

**TRAJECTOIRES
PROFESSIONNELLES :
SUIVI DES PROMOTIONS
2012 À 2016
(ENQUÊTES 2013-2018)**

RAPPORT D'ANALYSE – DÉCEMBRE 2018

DIRECTION DES RELATIONS ENTREPRISES – ARNAUD TROLLÉ

Table des matières

Table des figures	IV
Liste des tableaux	IV
Synthèse	1
1 Contexte de l'étude	3
2 Méthodologie	4
2.1 Critères de choix	4
2.2 Indicateurs pertinents	4
3 Résultats	5
3.1 Analyse des indicateurs nominaux	5
3.1.1 La situation professionnelle	5
3.1.2 Le type de contrat	7
3.1.3 Le type d'entreprise	9
3.1.4 Le lieu de travail	9
3.1.5 L'adéquation de l'emploi avec le secteur disciplinaire de la formation	14
3.2 Analyse des indicateurs métriques	16
Annexes	18
A Notions pré-requises	18
A.1 Données de panel / Données longitudinales	18
A.2 Modèles de régression multi-niveaux	18
B Méthodes d'analyse	19
B.1 Analyse des indicateurs nominaux	19
B.2 Analyse des indicateurs métriques	20
Bibliographie	22

Table des figures

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%).	6
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%).	8
3	Type d'entreprise - Séquences d'états par promotion. En ordonnée : Fréquence cumulée en %.	10
4	Type d'entreprise - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).	11
5	Lieu de travail - Séquences d'états par promotion. En ordonnée : Fréquence cumulée en %.	12
6	Lieu de travail - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).	13
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
8	Rémunération brute annuelle (avec primes) - Trajectoires moyennes estimées pour les femmes et les hommes des promotions 2012, 2013, 2014 et 2015 (avec leur intervalle de confiance à 95%).	16

Liste des tableaux

1	Taux de réponse, pour les promotions 2012 à 2016.	3
2	Nombres de diplômés qui ont répondu au moins à une enquête et qui ont répondu à l'ensemble des enquêtes les concernant, pour les promotions 2012 à 2016.	3

Synthèse

Ce document présente les résultats de l'analyse des données de panels de diplômés 2012, 2013, 2014, 2015 et 2016 de l'INSA Lyon. Les données étudiées concernent 6 indicateurs de leur trajectoire professionnelle, issus de l'enquête CGE. Parmi eux, figurent 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, les tendances diffèrent quelque peu d'une promotion à l'autre. Tout d'abord, un an après leur sortie de l'École, la proportion de femmes en activité professionnelle est, en ce qui concerne la promotion la plus ancienne (i.e. 2012), inférieure à la proportion d'hommes dans la même situation ; quant aux promotions suivantes (i.e. 2013, 2014, 2015 et 2016), ces proportions sont relativement similaires. Ensuite, plusieurs années après la sortie de l'École, on observe des évolutions divergentes : soit les proportions de femmes et d'hommes en activité professionnelle restent ou deviennent proches (promotions 2012, 2013 et 2015), soit la proportion de femmes en activité professionnelle devient inférieure à la proportion d'hommes dans la même situation (promotion 2014).
- Concernant le type de contrat, alors que, quelle que soit la promotion, un an après la sortie de l'École, la proportion de femmes en CDI est inférieure à celle des hommes en CDI, les femmes rattrapent leur retard sur les hommes après plusieurs années (promotions 2012, 2013 et 2014), ou alors, les dépassent légèrement même, et ce au bout d'un ou deux ans seulement (promotions 2015 et 2016).
- Concernant le type d'entreprise, quelle que soit la promotion, on observe, en proportion non négligeable (entre 6%, pour la promotion 2016, et 17%, pour la promotion 2013), des diplômés 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.
- Concernant le lieu de travail, quelle que soit la promotion, on observe, en proportion non négligeable (jusqu'à 20% environ pour les promotions 2013 et 2014), des transitions d'un lieu de travail à un autre, ce qui dénote une certaine mobilité géographique pour une frange des diplômés. En outre, quelle que soit la promotion, la proportion de diplômés qui travaillent à l'étranger tend à augmenter avec le nombre d'années après diplomation (par exemple, pour les promotions 2012 et 2013, cette proportion double en 5 ans, pour dépasser les 20%), ou, à défaut, à rester stable.
- Concernant l'adéquation de l'emploi avec le secteur disciplinaire de la formation, on observe qu'au fil du temps, les diplômés de la promotion 2012 ont tendance à s'écarter du secteur disciplinaire de leur formation. Pour les promotions 2013, 2014 et 2015, aucune tendance nette ne se dégage.
- 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 des promotions 2014 et 2015 convergent, i.e. seules les femmes des promotions 2014 et 2015 semblent être en mesure de rattraper leur retard sur les hommes diplômés la même année.

