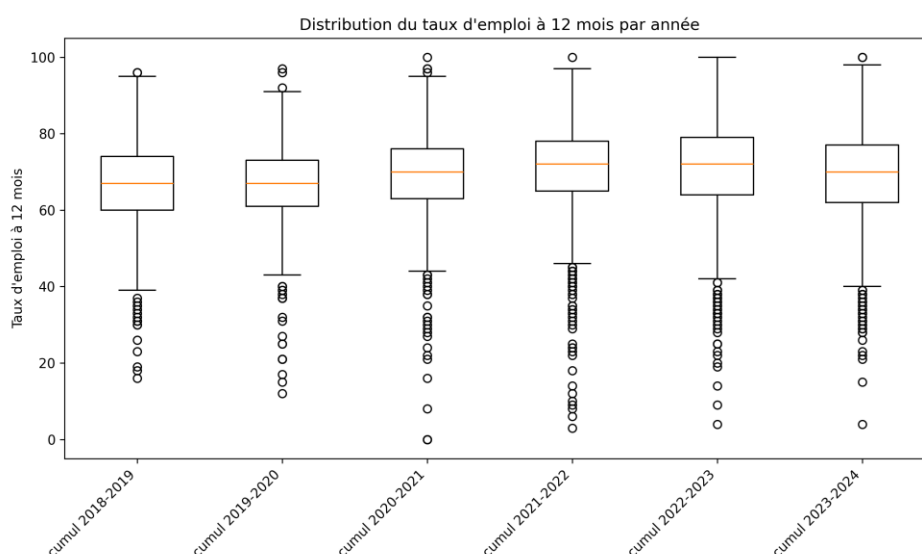


Figure 4 - Distribution du taux d'emploi à 12 mois par année

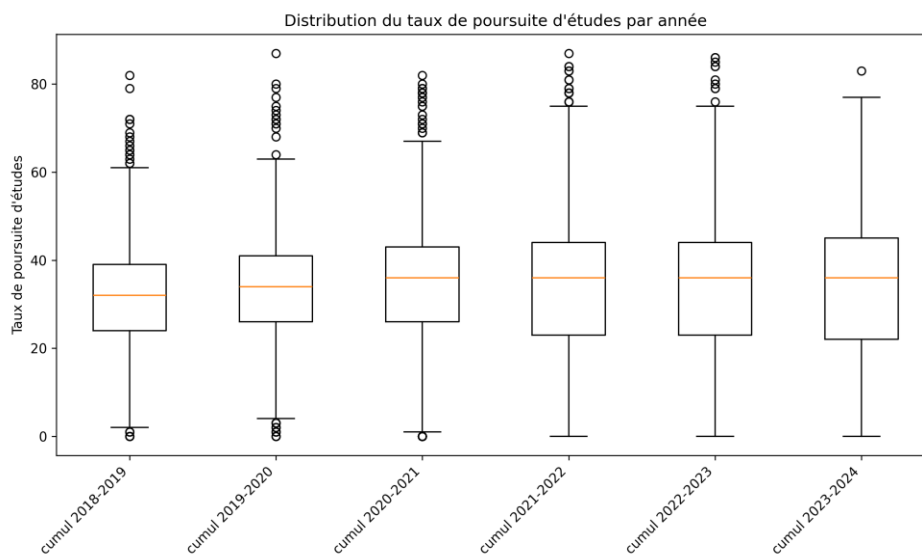


Figure 5 - Distribution du taux de poursuite d'études par année

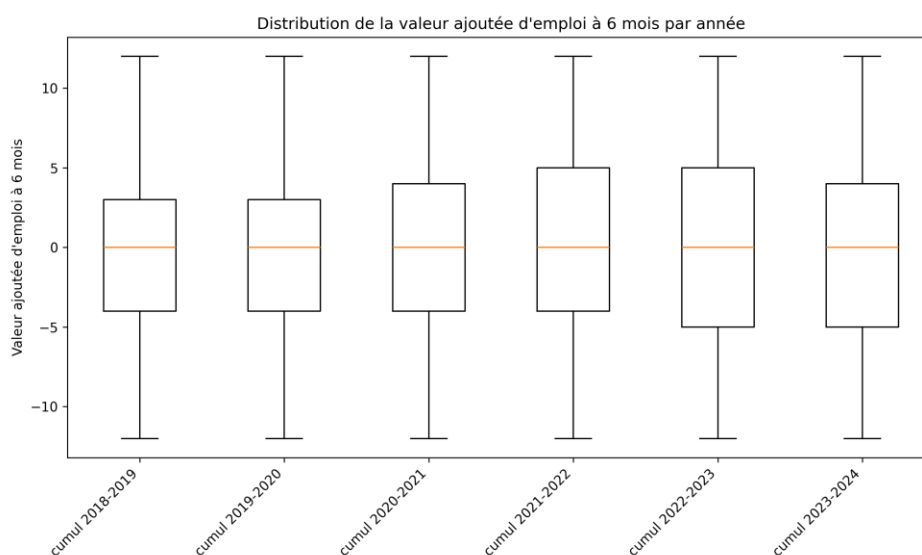


Figure 6 - Distribution de la valeur ajoutée d'emploi à 6 mois par année

Les figures 3 à 6 confirment la présence d'une forte hétérogénéité entre CFA. Les distributions des taux d'emploi à 6 mois et à 12 mois restent relativement stables d'un millésime à l'autre, avec un léger déplacement vers le haut autour des années intermédiaires. La distribution du taux de poursuite d'études apparaît également dispersée, ce qui renforce l'intérêt de cette variable dans l'analyse économétrique. Enfin, la valeur ajoutée d'emploi à 6 mois est centrée autour de zéro, conformément à sa définition, tout en laissant apparaître des écarts substantiels entre établissements.

Afin de compléter cette première lecture des données, il est utile de présenter la répartition des observations par année. Cette étape permet de vérifier que l'échantillon retenu couvre

l'ensemble des millésimes disponibles, mais aussi de documenter le caractère déséquilibré du panel.

Table 3 - Répartition de l'échantillon principal par année

Année	Nombre d'observations
cumul 2018-2019	739
cumul 2019-2020	779
cumul 2020-2021	1 045
cumul 2021-2022	1 502
cumul 2022-2023	1 795
cumul 2023-2024	1 973

Le tableau 3 montre que le panel est nettement déséquilibré, avec une montée progressive du nombre d'observations au fil des millésimes. Ce point importe pour l'interprétation des résultats : les effets fixes annuels capturent des différences moyennes entre années, mais ces moyennes sont calculées sur des compositions d'échantillon qui varient sensiblement dans le temps. Une partie des écarts annuels estimés peut donc refléter des changements de couverture statistique autant que de véritables évolutions d'insertion.

Afin de documenter plus directement cette transformation du panel, la figure 7 distingue les observations correspondant aux CFA historiques de celles relatives aux nouveaux CFA apparus dans la base après les premières vagues.

