

**ÉTUDE DE DONNÉES
DE PANEL
(ENQUÊTES 2013-2017)
PROMOTIONS 2012-2013-2014**

RAPPORT D'ANALYSE – SEPTEMBRE 2017

DIRECTION DES RELATIONS ENTREPRISES – ARNAUD TROLLÉ

Table des matières

Table des figures	V
Liste des tableaux	VII
1 Préambule	1
2 Méthodologie	3
2.1 Critères de choix	4
2.2 Indicateurs pertinents	4
3 Résultats	7
3.1 Analyse des indicateurs nominaux	8
3.1.1 La situation professionnelle	8
3.1.2 Le type de contrat	9
3.1.3 Le type d'entreprise	11
3.1.4 Le lieu de travail	13
3.1.5 L'adéquation de l'emploi avec le secteur disciplinaire de la formation	15
3.2 Analyse des indicateurs métriques	16
4 Conclusions	19
Annexes	21
A Notions pré-requises	23
A.1 Données de panel / Données longitudinales	23
A.2 Modèles de régression multi-niveaux	23
B Méthodes d'analyse	25
B.1 Analyse des indicateurs nominaux	25
B.2 Analyse des indicateurs métriques	25
Bibliographie	27

Table des figures

3.1	Situation professionnelle - Les distributions d'états, par promotion et par genre. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).	8
3.2	Type de contrat - Les distributions d'états, par promotion et par genre. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).	10
3.3	Type d'entreprise - Séquences d'états par promotion. En ordonnée : Fréquence cumulée en %.	11
3.4	Type d'entreprise - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).	12
3.5	Lieu de travail - Séquences d'états par promotion. En ordonnée : Fréquence cumulée en %.	13
3.6	Lieu de travail - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).	14
3.7	Adéquation de l'emploi avec le secteur disciplinaire de la formation - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).	15
3.8	Rémunération brute annuelle (avec primes) - Trajectoires individuelles groupées par promotion et par genre.	16
3.9	Rémunération brute annuelle (avec primes) - Trajectoires moyennes estimées pour les femmes et les hommes des promotions 2012, 2013 et 2014.	17

Liste des tableaux

1.1	Taux de réponse aux 3 enquêtes, pour les promotions 2012, 2013 et 2014.	1
1.2	Nombres de diplômé-e-s qui ont répondu au moins à une enquête et qui ont répondu aux 3 enquêtes, pour les promotions 2012, 2013 et 2014.	1

Chapitre 1

Préambule

Ce document présente les résultats de l'analyse des données de panels de diplômé-e-s 2012, 2013 et 2014 de l'INSA Lyon.

En 2017, pour la seconde fois, l'enquête d'insertion professionnelle a concerné 5 promotions de diplômé-e-s de l'INSA Lyon : 2012, 2013, 2014, 2015 et 2016. Ainsi, les promotions 2012, 2013 et 2014 ont été interrogées trois fois, i.e. une première fois à la sortie de l'École (en 2013, 2014 et 2015, respectivement), une seconde fois en 2016, et une troisième fois en 2017.

Le Tableau 1.1 présente, pour les promotions 2012, 2013 et 2014, les taux de réponse aux 3 enquêtes.

		Enquete n°1		Enquete n°2		Enquete n°3	
		Année	Taux de réponse	Année	Taux de réponse	Année	Taux de réponse
Promotion	2012	2013	81,9%	2016	52,8%	2017	40,0%
	2013	2014	85,2%	2016	54,8%	2017	40,6%
	2014	2015	85,2%	2016	64,6%	2017	51,1%

Tableau 1.1: Taux de réponse aux 3 enquêtes, pour les promotions 2012, 2013 et 2014.

Le Tableau 1.2 présente, pour les promotions 2012, 2013 et 2014, les nombres de diplômé-e-s qui ont répondu au moins à une enquête et qui ont répondu aux 3 enquêtes.

		Diplômé-e-s qui ont répondu au moins à une enquête		Diplômé-e-s qui ont répondu aux 3 enquêtes	
		Nombre	% de l'effectif	Nombre	% de l'effectif
Promotion	2012	576	60,1%	251	26,2%
	2013	565	59,2%	264	27,6%
	2014	609	60,8%	392	39,1%

Tableau 1.2: Nombres de diplômé-e-s qui ont répondu au moins à une enquête et qui ont répondu aux 3 enquêtes, pour les promotions 2012, 2013 et 2014.

Dans le cas le plus défavorable, on compte 251, 264 et 392 diplômé-e-s 2012, 2013 et 2014, respectivement, qui ont répondu aux 3 enquêtes. Ces tailles d'échantillon, qui restent substantielles, garantissent *a priori* des résultats statistiques robustes.

L'étude vise à analyser les données de panels de diplômé-e-s 2012, 2013 et 2014, dans l'objectif de dresser et comparer leurs trajectoires professionnelles, mesurées selon différents in-

dicateurs (situation professionnelle, type de contrat, type d'entreprise, localisation de l'emploi, etc.).

Chapitre 2

Méthodologie

Ce chapitre présente la méthodologie adoptée pour analyser les données de panels de diplômé-e-s 2012, 2013 et 2014.

Pour aider à la compréhension, il est vivement recommandé au lecteur de se reporter à l'annexe A, où certaines notions pré-requises sont rappelées.

Les trajectoires professionnelles peuvent être mesurées selon de multiples indicateurs. En effet, le questionnaire de l'enquête CGE a été conçu de façon à collecter un maximum d'informations relatives à la situation professionnelle des diplômé-e-s, leur rémunération, leur appréciation de l'École, etc. Il s'agit donc d'opérer un choix d'indicateurs pertinents. Tout d'abord, les critères de choix sont exposés (Cf. Section 2.1), puis les indicateurs jugés comme pertinents sont listés (Cf. Section 2.2).

Enfin, les méthodes d'analyse utilisées pour dresser ces trajectoires professionnelles sont présentées dans l'Annexe B.

2.1 Critères de choix

Deux critères ont guidé le choix d'indicateurs pertinents :

- i) *La robustesse statistique.* Certes, les tailles d'échantillon, i.e. 251, 264 et 392 pour les promotions 2012, 2013 et 2014, respectivement, si l'on considère les diplômé-e-s qui ont répondu aux 3 enquêtes, sont *a priori* substantielles. Mais ces tailles d'échantillon ne valent que pour une seule question, pour laquelle une réponse était obligatoire : "Quelle est votre situation ?". Pour les autres questions, la réponse n'était pas obligatoire. Ainsi, les tailles d'échantillon, i.e. les nombres de diplômé-e-s 2012, 2013 et 2014 qui ont répondu à une autre question donnée, peuvent dramatiquement chuter, jusqu'à rendre un indicateur peu robuste.
- ii) *L'intérêt métier.* L'indicateur doit revêtir un intérêt particulier vis-à-vis des questionnements de la Direction de l'INSA Lyon sur le devenir de ses diplômé-e-s.

En outre, on relève deux types d'indicateurs, selon le niveau d'information :

- i) Les *indicateurs nominaux.* Pour ces indicateurs, le niveau d'information est nominal. Généralement, l'indicateur implique le choix d'un item verbal (dans une liste) en réponse à une question fermée. Par exemple, à la question "Quelle est votre situation ?", les réponses verbales possibles étaient : "En activité professionnelle", "En volontariat", "En recherche d'emploi", etc.
- ii) Les *indicateurs métriques.* Pour ces indicateurs, le niveau d'information est métrique. L'indicateur implique la précision d'un montant, ou une notation sur une échelle numérique.