Ces tendances sont encore à confirmer. À cette fin, les données complémentaires collectées via la prochaine enquête CGE (i.e. l'enquête 2019) en ce qui concerne la promotion 2014 (la dernière qui concernera cette promotion), via les 2 prochaines enquêtes CGE (i.e. les enquêtes 2019 et 2020) quant à la promotion 2015 (les deux dernières qui concerneront cette promotion), etc., seront grandement utiles. Notons qu'il n'est plus possible de collecter de données complémentaires pour la promotion 2013 ; l'enquête 2018 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.

1 Contexte de l'étude

En 2018, pour la troisième fois, l'enquête d'insertion professionnelle a concerné 5 promotions de diplômés de l'INSA Lyon : 2013, 2014, 2015, 2016 et 2017. La promotion 2012, non concernée par l'enquête réalisée cette année, avait été interrogée auparavant trois fois. Les promotions 2013 et 2014 ont, quant à elles, été interrogées quatre fois ; la promotion 2015 a été interrogée trois fois ; enfin, la promotion 2016 a été interrogée deux fois.

Le Tableau 1 présente, pour les promotions 2012 à 2016, les taux de réponse aux enquêtes les concernant.

		Enquete n°1		Enquete n°2		Enquete n°3		Enquete n°4	
		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%	2018	39,8%
	2014	2015	85,2%	2016	64,6%	2017	51,1%	2018	46,9%
	2015	2016	81,0%	2017	52,9%	2018	41,5%		
	2016	2017	77,7%	2018	49,3%				

Tableau 1: Taux de réponse, pour les promotions 2012 à 2016.

Le Tableau 2 présente, pour les promotions 2012 à 2016, les nombres de diplômés qui ont répondu au moins à une enquête et qui ont répondu à l'ensemble des enquêtes les concernant.

		Diplômés qui ont répondu au moins à une enquête		Diplômés qui ont répondu à l'ensemble des enquêtes les concernant	
		Nombre	% de l'effectif	Nombre	% de l'effectif
Promotion	2012 (enquêtée 3 fois)	860	89,8%	251	26,2%
	2013 (enquêtée 4 fois)	877	91,8%	152	15,9%
	2014 (enquêtée 4 fois)	925	92,3%	254	25,3%
	2015 (enquêtée 3 fois)	824	85,9%	260	27,1%
	2016 (enquêtée 2 fois)	839	80,7%	462	44,4%

Tableau 2: Nombres de diplômés qui ont répondu au moins à une enquête et qui ont répondu à l'ensemble des enquêtes les concernant, pour les promotions 2012 à 2016.

Dans le cas le plus défavorable, on compte 251, 152, 254, 260 et 462 diplômés 2012, 2013, 2014, 2015 et 2016, respectivement, qui ont répondu à l'ensemble des enquêtes les concernant. Ces tailles d'échantillon, qui restent substantielles, garantissent *a priori* des résultats statistiques robustes.

L'étude vise à analyser les données de panels de diplômés 2012, 2013, 2014, 2015 et 2016 dans l'objectif de dresser et comparer leurs trajectoires professionnelles, mesurées selon différents indicateurs (situation professionnelle, type de contrat, type d'entreprise, localisation de l'emploi, etc.).

2 Méthodologie

Cette section présente la méthodologie adoptée pour analyser les données de panels de diplômés 2012, 2013, 2014, 2015 et 2016.

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 diplômés, 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, 152, 254, 260 et 462 pour les promotions 2012, 2013, 2014, 2015 et 2016, respectivement, si l'on considère les diplômés qui ont répondu à l'ensemble des enquêtes les concernant, 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 diplômés 2012 à 2016 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 diplômés.

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, par exemple, la précision d'un montant.

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 (VIA, VIE, ...)”, “En recherche d’emploi”, “En poursuite d’études (hors thèse)”, “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).

3 Résultats

Cette section présente les résultats de l’analyse des données de panels de diplômés 2012, 2013, 2014, 2015 et 2016, opérée suivant la méthodologie abordée dans 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.

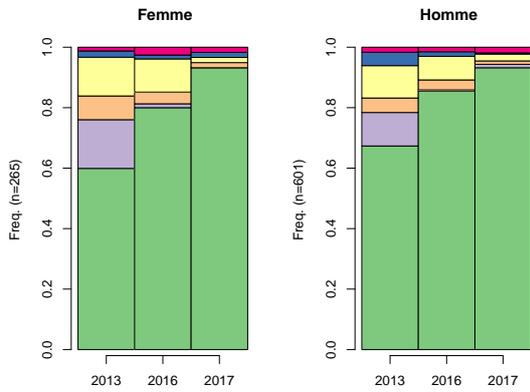
Ensuite, les résultats de l’analyse des indicateurs métriques sont relatés (Cf. Section 3.2). Pour rappel, un seul indicateur métrique a été étudié : la rémunération brute annuelle (avec primes).