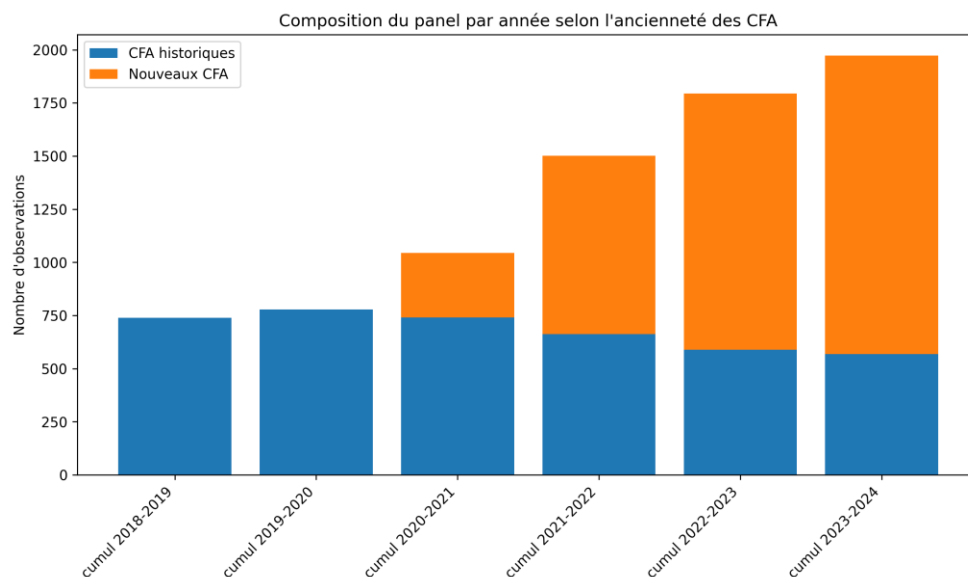


Figure 7 - Composition du panel par année selon l'ancienneté des CFA

La figure 7 met en évidence que l'élargissement progressif de l'échantillon tient largement à l'entrée de nouveaux établissements dans la base au fil du temps. Cette évolution est cohérente avec le changement institutionnel intervenu après la réforme de 2018 et justifie que l'analyse économétrique distingue explicitement les CFA historiques des nouveaux CFA.

Elle motive également la mise en place de tests de robustesse destinés à vérifier que les résultats ne sont pas uniquement portés par la transformation progressive de la composition du panel.

Les écarts territoriaux apparaissent également marqués, ce qui justifie l'introduction d'effets fixes régionaux dans certaines spécifications.

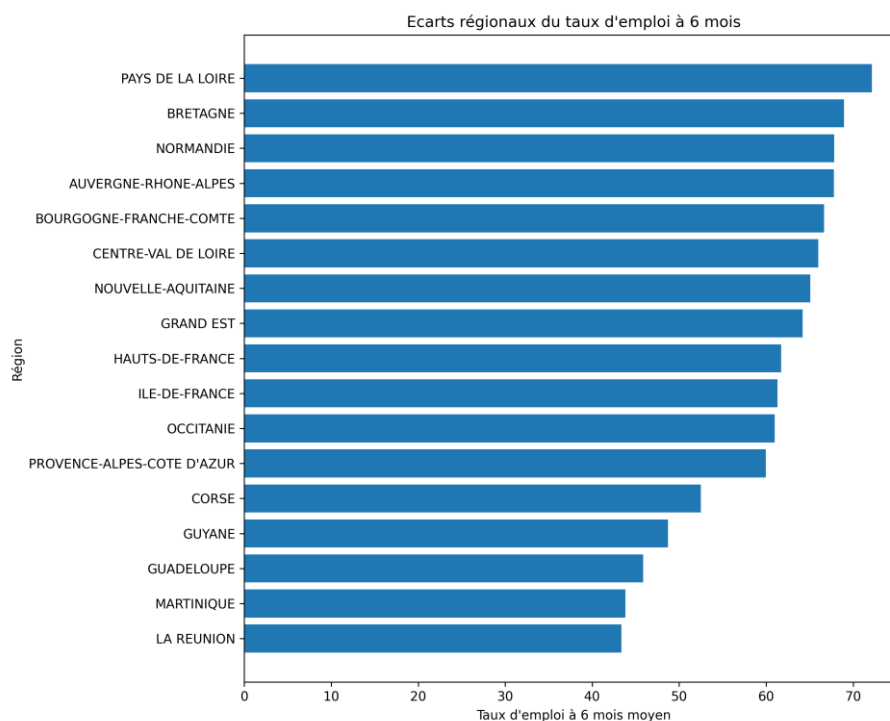


Figure 8 - Écarts régionaux du taux d'emploi à 6 mois

La figure 8 montre que les taux moyens d'emploi à 6 mois diffèrent sensiblement selon les régions. Certaines régions métropolitaines de l'Ouest et du Nord-Ouest apparaissent mieux positionnées, tandis que plusieurs territoires ultramarins présentent des niveaux moyens d'insertion plus faibles. Ces écarts confortent la pertinence d'un contrôle explicite des effets territoriaux dans les modèles estimés.