2.2 Indicateurs pertinents

Au final, les indicateurs qui ont été jugés comme pertinents sont :

- Indicateurs nominaux :
 - *Situation professionnelle.* Les *modalités* (ou *états*) possibles, au nombre de 6, sont : "En activité professionnelle", "En volontariat", "En recherche d'emploi", "En poursuite d'études (y compris préparation de concours)", "En thèse" et "Sans activité volontairement".
 - *Type de contrat.* Les états possibles, au nombre de 3, sont : "CDI", "CDD" et "Intérim et autre".
 - *Type d'entreprise.* Les états possibles, au nombre de 2, sont : "ETI/Grande entreprise" et "TPE/PME".
 - *Lieu de travail.* Les états possibles, au nombre de 3, sont : "Province", "Île-de-France" et "Étranger".
 - *Adéquation de l'emploi avec le secteur disciplinaire de la formation.* Les états possibles, au nombre de 2, sont : "Oui" et "Non".
- Indicateurs métriques :

- *Rémunération brute annuelle (avec primes)*. Cet indicateur a impliqué la saisie d'un montant (en Euros).

Chapitre 3

Résultats

Ce chapitre présente les résultats de l'analyse des données de panels de diplômé-e-s 2012, 2013 et 2014, opérée suivant la méthodologie abordée dans le chapitre 2 et l'Annexe B.

Tout d'abord, les résultats de l'analyse des indicateurs nominaux sont présentés (Cf. Section 3.1). Pour rappel, les indicateurs nominaux jugés comme pertinents sont au nombre de 5 : la situation professionnelle, le type de contrat, le type d'entreprise, le lieu de travail et l'adéquation de l'emploi avec le secteur disciplinaire de la formation.

Ensuite, les résultats de l'analyse des indicateurs métriques sont relatés (Cf. Section 3.2). Pour rappel, un seul indicateur métrique a été étudié : la rémunération brute annuelle (avec primes).

3.1 Analyse des indicateurs nominaux

3.1.1 La situation professionnelle

La Figure 3.1 présente les distributions d'états, par promotion et par genre.

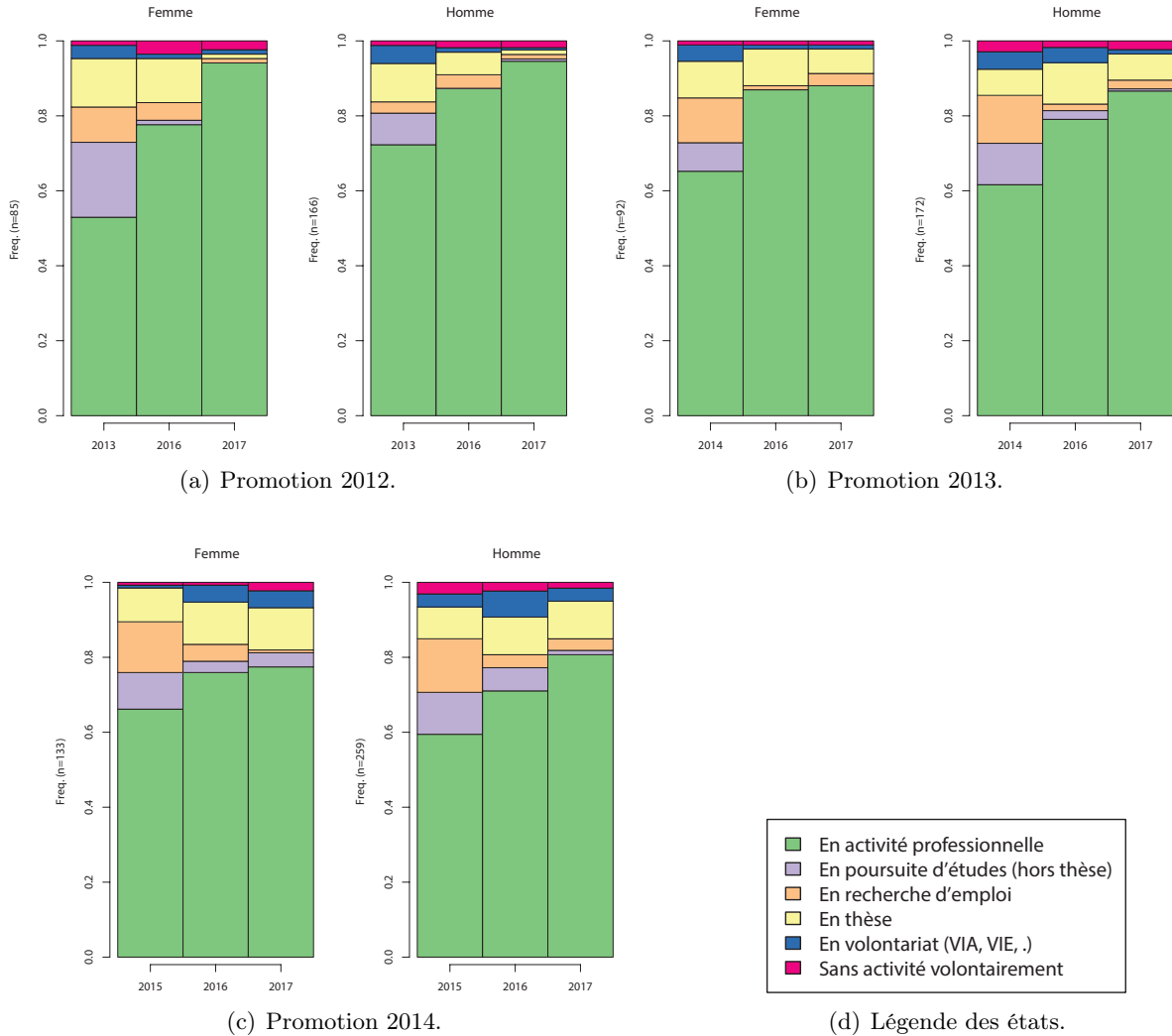


Figure 3.1: Situation professionnelle - Les distributions d'états, par promotion et par genre. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).

Tout d'abord, observons les distributions d'états promotion par promotion. Pour la promotion 2012, la proportion de femmes en activité professionnelle en 2013 (i.e. soit un an après la sortie de l'École) est nettement inférieure à la proportion d'hommes dans la même situation. Cinq ans après la sortie de l'École (i.e. en 2017), les proportions de femmes et d'hommes en activité professionnelle sont équivalentes. Pour les promotions 2013 et 2014, à l'inverse, la proportion de femmes en activité professionnelle un an après la sortie de l'École (2014 et 2015, respectivement) est supérieure à la proportion d'hommes dans la même situation. Pour ces 2 promotions, respectivement 4 et 3 ans après la sortie de l'École, les proportions de femmes et

d'hommes en activité professionnelle deviennent similaires.

Au travers des 3 promotions, on retrouve certaines tendances logiques : les proportions de femmes ou d'hommes en recherche d'emploi, ainsi que les proportions de femmes ou d'hommes en poursuite d'études (hors thèse), diminuent avec le nombre d'années après diplôme.

Enfin, concernant les proportions de femmes ou d'hommes en thèse, on observe une tendance tout aussi logique : i) pour la promotion 2012, cinq ans après la sortie de l'École (i.e. en 2017), ces proportions sont très faibles, marquant ainsi l'achèvement de la thèse en 3 voire 4 ans pour la quasi-totalité de ces diplômé-e-s ; ii) pour la promotion 2014, interrogée 3 années consécutives, ces proportions restent stables entre 2015 et 2017, puisque les diplômé-e-s concerné-e-s n'ont pas encore achevé leur thèse.

3.1.2 Le type de contrat

La Figure 3.2 présente les distributions d'états, par promotion et par genre.