3.1 Analyse des indicateurs nominaux

3.1.1 La situation professionnelle

La Figure 1 présente les distributions d’états, par promotion et par genre.

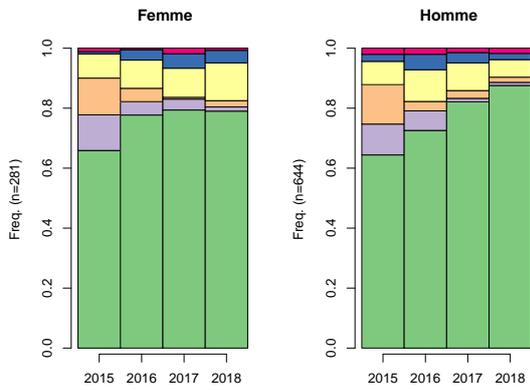
Tout d’abord, observons les distributions d’états promotion par promotion. Pour la promotion 2012, la proportion de femmes en activité professionnelle un an après la sortie de l’École (i.e. en 2013) est inférieure à la proportion d’hommes dans la même situation. Cinq ans après la sortie de l’École (i.e. en 2017), les proportions de femmes et d’hommes en activité professionnelle sont équivalentes. Pour les promotions 2013, 2014, 2015 et 2016, les proportions de femmes et d’hommes en activité professionnelle un an après la sortie de l’École (i.e. en 2014, 2015, 2016 et 2017, respectivement) sont relativement similaires. Alors que, pour la promotion 2013, les proportions de femmes et d’hommes en activité professionnelle 4 ans (et aussi 5 ans) après la sortie de l’École restent similaires, pour la promotion 2014, la proportion de femmes en activité professionnelle 4 ans après la sortie de l’École devient inférieure à la proportion d’hommes dans la même situation. Pour la promotion 2015, les proportions de femmes et d’hommes en activité professionnelle 3 ans après la sortie de l’École restent proches. Pour la



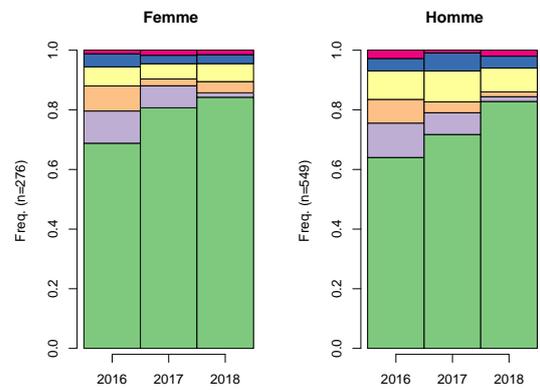
(a) Promotion 2012.



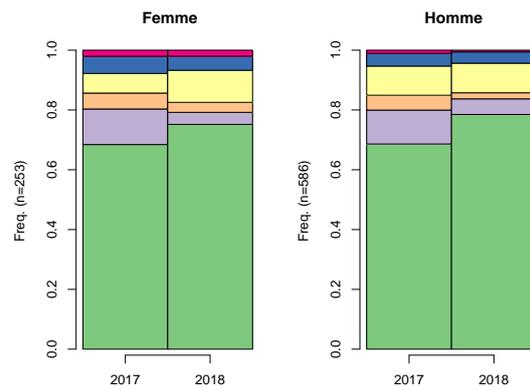
(b) Promotion 2013.



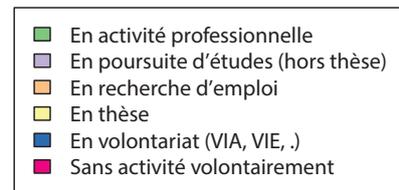
(c) Promotion 2014.



(d) Promotion 2015.



(e) Promotion 2016.



(f) Légende des états.

Figure 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%).

promotion 2016, la proportion de diplômés en activité professionnelle 2 ans après la sortie de l'École est déjà plus élevée chez les hommes que chez les femmes.