3.9 Présentation des résultats économétriques

Les résultats économétriques sont présentés en quatre temps. Le tableau 4 reporte d'abord les estimations avec effets fixes CFA et année en distinguant explicitement les CFA historiques et les nouveaux CFA par le biais d'un terme d'interaction. Le tableau 5 présente ensuite les robustesses en modèle fractionnaire pour les taux d'emploi à 6 mois et à 12 mois. Enfin, les résultats sont discutés à la lumière des estimations séparées sur sous-échantillons historiques et nouveaux CFA, utilisées comme vérification complémentaire.

Table 4 - Associations entre poursuite d'études et insertion : interaction entre CFA historiques et nouveaux

	Emp. 6m	VA 6m	Emp. 12m
Association pour CFA historiques	-0.0357 (0.0402)	-0.0277 (0.0218)	0.0076 (0.0422)
Supplément pour nouveaux CFA	0.0422 (0.0541)	0.0225 (0.0289)	0.0126 (0.0560)
Association totale pour nouveaux CFA	0.0065 (0.0366)	-0.0052 (0.0191)	0.0202 (0.0374)
Effets fixes CFA	Oui	Oui	Oui
Effets fixes année	Oui	Oui	Oui
Observations	7,833	7,833	7,833
R^2	0.830	0.702	0.831

Notes : erreurs standards clusterisées au niveau du CFA entre parenthèses. Les modèles incluent des effets fixes CFA et année. Les CFA historiques sont définis comme les établissements présents dès les deux premières vagues (*cumul 2018-2019* ou *cumul 2019-2020*) ; les nouveaux CFA apparaissent pour la première fois à partir de *cumul 2020-2021*. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table 5 - Robustesse en modèle fractionnaire pour les taux d'emploi

	Emp. 6m	Emp. 12m
Within-CFA historiques	-0.0181 (0.0147)	0.0021 (0.0166)
Supplément within pour nouveaux CFA	0.0231 (0.0197)	0.0091 (0.0219)
Within-CFA nouveaux (total)	0.0050 (0.0133)	0.0112 (0.0145)
Entre CFA (moyenne de poursuite)	-0.0137* (0.0072)	-0.0029 (0.0073)
Effets fixes année	Oui	Oui
Effets fixes région	Oui	Oui
Observations	7,833	7,833
Pseudo- R^2	0.231	0.271

Notes : modèles de type fractional logit estimés sur les taux d'emploi ramenés à l'intervalle [0,1], avec erreurs standards clusterisées au niveau du CFA. La variable de poursuite d'études

est divisée par 10 : les coefficients correspondent donc à une variation associée à 10 points de pourcentage de poursuite d'études. La décomposition *within/between* distingue la variation intra-CFA dans le temps de l'hétérogénéité moyenne entre CFA. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Afin de synthétiser visuellement les estimations principales, la figure 9 présente les coefficients associés au taux de poursuite d'études selon les variables dépendantes et selon le statut des établissements.

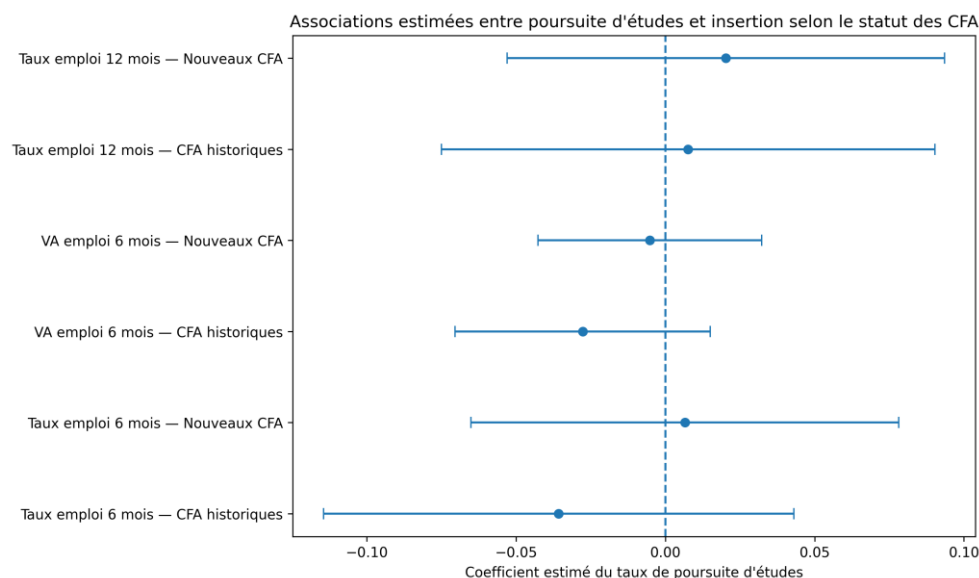


Figure 9 - Associations estimées entre poursuite d'études et insertion selon le statut des CFA