Tout d'abord, examinons encore les distributions d'états promotion par promotion. Pour la promotion 2012, la proportion de femmes en CDI en 2013 est largement inférieure à la proportion d'hommes en CDI. Cinq ans après la sortie de l'École (i.e. en 2017), le retard est quasiment rattrapé : les proportions de femmes et d'hommes en CDI deviennent similaires. Pour les promotions 2013 et 2014, on observe des tendances quasi-similaires, si ce n'est que l'écart entre les proportions de femmes et d'hommes en CDI est initialement (i.e. en 2014 et 2015, respectivement) moindre. Par ailleurs, notons que quasiment tous les hommes sont en CDI 2 et 3 ans après leur sortie de l'École, respectivement pour les promotions 2013 et 2014.

Globalement, plus l'année de promotion est récente, plus les proportions de femmes et d'hommes en CDI sont élevées un an après la sortie de l'École. En outre, sans surprise, ces proportions augmentent avec le nombre d'années après diplôme, et ce quelle que soit l'année de promotion.

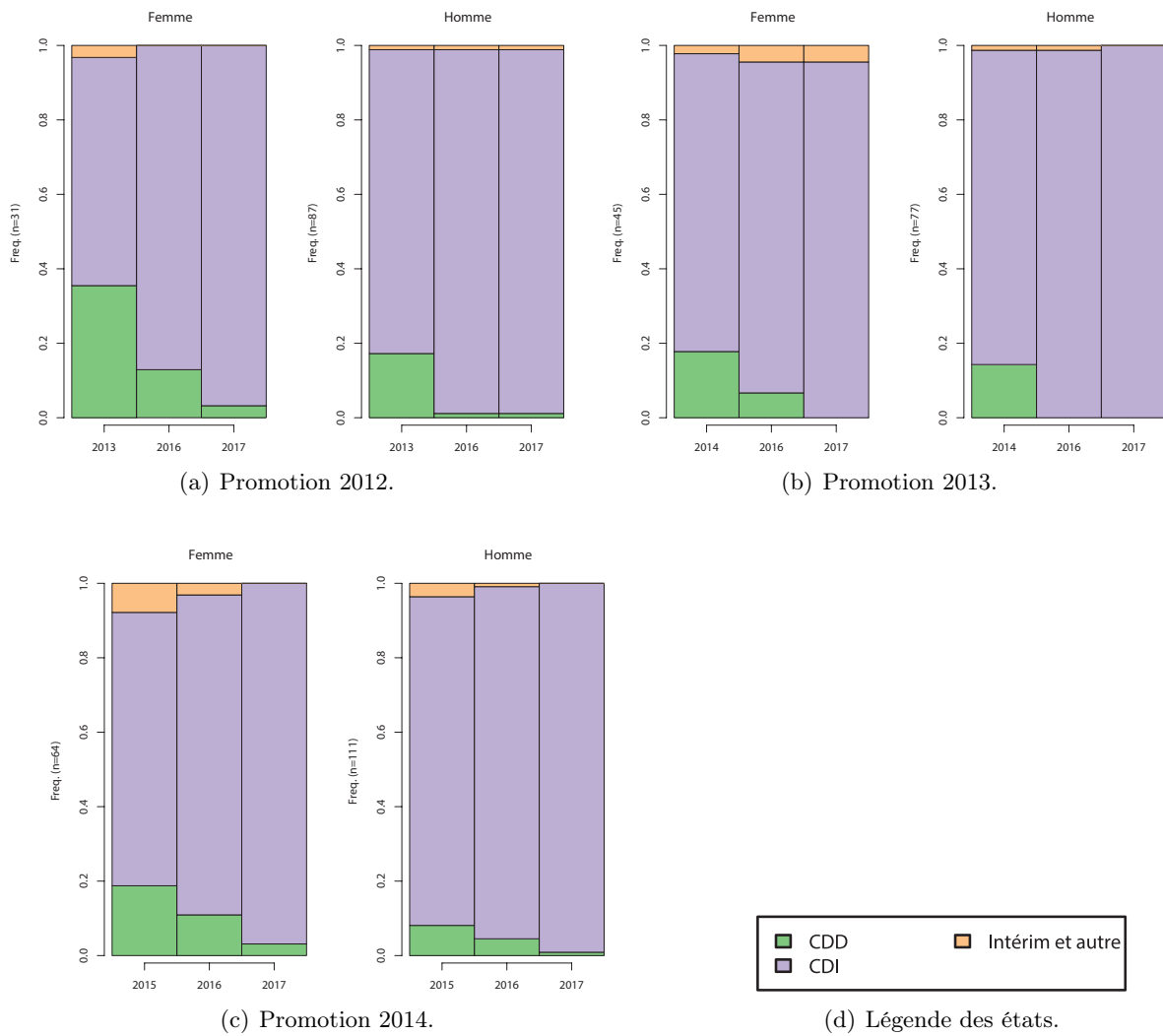


Figure 3.2: Type de contrat - Les distributions d'états, par promotion et par genre. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).

3.1.3 Le type d'entreprise

La Figure 3.3 présente les séquences d'états par promotion.

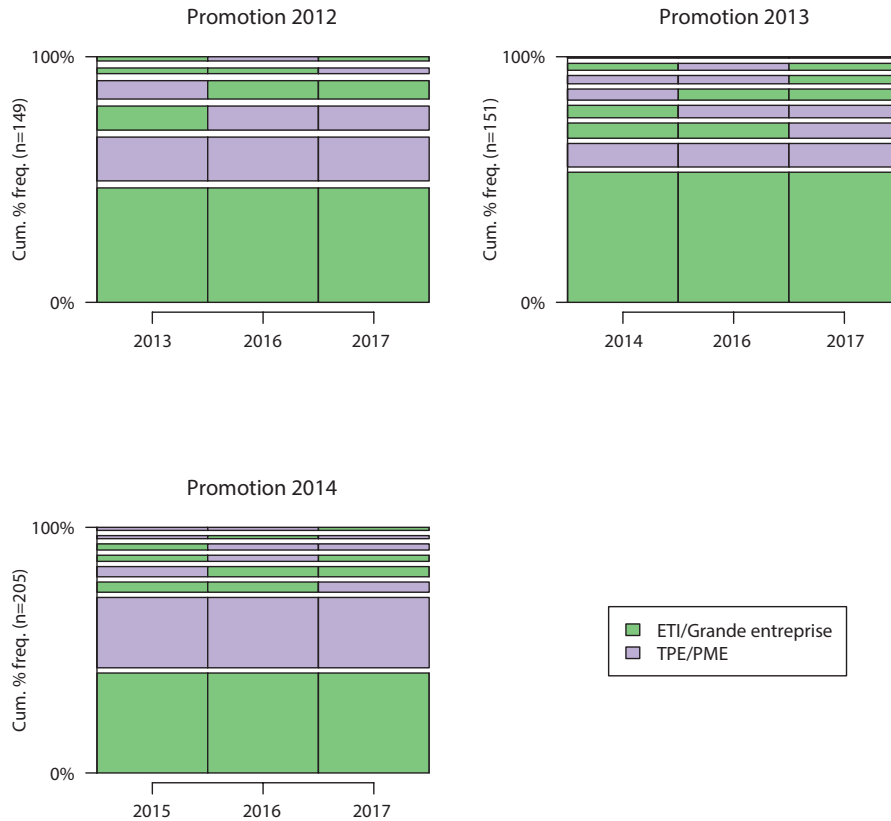


Figure 3.3: Type d'entreprise - Séquences d'états par promotion. En ordonnée : Fréquence cumulée en %.