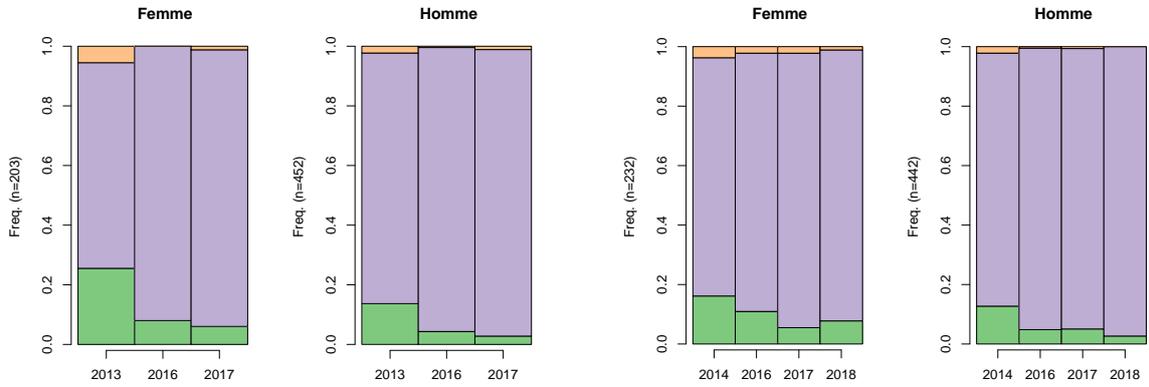
Au travers des 5 promotions, on retrouve certaines tendances logiques : les proportions de femmes ou d'hommes en recherche d'emploi, ainsi que les proportions de femmes ou d'hommes en poursuite d'études (hors thèse), diminuent avec le nombre d'années après diplomation.

3.1.2 Le type de contrat

La Figure 2 présente les distributions d'états, par promotion et par genre.

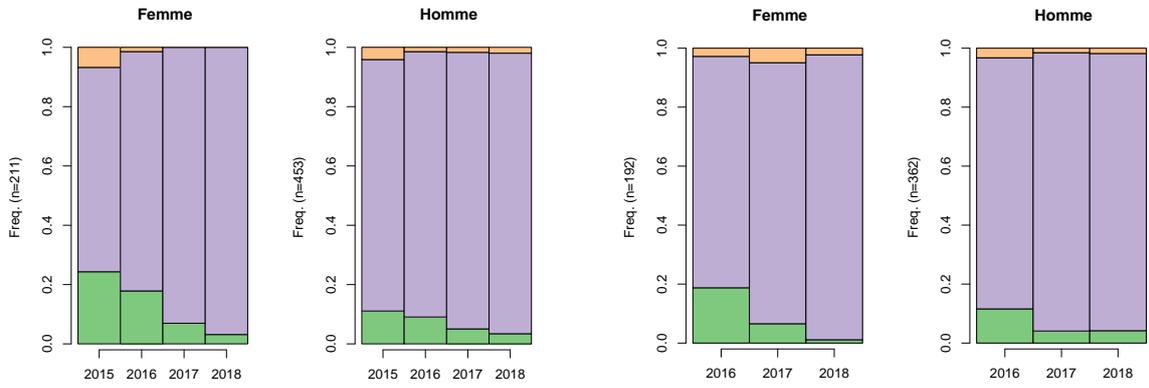
Tout d'abord, examinons encore les distributions d'états promotion par promotion. Dans l'ensemble, pour les promotions 2012, 2013, 2014, 2015 et 2016, la proportion de femmes en CDI un an après la sortie de l'École (i.e. en 2013, 2014, 2015, 2016 et 2017, respectivement) est inférieure, de façon plus ou moins importante, à la proportion d'hommes en CDI. Pour les promotions 2012, 2013 et 2014, 4 ou 5 ans après la sortie de l'École, le retard est quasiment rattrapé : les proportions de femmes et d'hommes en CDI sont relativement proches. Mieux, pour les promotions 2015 et 2016, la proportion de femmes en CDI devient légèrement supérieure à la proportion d'hommes en CDI, 3 et 2 ans, respectivement, après la sortie de l'École.

Globalement, les proportions de femmes et d'hommes en CDI augmentent, sans surprise, avec le nombre d'années après diplomation, et ce quelle que soit l'année de promotion.



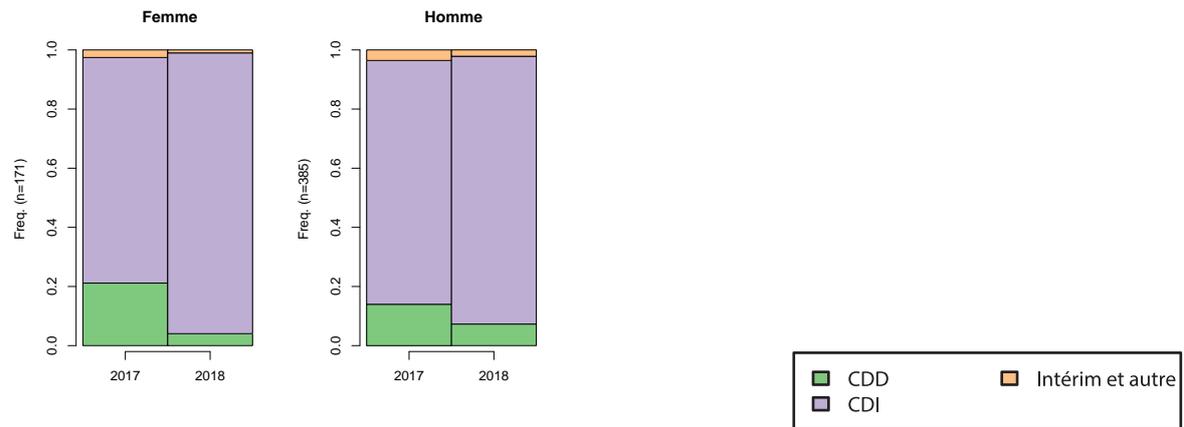
(a) Promotion 2012.