3.10 Commentaire des résultats

Le tableau 4 conduit à une lecture plus prudente que celle suggérée par les premières spécifications. Une fois introduits les effets fixes établissement et une distinction explicite entre CFA historiques et nouveaux CFA, aucune association statistiquement robuste n'apparaît entre variation du taux de poursuite d'études et variation des indicateurs d'insertion.

Pour le taux d'emploi à 6 mois, l'association estimée pour les CFA historiques est négative ($\beta = -0.0357$), mais elle n'est pas statistiquement significative. Le supplément associé aux nouveaux CFA est positif ($\beta = 0.0422$), mais lui aussi non significatif. L'association totale estimée pour les nouveaux CFA est ainsi très proche de zéro ($\beta = 0.0065$). Autrement dit, les données ne mettent pas en évidence de différence robuste entre établissements historiques et nouveaux établissements sur ce point.

Le même constat s'impose pour la valeur ajoutée d'emploi à 6 mois. L'association estimée pour les CFA historiques est négative, tandis que le différentiel propre aux nouveaux CFA est positif, mais aucun de ces coefficients n'est significatif aux seuils usuels. Le résultat suggère donc que, une fois contrôlée l'hétérogénéité fixe des établissements, le taux de poursuite d'études n'est pas associé de manière robuste aux variations de performance relative mesurées par la valeur ajoutée.

Enfin, pour le taux d'emploi à 12 mois, l'association estimée pour les CFA historiques est très faible et non significative, tout comme le différentiel pour les nouveaux CFA. L'association totale pour ces derniers demeure elle aussi non significative. Les résultats ne montrent donc pas que la poursuite d'études serait associée, à moyen terme, à des écarts d'insertion robustes au sein des établissements.

Pris ensemble, ces résultats suggèrent que les corrélations les plus marquées observées dans les spécifications les plus simples reflétaient principalement des différences transversales entre établissements. Dès lors que l'on exploite prioritairement la variation intra-CFA dans le temps et que l'on distingue les établissements historiques des nouveaux CFA, l'association entre poursuite d'études et insertion devient statistiquement fragile.

La figure 9 confirme visuellement ce diagnostic. Les intervalles de confiance associés aux coefficients estimés recouvrent systématiquement zéro, aussi bien pour les CFA historiques que pour les nouveaux CFA. Cette représentation graphique renforce l'idée selon laquelle aucune association robuste ne peut être mise en évidence entre variation de la poursuite d'études et variation des indicateurs d'insertion dans les spécifications privilégiées.

3.11 Robustesse en modèle fractionnaire

Le tableau 5 confirme cette lecture. Pour le taux d'emploi à 6 mois, l'effet *within* estimé pour les CFA historiques est négatif, mais non significatif. Le différentiel *within* associé aux nouveaux CFA n'est pas significatif non plus, et l'effet total estimé pour ces derniers demeure proche de zéro. La prise en compte explicite du caractère borné des variables de taux ne modifie donc pas substantiellement le diagnostic issu des modèles linéaires avec effets fixes établissement.

Pour le taux d'emploi à 12 mois, le constat est similaire : ni l'effet *within* pour les CFA historiques, ni le différentiel associé aux nouveaux CFA, ni l'effet total pour ces derniers n'apparaissent significatifs. Les modèles fractionnaires ne mettent donc pas en évidence d'association robuste entre variation de la poursuite d'études et variation des taux d'emploi au sein d'un même établissement.