Quelle que soit la promotion, les 2 séquences les plus fréquentes correspondent : i) aux diplômé-e-s qui travaillent dans une ETI/Grande entreprise et qui restent dans une telle structure, et ii) aux diplômé-e-s qui travaillent dans une TPE/PME et qui restent dans une telle structure. Il est intéressant de noter que, quelle que soit la promotion, la 3^{ème} séquence la plus fréquente correspond à des diplômé-e-s qui travaillaient dans une ETI/Grande entreprise mais qui, à un moment donné, sont passés dans une TPE/PME (et y sont resté-e-s). Le chemin inverse (i.e. le passage d'une TPE/PME à une ETI/Grande entreprise) existe, mais dans une moindre mesure.

La Figure 3.4 présente les distributions d'états par promotion.

Concernant les promotions 2012 et 2013, avec le nombre d'années après diplôme, on observe une augmentation de la proportion de diplômé-e-s qui travaillent dans une TPE-PME ; pour la promotion 2014, cette proportion reste assez stable.

En outre, alors que, quelle que soit l'année d'enquête, la proportion de diplômé-e-s qui travaillent dans une TPE-PME est moindre pour la promotion 2013 relativement à la promotion 2012, cette proportion est plus élevée pour la promotion 2014 relativement à la promotion 2012.

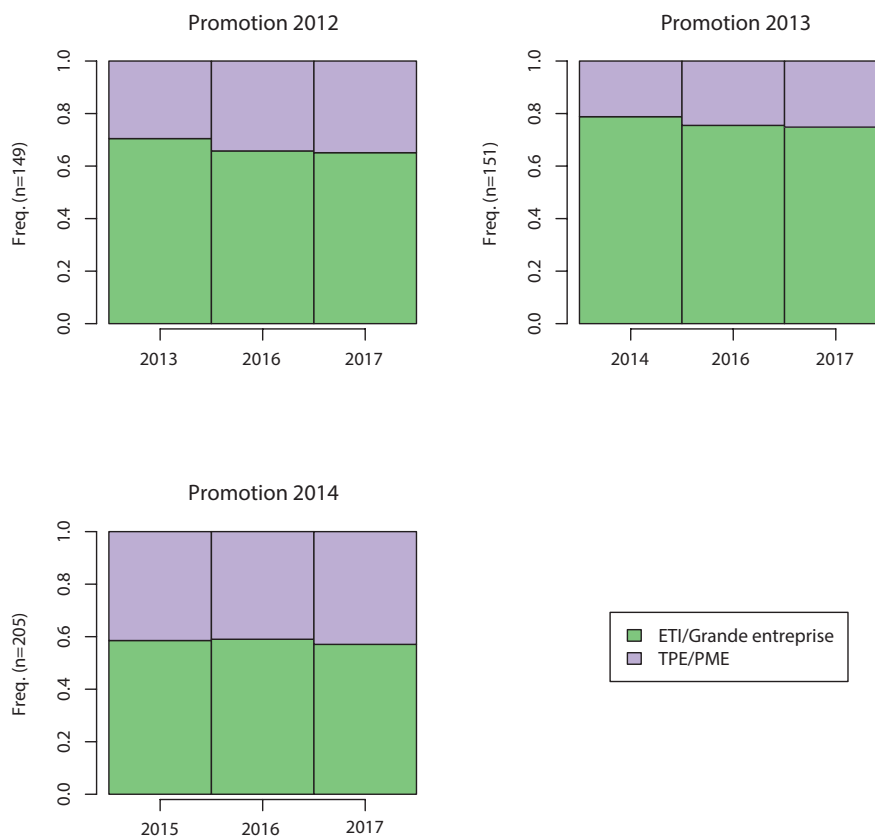


Figure 3.4: Type d'entreprise - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).

3.1.4 Le lieu de travail

La Figure 3.5 présente les 10 séquences d'états les plus fréquentes, par promotion.

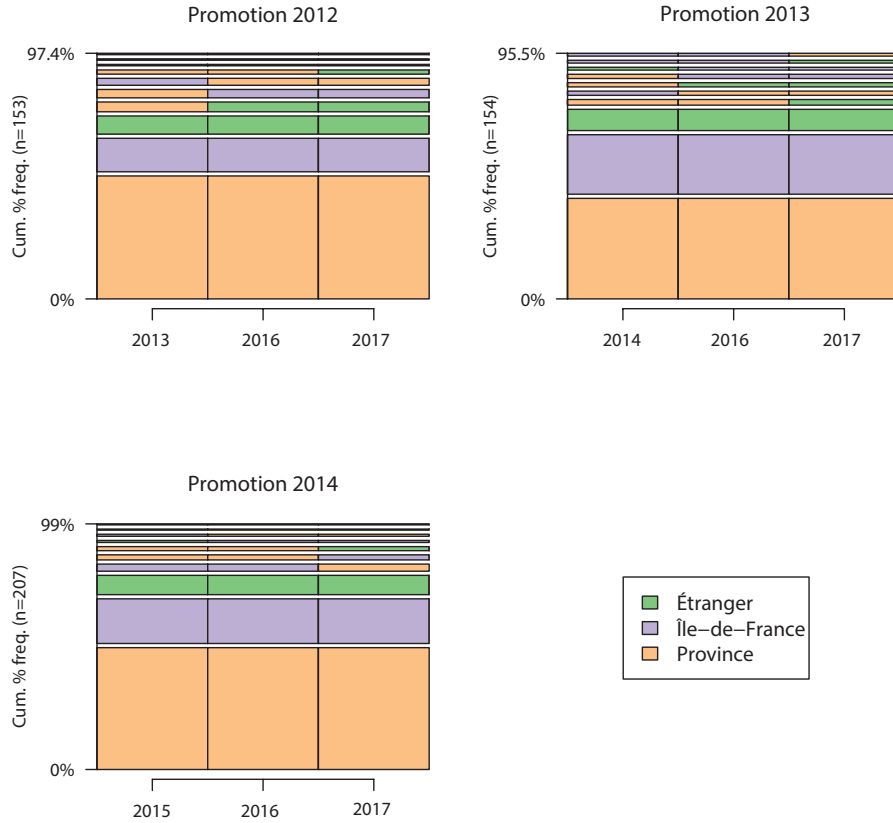


Figure 3.5: Lieu de travail - Séquences d'états par promotion. En ordonnée : Fréquence cumulée en %.

Quelle que soit la promotion, les 3 séquences les plus fréquentes correspondent aux diplômé-e-s qui travaillent, respectivement, en Province, en Île-de-France et à l'étranger, et qui ne changent pas de lieu de travail. Néanmoins, quelle que soit la promotion, on observe, en nombre non négligeable, des transitions d'un lieu de travail à un autre, ce qui dénote une certaine mobilité géographique pour une frange des diplômé-e-s.

La Figure 3.6 présente les distributions d'états par promotion.

Pour les promotions 2012 et 2013, la proportion de diplômé-e-s qui travaillent à l'étranger augmente avec le nombre d'années après diplôme, tandis que la proportion de diplômé-e-s qui travaillent en Province diminue ; la proportion de diplômé-e-s qui travaillent en Île-de-France reste stable. Concernant la promotion 2014, les proportions de diplômé-e-s qui travaillent à l'étranger, en Province et en Île-de-France restent relativement stables.

En outre, quelle que soit l'année d'enquête, la proportion de diplômé-e-s qui travaillent en Île-de-France reste plus élevée pour la promotion 2013, relativement aux promotions 2012 et 2014.