(b) Promotion 2013.



(c) Promotion 2014.

(d) Promotion 2015.



(e) Promotion 2016.

(f) Légende des états.

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

Quelle que soit la promotion, la première séquence la plus fréquente correspond aux diplômés qui travaillent dans une ETI/Grande entreprise un an après la sortie de l'École et qui restent dans une telle structure ; la seconde séquence la plus fréquente correspond aux diplômés qui travaillent dans une TPE/PME un an après la sortie de l'École et qui restent dans une telle structure, hormis pour la promotion 2013. Pour cette promotion, la seconde séquence la plus fréquente correspond à des diplômés qui travaillaient dans une ETI/ Grande entreprise un an après la sortie de l'École mais qui, plusieurs années plus tard, sont passés dans une TPE/PME. Cette observation est à prendre avec précaution, car la taille de l'échantillon de diplômés 2013 est assez faible, comparativement aux autres tailles d'échantillons. En outre, relevons que, quelle que soit la promotion, la 3^{ème} séquence la plus fréquente correspond à des diplômés qui travaillaient dans une ETI/ Grande entreprise un an après la sortie de l'École mais qui, seulement un an plus tard, sont passés dans une TPE/PME (et y sont restés). Le chemin inverse (i.e. le passage d'une TPE/PME à une ETI/Grande entreprise) existe, mais dans une moindre mesure.

La Figure 4 présente les distributions d'états, par promotion.

Concernant les promotions 2012 et 2013, avec le nombre d'années après diplomation, on observe une augmentation de la proportion de diplômés qui travaillent dans une TPE-PME ; pour les promotions 2014, 2015 et 2016, cette proportion reste assez stable.

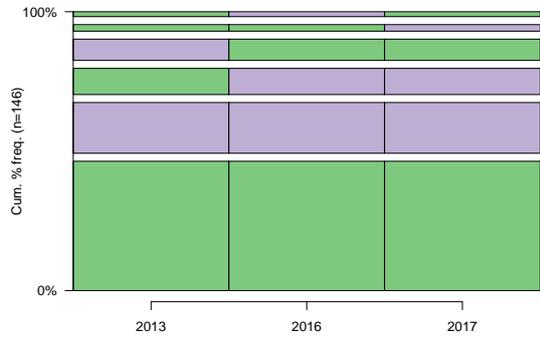
3.1.4 Le lieu de travail

La Figure 5 présente les séquences d'états les plus fréquentes, par promotion.

Quelle que soit la promotion, les 3 séquences les plus fréquentes correspondent aux diplômés 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 diplômés.

La Figure 6 présente les distributions d'états, par promotion.

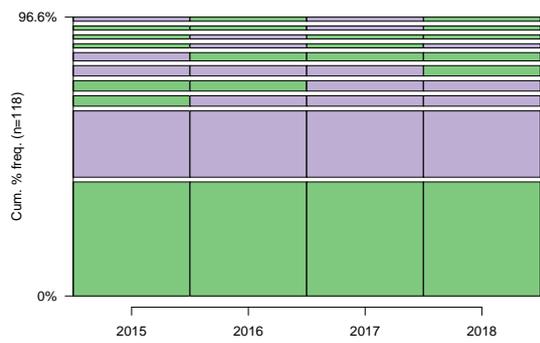
Pour la promotion 2012, la proportion de diplômés qui travaillent à l'étranger augmente avec le nombre d'années après diplomation, tandis que la proportion de diplômés qui travaillent en Province diminue ; la proportion de diplômés qui travaillent en Île-de-France augmente entre 2013 et 2016, puis diminue l'année d'après (i.e. en 2017). Pour la promotion 2013, la proportion de diplômés qui travaillent à l'étranger augmente entre 2014 et 2017, puis reste stable l'année d'après (i.e. en 2018) ; la proportion de diplômés qui travaillent en Province diminue entre 2014 et 2016, puis se stabilise les deux années suivantes (i.e. en 2017 et 2018) ; la proportion de diplômés qui travaillent en Île-de-France reste stable entre 2014 et 2016, puis diminue l'année d'après (i.e. en 2017), pour se stabiliser ensuite (i.e. en 2018). Concernant la promotion 2014, les proportions de diplômés qui travaillent à l'étranger, en Province et en Île-de-France restent relativement stables. Pour les promotions 2015 et 2016, avec le nombre d'années après diplomation, la proportion de diplômés qui travaillent à l'étranger tend à augmenter, la proportion de diplômés qui travaillent en Province à diminuer et la proportion de diplômés qui travaillent en Île-de-France à rester stable.