Le seul signal plus net concerne la composante *between* du modèle fractionnaire à 6 mois, qui apparaît négative et à la limite des seuils usuels de significativité. Ce résultat suggère surtout que les CFA qui présentent, en moyenne sur la période, des niveaux plus élevés de poursuite d'études tendent aussi à afficher, en moyenne, des taux d'emploi à 6 mois légèrement plus faibles. Il ne permet toutefois pas d'inférer qu'une hausse de la poursuite d'études à l'intérieur d'un même CFA entraînerait mécaniquement une baisse de son insertion.

3.12 Lecture d'ensemble

Les résultats convergent vers un même constat. Premièrement, les associations négatives mises en évidence dans les spécifications les plus simples ne résistent pas à l'introduction d'effets fixes établissement. Deuxièmement, la distinction entre CFA historiques et nouveaux CFA ne révèle pas d'hétérogénéité robuste de ces associations. Troisièmement, les modèles fractionnaires, mieux adaptés à la nature bornée des taux d'emploi, ne conduisent pas à une interprétation différente. Quatrièmement, la robustesse par variables instrumentales

générées en interne confirme que ce diagnostic ne s'explique pas par un biais d'endogénéité de l'estimateur OLS *within*.

L'enseignement principal du papier devient dès lors plus circonscrit mais aussi plus solide : les écarts d'insertion observés entre établissements sont fortement liés à des différences structurelles entre CFA, et non à un lien robuste entre variation de la poursuite d'études et variation de l'insertion à l'intérieur d'un même établissement. Dans cette perspective, les coefficients estimés doivent être lus comme des associations conditionnelles fragiles plutôt que comme l'indice d'un mécanisme stable et généralisable.

3.13 Robustesse par variables instrumentales (Lewbel)

Afin d'évaluer directement l'ampleur du biais d'endogénéité potentiel du taux de poursuite d'études, une stratégie d'identification par variables instrumentales générées en interne est mise en oeuvre, suivant. Cette approche exploite l'hétéroscédasticité des résidus de l'équation de premier stade pour construire des instruments valides sans recourir à une source de variation externe au panel.

La condition préalable d'hétéroscédasticité est confirmée par un test de Breusch-Pagan significatif au seuil de 1 % (statistique = 44.6). Deux instruments sont retenus parmi les instruments candidats sur la base de leur force au premier stade ; la statistique de Fisher partielle associée s'élève à 150.0, ce qui écarte tout problème d'instruments faibles. Le test de suridentification de Hansen confirme la validité des instruments retenus ($J = 1.90$, $p = 0.168$).

Le tableau 6 présente les résultats de cette estimation et les compare à l'estimateur OLS *within* de référence.

Table 6 - Comparaison OLS *within* et 2SLS Lewbel — Taux d'emploi à 6 mois

Estimateur	Coefficient	Écart-type	IC 95 %
OLS <i>within</i>	-0.012	0.016	[-0.043 ; 0.020]
2SLS Lewbel	-0.049	0.076	[-0.197 ; 0.099]

Diagnostics

Breusch-Pagan = 44.6, $p < 0.01$

(hétéroscédasticité)

Fisher premier stade = 150.0, $p < 0.01$

Hansen J (suridentification) = 1.90, $p = 0.168$

Notes : erreurs standards clusterisées au niveau du CFA. Les deux estimateurs incluent des effets fixes CFA et année. Les instruments Lewbel sont construits à partir des résidus hétéroscédastiques du premier stade interagis avec les dummies année centrées ; deux instruments sont retenus sur la base de leur t-statistique au premier stade.

Le coefficient 2SLS (-0.049) est non significatif et proche du coefficient OLS *within* (-0.012), la différence entre les deux estimateurs s'élevant à -0.037, inférieure à deux fois l'écart-type maximal des deux estimateurs. Ce diagnostic informel de Hausman suggère que le biais d'endogénéité est probablement limité dans ce contexte.

Ces résultats renforcent la lecture d'ensemble du papier, l'absence d'association robuste entre poursuite d'études et insertion ne semble pas attribuable à un biais d'endogénéité de l'estimateur OLS *within*, mais reflète bien une absence de relation substantielle au sein des établissements une fois leur hétérogénéité fixe contrôlée.