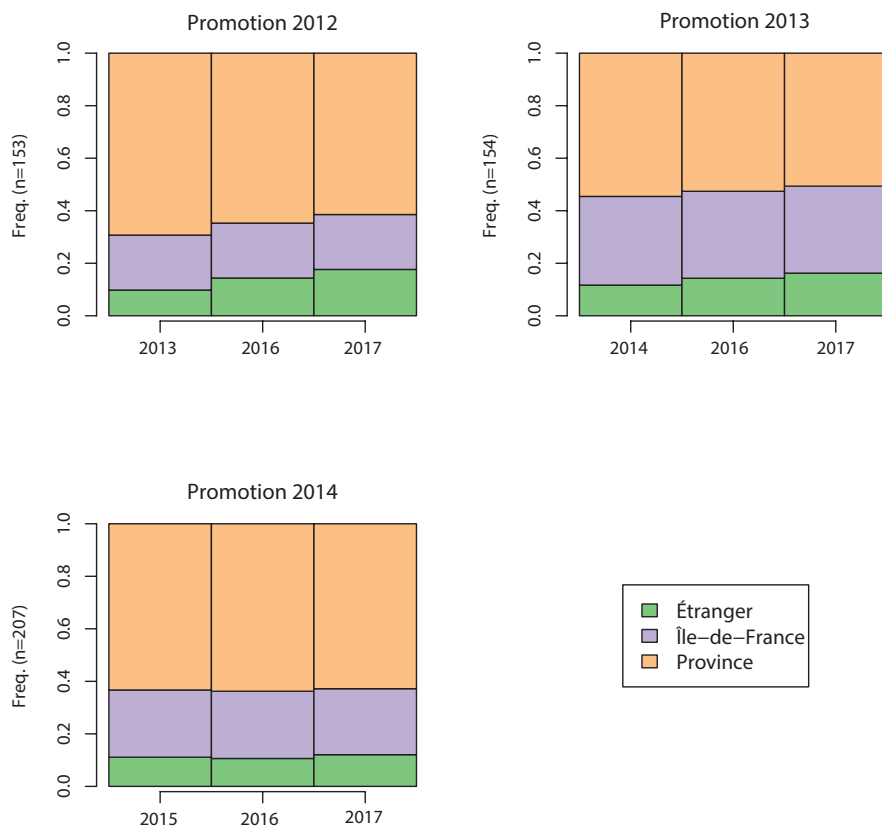


Figure 3.6: Lieu de travail - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).

3.1.5 L'adéquation de l'emploi avec le secteur disciplinaire de la formation

La Figure 3.7 présente les distributions d'états par promotion.

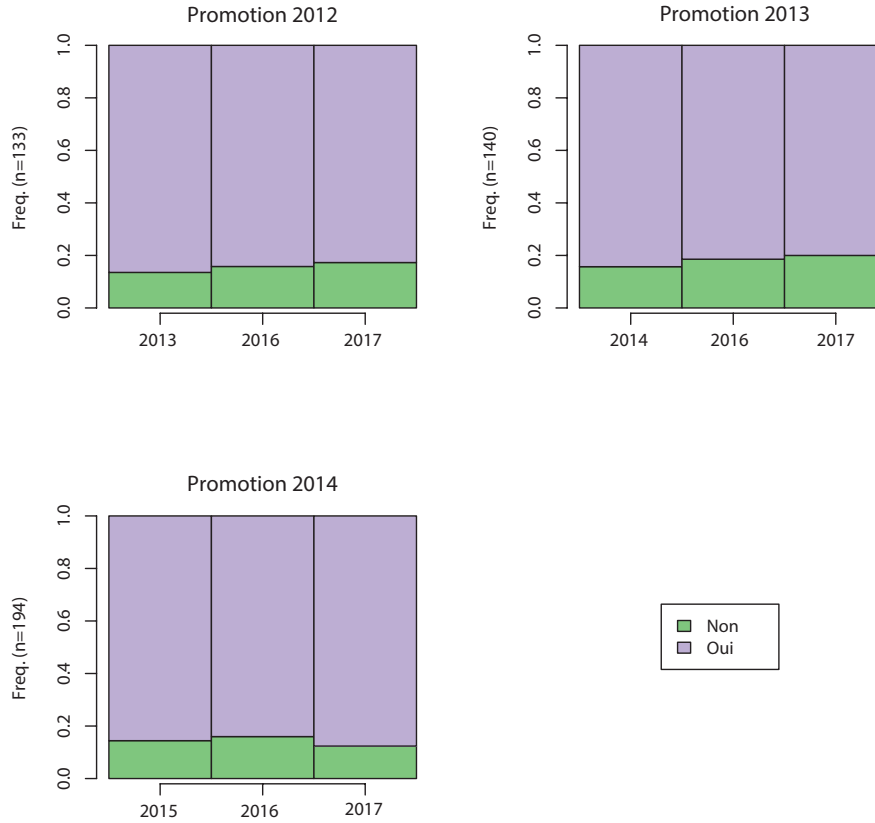


Figure 3.7: Adéquation de l'emploi avec le secteur disciplinaire de la formation - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).

Concernant les promotions 2012 et 2013, avec le nombre d'années après diplôme, on note une légère augmentation de la proportion de diplômé-e-s qui ne jugent plus leur emploi en adéquation avec le secteur disciplinaire de leur formation, i.e. au fil du temps, les diplômé-e-s des plus anciennes promotions ont tendance à s'écarter du secteur disciplinaire de leur formation. Pour la promotion 2014, interrogée ces 3 dernières années, cette proportion reste assez stable.

3.2 Analyse des indicateurs métriques

La Figure 3.8 montre, pour la rémunération brute annuelle (exprimée en kEuros constants), les trajectoires individuelles, groupées par promotion et par genre. À quelques exceptions près, la rémunération des diplômé-e-s a – sans surprise – augmenté au fil des années après leur diplôme.