(a) Promotion 2012.



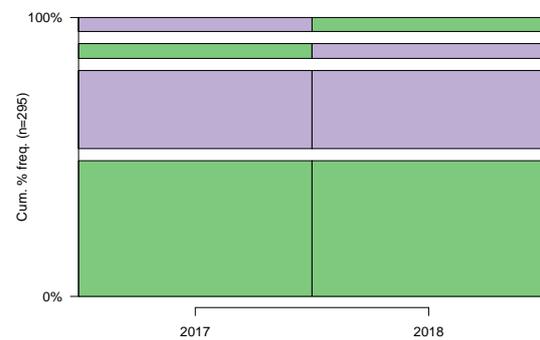
(b) Promotion 2013.



(c) Promotion 2014.



(d) Promotion 2015.

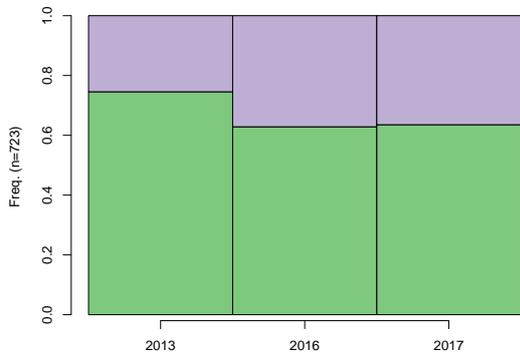


(e) Promotion 2016.

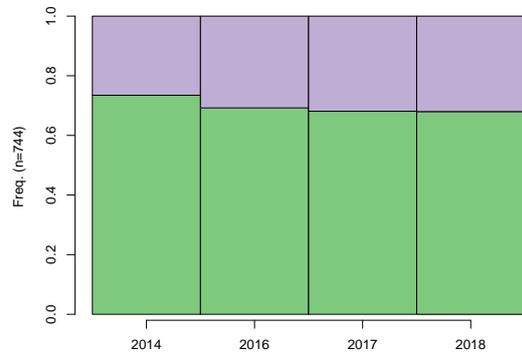


(f) Légende des états.

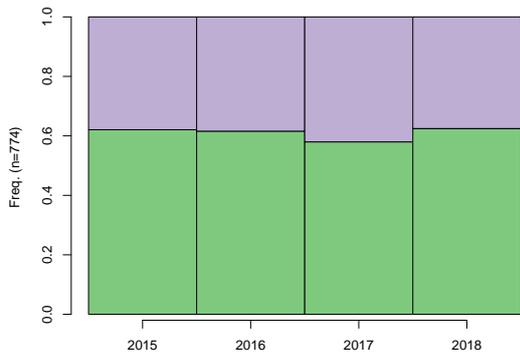
Figure 3: Type d'entreprise - Séquences d'états par promotion. En ordonnée : Fréquence cumulée en %.



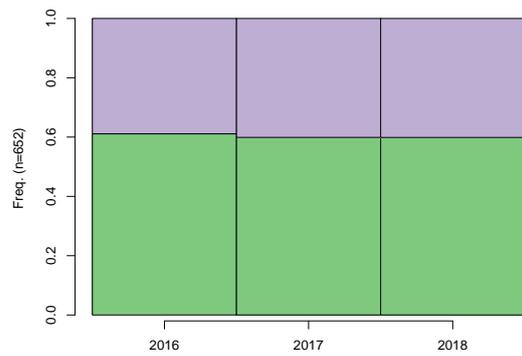
(a) Promotion 2012.



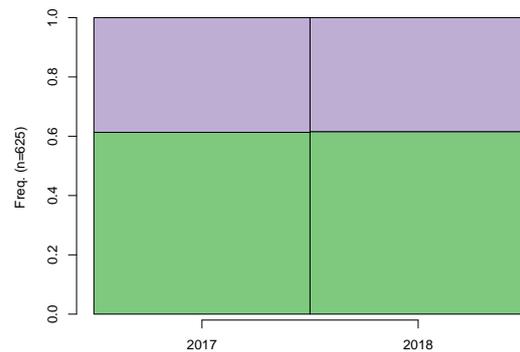
(b) Promotion 2013.