Lecture des coefficients de détermination

Les niveaux de R^2 et de pseudo- R^2 ne doivent pas être surinterprétés. Dans les modèles linéaires avec effets fixes établissement, une part importante de la variance est mécaniquement absorbée par l'introduction d'un grand nombre d'effets fixes CFA. Dans les modèles fractionnaires, la logique d'interprétation diffère encore, le pseudo- R^2 ne constituant pas l'équivalent direct du coefficient de détermination des moindres carrés ordinaires.

Dans le cadre présent, l'objectif n'est pas de maximiser la performance prédictive des modèles, mais d'évaluer la robustesse des associations estimées lorsque l'on modifie à la fois le degré de contrôle de l'hétérogénéité inobservée, la composition de l'échantillon et la forme fonctionnelle retenue pour les variables dépendantes. À cet égard, l'affaiblissement des coefficients estimés dans les spécifications les plus exigeantes est plus informatif que le niveau absolu des indicateurs d'ajustement.

Au-delà de la seule significativité statistique, l'ampleur économique des coefficients estimés doit également être discutée. Dans les spécifications les plus simples comme dans les modèles plus exigeants, les coefficients associés au taux de poursuite d'études demeurent de faible magnitude. Même lorsque leur signe est négatif, ils correspondent à des variations relativement modestes des indicateurs d'insertion pour une hausse de 10 points de pourcentage du taux de poursuite d'études. Cette faible ampleur économique renforce l'idée que les associations mises en évidence ne sauraient être surinterprétées. Dans les spécifications intégrant les effets fixes CFA, les interactions selon l'ancienneté des établissements et les robustesses en modèle fractionnaire, les coefficients deviennent en outre statistiquement fragiles, ce qui conduit à conclure à l'absence d'effet économiquement substantiel et robuste de la poursuite d'études sur les indicateurs d'insertion considérés.

4. Conclusion

Cet article s'est attaché à déconstruire les déterminants de la performance d'insertion des centres de formation d'apprentis (CFA) en France, en exploitant un panel inédit issu du dispositif InserJeunes sur la période 2018-2024. En mobilisant une stratégie d'identification rigoureuse, combinant des effets fixes établissement et temporels, des modèles fractionnaires et une approche par variables instrumentales générées en interne (Lewbel, 2012), notre démarche visait à isoler l'effet des comportements de poursuite d'études sur l'insertion professionnelle à court et moyen termes.

Notre analyse empirique apporte une contribution clarificatrice à l'économie de la formation professionnelle, articulée autour de trois résultats majeurs. Premièrement, bien que l'analyse descriptive initiale suggère une association négative entre le taux de poursuite d'études et le taux d'emploi (à 6 ou 12 mois), nos estimations démontrent que cette corrélation s'évanouit totalement dès lors que l'hétérogénéité structurelle invariante des établissements est

neutralisée. Deuxièmement, l'absence de relation intra-établissement robuste, confirmée par nos tests différenciant les CFA historiques des nouveaux entrants, indique que les écarts d'insertion reflètent d'abord des disparités institutionnelles, territoriales et de spécialisation persistantes, plutôt qu'un effet d'éviction mécanique lié à la poursuite d'études. Troisièmement, le recours aux instruments de Lewbel atteste que cette absence de corrélation ne relève pas d'un biais d'endogénéité de l'estimateur OLS *within*, le coefficient 2SLS (-0.049) s'avérant non significatif et proche de l'estimation standard (-0.012). Par ailleurs, l'approche en modèle fractionnaire confirme la stabilité de ce diagnostic face à la nature bornée des taux d'emploi.

Sur le plan des politiques publiques, ces conclusions appellent à une refonte de l'évaluation institutionnelle des CFA. L'utilisation des taux bruts d'insertion comme boussole de pilotage ou outil de comparaison s'avère trompeuse, puisqu'elle impute aux établissements des performances qui relèvent largement de facteurs exogènes (environnement macroéconomique local, profil sociodémographique captif, spécialisation sectorielle). Afin d'appréhender la véritable efficacité pédagogique et institutionnelle, cette étude démontre la nécessité pour les décideurs publics de s'appuyer prioritairement sur des indicateurs de valeur ajoutée neutralisant ces variables de confusion, à l'instar de la VA emploi à 6 mois documentée par InserJeunes.