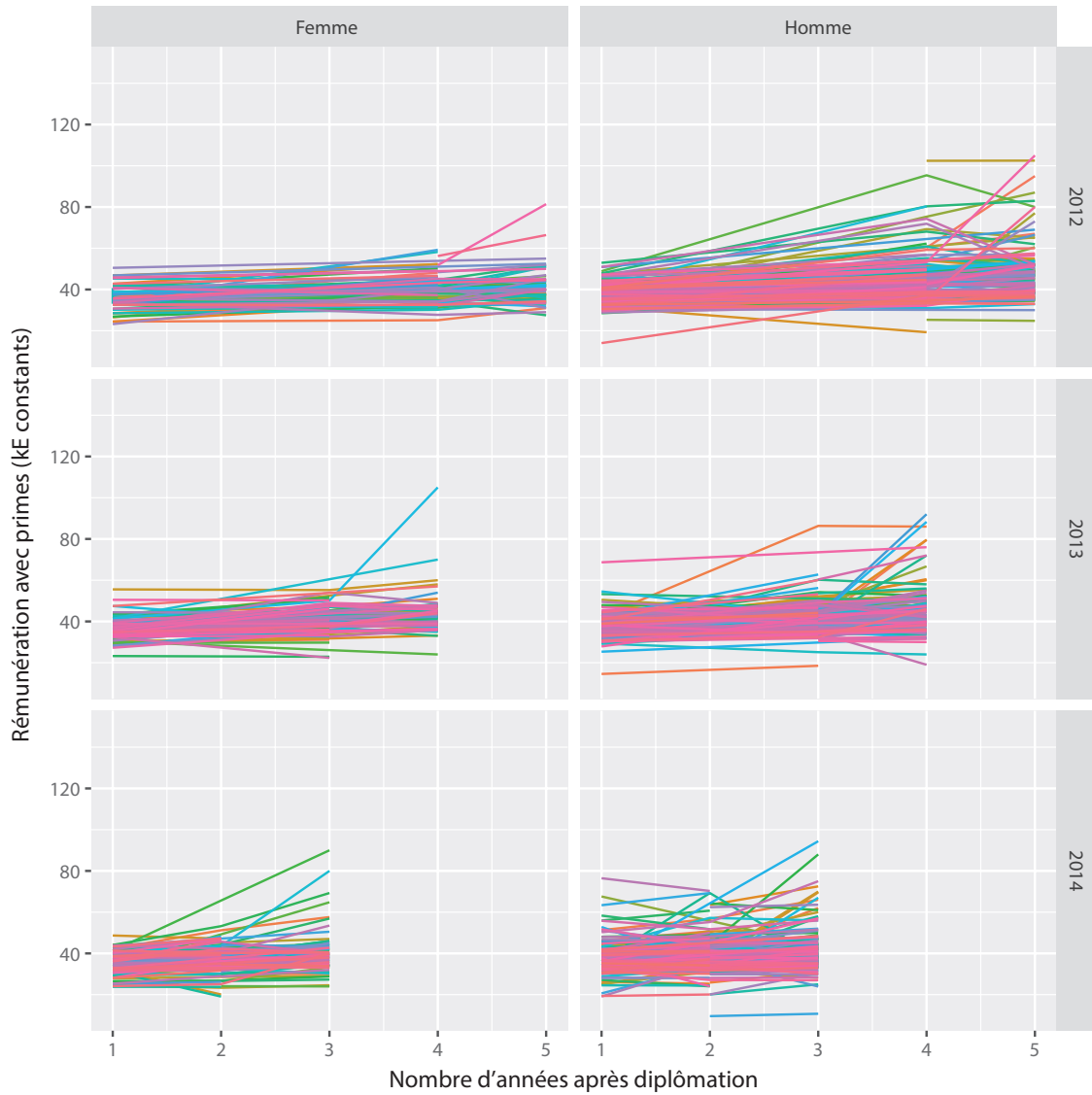


Figure 3.8: Rémunération brute annuelle (avec primes) - Trajectoires individuelles groupées par promotion et par genre.

Toujours pour la rémunération brute annuelle (exprimée en kEuros constants), la Figure 3.9 montre les trajectoires moyennes estimées, pour les femmes et les hommes des promotions 2012, 2013 et 2014, par le modèle de régression multi-niveaux, avec leur intervalle de confiance à 95%.

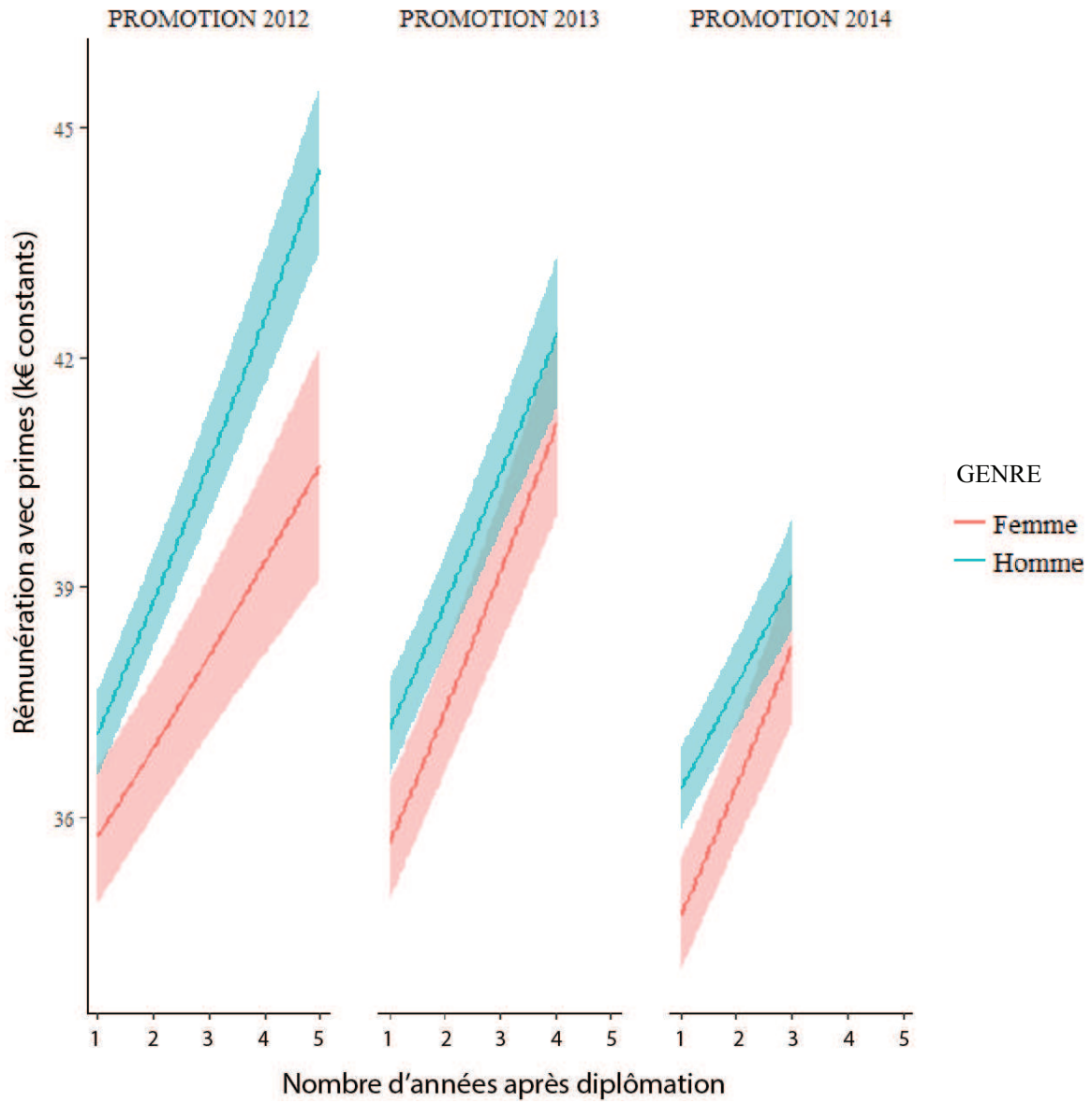


Figure 3.9: Rémunération brute annuelle (avec primes) - Trajectoires moyennes estimées pour les femmes et les hommes des promotions 2012, 2013 et 2014.

On note que :

- i) Sans surprise, la rémunération brute annuelle augmente avec le nombre d'années après diplôme, quelle que soit la promotion et quel que soit le genre. Cette augmentation est significative, d'après le résultat du test statistique sur le coefficient de régression de la variable T_{ti} (i.e. β_{10} , Cf. Section B.2)
- ii) Un an après la sortie de l'École, la rémunération brute annuelle des femmes, quelle que soit l'année de promotion, est inférieure à celle des hommes. Cette différence de rémunération initiale est significative, d'après le résultat du test statistique sur le coefficient de régression de la variable G_i (i.e. β_{01} , Cf. Section B.2).
- iii) Un an après la sortie de l'École, la rémunération brute annuelle des diplômé-e-s 2014, quel que soit le genre, est inférieure à celle des diplômé-e-s 2012 et 2013. Cette différence de rémunération initiale est significative, d'après le résultat du test statistique sur le coefficient de régression de la variable P_i (i.e. β_{02} , Cf. Section B.2).
- iv) L'augmentation de la rémunération des femmes de la promotion 2014 en fonction du nombre d'années après diplôme est plus importante, relativement aux femmes des promotions 2012 et 2013. Cette différence d'évolution de rémunération est significative, d'après le résultat du test statistique sur le coefficient de régression de la variable P_i (i.e. β_{12} , Cf. Section B.2).
- v) *A contrario*, l'augmentation de la rémunération des hommes de la promotion 2014 en fonction du nombre d'années après diplôme est moindre, relativement aux hommes des promotions 2012 et 2013. Cette différence d'évolution de rémunération est significative, d'après le résultat du test statistique sur le coefficient de régression de l'interaction $G_i \times P_i$ (i.e. β_{13} , Cf. Section B.2).