(c) Promotion 2014.



(d) Promotion 2015.

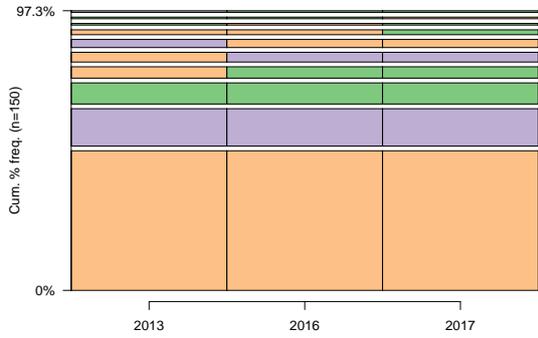


(e) Promotion 2016.

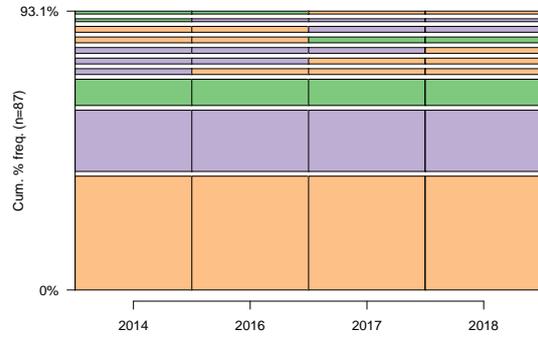


(f) Légende des états.

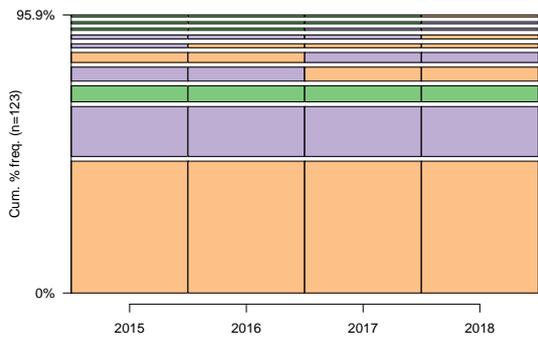
Figure 4: Type d'entreprise - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).



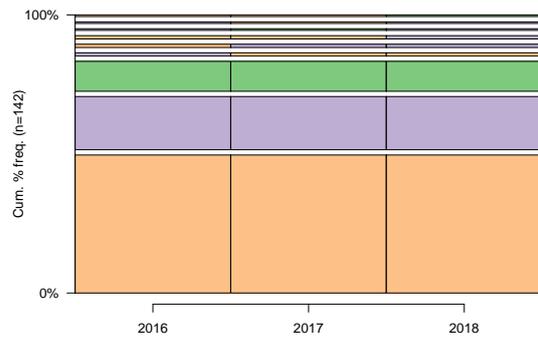
(a) Promotion 2012.



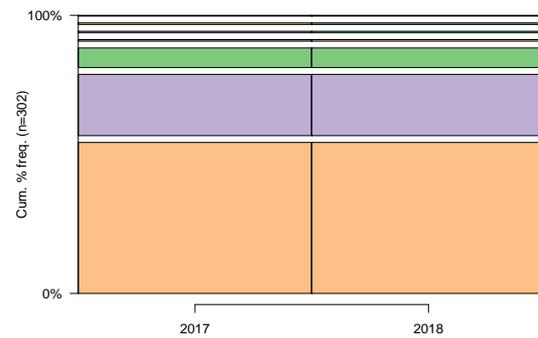
(b) Promotion 2013.



(c) Promotion 2014.



(d) Promotion 2015.

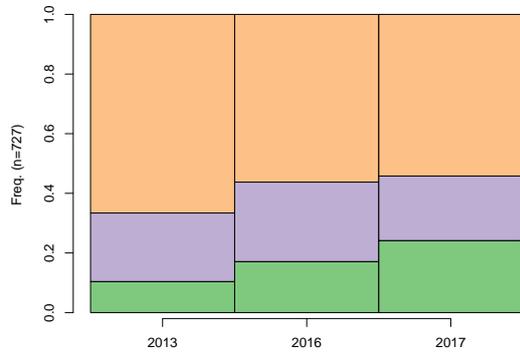


(e) Promotion 2016.

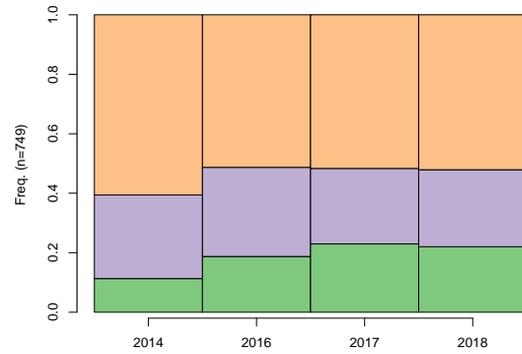


(f) Légende des états.