Si cette recherche clarifie l'interprétation des différentiels de placement, elle présente toutefois des limites qui dessinent de fertiles perspectives de recherche. D'une part, la nature macro-institutionnelle de la base contraint l'analyse et empêche de distinguer finement les profils individuels ou d'identifier la nature exacte des poursuites d'études (choix d'accumulation de capital humain *versus* poursuite contrainte face aux frictions du marché du travail). D'autre part, en l'absence d'une variable instrumentale externe strictement exogène, et face aux diagnostics défavorables des tentatives de modélisation dynamique (GMM) sur un panel court et déséquilibré, les coefficients estimés doivent conserver une prudence interprétative quant à leur dimension purement causale. Enfin, la mesure de l'insertion demeurant étroitement dépendante du champ statistique couvert par InserJeunes, de futurs travaux gagneraient à mobiliser des données de panel administratives individuelles pour décomposer la "boîte noire" des stratégies post-apprentissage.

En définitive, notre travail souligne l'importance décisive de l'hétérogénéité invariante des CFA dans l'analyse de leurs performances, et invite à dépasser les lectures naïves de l'insertion pour construire une évaluation microéconomique plus rigoureuse des politiques de l'alternance.

Références

- Acemoglu, D., and J.-S. Pischke. 1999. "Beyond Becker: Training in Imperfect Labour Markets." *The Economic Journal* 109 (453): 112–42.
- Angrist, J. D., and J.-S. Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Arellano, M. 2003. *Panel Data Econometrics*. Oxford: Oxford University Press.
- Arrighi, J.-J., and C. Gasquet. 2010. "La voie professionnelle et l'apprentissage : des formations qui mènent à l'emploi." *Bref CEREQ* 271.
- Bentolila, S., A. Cabrales, and M. Jansen. 2023. "Does Dual Vocational Education and Training Pay Off?" IZA Discussion Paper 16688.
- CEREQ. 2021. "Enquête Génération 2017 - Interrogation 2020 : S'insérer dans un marché du travail en crise." *Bref CEREQ* 416.
- Chetty, R., J. N. Friedman, and J. E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–632.
- DARES. 2023. "L'apprentissage en 2022 : forte progression des entrées pour la cinquième année consécutive." *DARES Résultats* 30.
- DEPP. n.d. "Document de travail InserJeunes - la valeur ajoutée des établissements sur le taux d'emploi." Ministère de l'Éducation nationale.
- DEPP. 2024. "Les indicateurs du dispositif InserJeunes : définitions, méthodes de calcul et règles de diffusion." Ministère de l'Éducation nationale.
- Greene, W. H. 2018. *Econometric Analysis*. London: Pearson.
- Hanushek, E. A. 1986. "The Economics of Schooling: Production and Efficiency in Public Schools." *Journal of Economic Literature* 24 (3): 1141–77.
- Kemper, J., A. Au Yong Lyn, P. McDonald, and U. Renold. 2026. "Labor Market Outcomes of Formal Vocational Education and Training: a Meta-Analysis." *Education Economics* 34 (2): 267–91.
- Lewbel, A. 2012. "Using Heteroscedasticity to Identify and Estimate Mismeasured and Endogenous Regressor Models." *Journal of Business & Economic Statistics* 30 (1): 67–80.
- McGuinness, S., L. Kelly, A. Devlin, and A. Whelan. 2025. "Vocational Education, Earnings and Job Satisfaction in Europe." *Education Economics*.
- Michaelis, C., and S. Findeisen. 2024. "Long-term Effects of Different VET-to-Labor Market Transition Patterns on Subjective Well-Being." *Zeitschrift für Erziehungswissenschaft* 27: 393–419.

Ministère de l'Éducation nationale. n.d. "InserJeunes Apprentissage par centre de formation d'apprentis (CFA)." Portail des données de l'Éducation nationale.

Ministère de l'Éducation nationale. 2021. "InserJeunes : un service d'aide à l'orientation des jeunes en voie professionnelle."

Ryan, P. 2001. "The School-to-Work Transition: A Cross-National Perspective." *Journal of Economic Literature* 39 (1): 34–92.

Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.