Relativement aux points iv) et v), il est intéressant de relever que, tandis que les trajectoires des femmes et des hommes des promotions 2012 et 2013 en termes de rémunération brute annuelle sont parallèles (pour la promotion 2013) voire divergent (pour la promotion 2012), les trajectoires des femmes et des hommes de la promotion 2014 convergent, i.e. seules les femmes de la promotion 2014 sont en mesure de rattraper leur retard sur les hommes diplômés la même année. Cette tendance est à confirmer par les résultats des 2 prochaines enquêtes sur la promotion 2014.

Chapitre 4

Conclusions

Dans ce document, des données de panels de diplômé-e-s 2012, 2013 et 2014 de l'INSA Lyon ont été étudiées. Elles concernaient 6 indicateurs de leur situation professionnelle, issus de l'enquête CGE. Parmi eux, figuraient 5 indicateurs nominaux, à savoir la situation professionnelle, le type de contrat, le type d'entreprise, le lieu de travail et l'adéquation de l'emploi avec le secteur disciplinaire de la formation, et 1 indicateur métrique, i.e. la rémunération brute annuelle (avec primes).

Voici les résultats principaux :

- Concernant la situation professionnelle, il a été observé que, malgré des différences notables un an après la sortie de l'École, les femmes et les hommes sont en activité professionnelle en proportions équivalentes, 5, 4 et 3 ans après la sortie de l'École, pour les promotions 2012, 2013 et 2014, respectivement.
- Concernant le type de contrat, alors qu'un an après la sortie de l'École, la proportion de femmes en CDI est nettement inférieure à celle des hommes en CDI, les femmes rattrapent en très grande partie leur retard sur les hommes, mais tout de même 5, 4 et 3 ans après la sortie de l'École, pour les promotions 2012, 2013 et 2014, respectivement.
- Concernant le type d'entreprise, quelle que soit la promotion, on observe, en nombre non négligeable, des diplômé-e-s qui travaillaient dans une ETI/ Grande entreprise passer à un moment donné dans une TPE/PME (et y rester). Le chemin inverse (i.e. le passage d'une TPE/PME à une ETI/Grande entreprise) existe, mais dans une moindre mesure. En outre, quelle que soit l'année d'enquête, les diplômé-e-s 2014 ont davantage tendance à travailler dans une TPE-PME, relativement aux diplômé-e-s 2012 et 2013.
- Concernant le lieu de travail, quelle que soit la promotion, on observe, en nombre non négligeable, des transitions d'un lieu de travail à un autre, ce qui dénote une certaine mobilité géographique pour une frange des diplômé-e-s. Par ailleurs, pour les 2 plus anciennes promotions (i.e. 2012 et 2013), la proportion de diplômé-e-s qui travaillent à l'étranger augmente avec le nombre d'années après diplôme, tandis que la proportion de diplômé-e-s qui travaillent en Province diminue.
- Concernant l'adéquation de l'emploi avec le secteur disciplinaire de la formation, on observe qu'au fil du temps, les diplômé-e-s des 2 plus anciennes promotions (i.e. 2012 et 2013) ont tendance à s'écarter du secteur disciplinaire de leur formation.

- Concernant la rémunération brute annuelle (avec primes), il a été relevé que : i) il existe un écart initial (i.e. un an après la sortie de l'École) important entre les femmes et les hommes, et ce quelle que soit la promotion, et ii) seules les trajectoires des femmes et des hommes de la promotion 2014 convergent, i.e. seules les femmes de la promotion 2014 sont en mesure de rattraper leur retard sur les hommes diplômés la même année.

Ces premières tendances sont à confirmer. À cette fin, les données complémentaires collectées via la prochaine enquête CGE (i.e. l'enquête 2018) en ce qui concerne la promotion 2013 (la dernière qui concernera cette promotion), via les 2 prochaines enquêtes CGE (i.e. les enquêtes 2018 et 2019) quant à la promotion 2014 (les deux dernières qui concerneront cette promotion) seront grandement utiles. Notons qu'il n'est plus possible de collecter de données complémentaires pour la promotion 2012 ; l'enquête 2017 était la dernière à la concerner. Les données déjà collectées relativement à cette promotion n'en demeureront pas moins un acquis fort utile à des fins de comparaison.

Annexes

Annexe A

Notions pré-requises

A.1 Données de panel / Données longitudinales

Les *données de panel* sont un type de *données longitudinales*, i.e. des données collectées à différents points dans le temps. Il existe 3 types de données longitudinales [1] :

- Les *séries temporelles* (*time series* en anglais). Par série temporelle, on désigne de nombreuses observations dans le temps (T grand) de quelques *mêmes* unités statistiques (N petit).
- Les *données transversales répétées* (*repeated cross-sectional data* en anglais). De telles données concernent *différents* échantillons de nombreuses unités statistiques (N grand), observés à différents points dans le temps.
- Les *données de panel* (*panel data* en anglais). Ces données correspondent à 2 observations ou plus dans le temps (T petit) de nombreuses *mêmes* unités statistiques (N grand).

La collecte de données de panel est tout indiquée pour i) décrire le *changement* dans le temps (e.g. changement de comportement, relations sociales, trajectoires de carrière, etc.), et ii) estimer des modèles *causaux*.

Les réponses données par les diplômé-e-s des promotions 2012, 2013 et 2014 représentent ainsi des données de panel.

A.2 Modèles de régression multi-niveaux

Les modèles de régression multi-niveaux (ou *hiérarchiques*) sont adaptés à l'analyse de données avec une structure hiérarchique, telles les données de panel dont nous disposons. Les données de panel peuvent être vues comme des données hiérarchiques, avec des mesures répétées (dans le temps) subordonnées à des individus [2]. Dans sa forme la plus simple, le modèle hiérarchique est à 2 niveaux, avec une série de mesures répétées au niveau le plus bas, et les individus au niveau le plus haut. Les mesures répétées peuvent être prises à des occasions fixes ou variables, comme dans notre cas. La modélisation multi-niveaux ne nécessite pas de données balancées, i.e. ce n'est pas un problème que le nombre de mesures répétées ne soit pas le même pour l'ensemble des individus.

Relativement à la régression classique¹, la régression multi-niveaux n'implique pas de tests

¹Pour rappel, en régression classique (i.e. à un seul niveau), on cherche à relier une variable dépendante à un ensemble de variables explicatives.

statistiques qui reposent fortement sur l'hypothèse d'indépendance des observations (qui n'est pratiquement jamais valable pour des données hiérarchiques), et permet d'obtenir des estimations non biaisées des erreurs-types² des coefficients de la régression, contrairement à la régression classique pour laquelle ces estimations sont souvent beaucoup trop faibles, aboutissant à des tests statistiques avec des résultats douteusement significatifs [2].

Pour des données de panel, au niveau le plus bas, le modèle multi-niveaux s'écrit [2] :

$$Y_{ti} = \pi_{0i} + \pi_{1i} T_{ti} + \pi_{2i} X_{ti} + e_{ti} \quad (\text{A.1})$$

où Y_{ti} est la variable dépendante mesurée à l'occasion t pour l'individu i ; T_{ti} est la variable temporelle qui indique l'occasion de mesure (le nombre d'années après diplôme, dans notre cas) ; X_{ti} est une variable explicative dont la valeur peut fluctuer dans le temps ; et e_{ti} est le terme d'erreur résiduelle usuel (de variance σ_e^2).

Au niveau le plus haut, le modèle multi-niveaux s'écrit :

$$\pi_{0i} = \beta_{00} + \beta_{01} Z_i + u_{0i} \quad (\text{A.2})$$

$$\pi_{1i} = \beta_{10} + \beta_{11} Z_i + u_{1i} \quad (\text{A.3})$$

$$\pi_{2i} = \beta_{20} + \beta_{21} Z_i + u_{2i} \quad (\text{A.4})$$

où les termes β_{00} , β_{10} et β_{20} sont des coefficients *fixes* (i.e. qui ne varient pas selon l'individu) ; Z_i est une variable explicative invariable dans le temps (e.g. le genre d'un individu) ; les termes β_{01} , β_{11} et β_{21} sont les coefficients de régression fixes pour la variable explicative Z_i ; et u_{0i} , u_{1i} et u_{2i} sont des termes d'erreur résiduelle (aléatoires) au niveau des individus. Ces termes d'erreur sont supposés avoir une moyenne égale à 0, et être indépendants des termes e_{ti} au niveau le plus bas. Les variances des termes u_{0i} , u_{1i} et u_{2i} sont respectivement notées $\sigma_{u_0}^2$, $\sigma_{u_1}^2$ et $\sigma_{u_2}^2$; les covariances entre ces termes sont notées $\sigma_{u_{01}}$, $\sigma_{u_{02}}$ et $\sigma_{u_{12}}$, et sont supposées être différentes de 0.

La modélisation multi-niveaux consiste à estimer les différents coefficients de régression fixes (i.e. β_{00} , β_{10} , β_{20} , β_{01} , β_{11} et β_{21}) et les variances/covariances des termes d'erreur résiduelle (i.e. σ_e^2 , $\sigma_{u_0}^2$, $\sigma_{u_1}^2$, $\sigma_{u_2}^2$, $\sigma_{u_{01}}$, $\sigma_{u_{02}}$ et $\sigma_{u_{12}}$).

²L'erreur-type représente une mesure d'incertitude sur un coefficient de régression.

Annexe B

Méthodes d'analyse

La méthode d'analyse diffère selon le type d'indicateur (nominal, Cf. Section B.1, ou métrique, Cf. Section B.2).

B.1 Analyse des indicateurs nominaux

L'analyse des données nominales est descriptive. Elle concerne les diplômé-e-s qui ont participé aux 3 enquêtes (d'où l'absence de valeurs manquantes dans les séquences ou distributions d'états, voir ci-après).

Les *séquences d'états*¹ sont étudiées. En particulier, à l'aide du logiciel R et du package `TraMineR` [3], on représente graphiquement les séquences d'états sous forme de barres horizontales, d'autant plus épaisses que leur fréquence est élevée. Via les séquences d'états, on visualise les possibles transitions d'un état à un autre d'un groupe d'individus qui partagent la même séquence. Il est possible de grouper les séquences d'états selon l'année de promotion et le genre.

Les *distributions d'états*² sont aussi examinées. À cet effet, le package `TraMineR` permet de représenter graphiquement la distribution d'états pour chaque position temporelle. Les distributions d'états rendent compte des proportions d'individus qui se trouvent dans un état donné, à une position temporelle donnée ; contrairement aux séquences d'états, on ne peut visualiser les transitions d'un état à un autre. Il est aussi possible de grouper les distributions d'états en fonction de l'année de promotion et du genre.

B.2 Analyse des indicateurs métriques

L'analyse des données métriques est à la fois descriptive et inférentielle. Elle concerne les diplômé-e-s qui ont participé au moins à une enquête (d'où la présence possible de valeurs manquantes).

Tout d'abord, l'évolution temporelle des données individuelles est représentée graphiquement à l'aide du logiciel R et du package `ggplot2` (aspect descriptif). Les données individuelles peuvent être regroupées par promotion et par genre.

¹Dans le cas présent, la *séquence d'états* d'un individu est composée de trois états, un premier état pour l'année de la première enquête (i.e. 2013, 2014 ou 2015), un second pour l'année de la seconde enquête (i.e. 2016), et un troisième pour l'année de la troisième enquête (i.e. 2017).

²La *distribution d'états* correspond à une distribution *transversale* des états, à chaque position temporelle.

Ensuite, les données individuelles sont modélisées (aspect inférentiel) dans le but – notamment – de déterminer si i) l'indicateur évolue avec le temps, et ii) certaines variables de groupage des individus (année de promotion, genre) influencent cette évolution temporelle.

Dans la pratique, une modélisation multi-niveaux (Cf. Section A.2) est réalisée à l'aide du logiciel R et du package `brms`. Le modèle *linéaire*³, à 2 niveaux, incorpore au niveau le plus haut les 2 variables (nominales) invariables dans le temps que sont le genre et la promotion, ainsi que leur interaction. Au niveau le plus bas, il s'écrit :

$$Y_{ti} = \pi_{0i} + \pi_{1i} T_{ti} + e_{ti} \quad (\text{B.1})$$

où Y_{ti} est la rémunération brute annuelle (avec primes) mesurée t années après diplôme ($t = 1, 2, 3, 4$ ou 5) pour l'individu i ; T est le nombre d'années après diplôme ; et e_{ti} est le terme d'erreur résiduelle.

Au niveau le plus haut, le modèle s'écrit :

$$\pi_{0i} = \beta_{00} + \beta_{01} G_i + \beta_{02} P_i + \beta_{03} G_i \times P_i + u_{0i} \quad (\text{B.2})$$

$$\pi_{1i} = \beta_{10} + \beta_{11} G_i + \beta_{12} P_i + \beta_{13} G_i \times P_i + u_{1i} \quad (\text{B.3})$$

où les termes β_{00} et β_{10} sont les coefficients de régression fixes, respectivement, l'ordonnée à l'origine et la pente d'évolution en fonction du nombre d'années après diplôme ; G_i est le genre de l'individu i (femme ou homme) ; les termes β_{01} et β_{11} sont les coefficients de régression fixes pour la variable explicative G_i ; P_i est l'année de promotion de l'individu i (2012, 2013 ou 2014) ; les termes β_{02} et β_{12} sont les coefficients de régression fixes pour la variable explicative P_i ; $G_i \times P_i$ représente le terme d'interaction entre l'année de promotion et le genre de l'individu i ; les termes β_{03} et β_{13} sont les coefficients de régression fixes pour l'interaction $G_i \times P_i$; et u_{0i} et u_{1i} sont des termes d'erreur résiduelle (aléatoires) au niveau des individus.

³Le modèle testé est linéaire, i.e. on cherche à déterminer une équation qui relie linéairement la rémunération brute annuelle au nombre d'années après diplôme. Notons que l'insertion d'un terme quadratique n'est pas encore possible, puisque nous disposons cette année seulement de 3 mesures répétées dans le temps ; dans le cas présent, l'insertion d'un terme quadratique dans l'équation de régression résulterait en une modélisation parfaite.

Bibliographie

- [1] P.A. McManus (2011). *Introduction to Regression Models for Panel Data Analysis*. Indiana University. URL : https://ssrc.indiana.edu/doc/wimdocs/2011-10-07_mcmanus_panel_slides.pdf. 23
- [2] J. J. Hox (2010). *Multilevel Analysis. Techniques and Applications. Second Edition*. New York and Hove: Routledge, 382 pages. 23, 24
- [3] A. Gabadinho, G. Ritschard, N.S. Muller and M. Studer (2011). “Analyzing and Visualizing State Sequences in R with TraMineR.” *Journal of Statistical Software*, 40(4), 1–37. URL : <http://www.jstatsoft.org/v40/i04>. 25