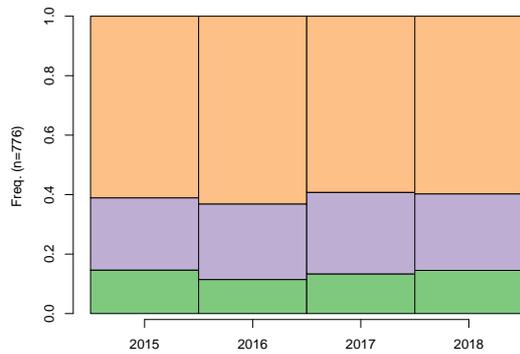
Figure 5: Lieu de travail - Séquences d'états par promotion. En ordonnée : Fréquence cumulée en %.



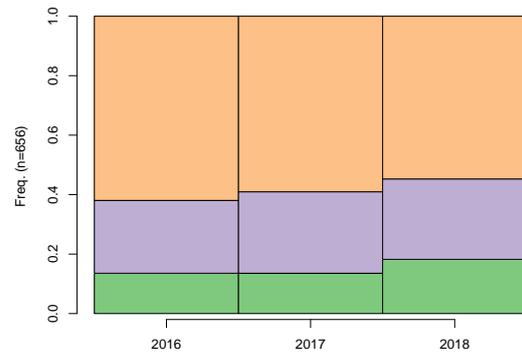
(a) Promotion 2012.



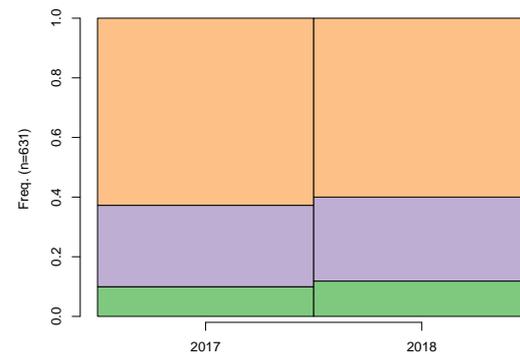
(b) Promotion 2013.



(c) Promotion 2014.



(d) Promotion 2015.



(e) Promotion 2016.



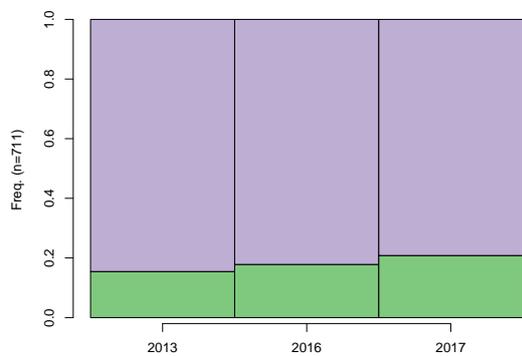
(f) Légende des états.

Figure 6: Lieu de travail - Distributions d'états par promotion. En ordonnée : Fréquence cumulée (entre 0, i.e. 0%, et 1, i.e. 100%).

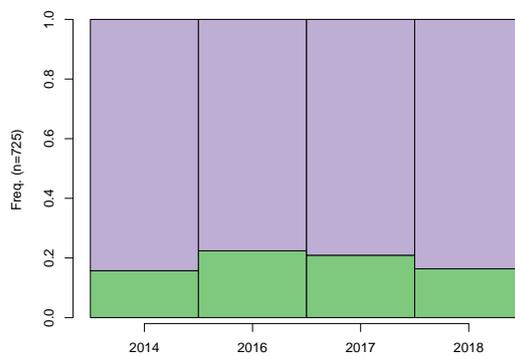
3.1.5 L'adéquation de l'emploi avec le secteur disciplinaire de la formation

La Figure 7 présente les distributions d'états, par promotion.

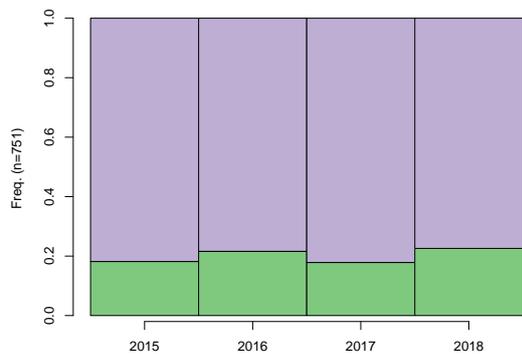
Concernant la promotion 2012, avec le nombre d'années après diplomation, on note une légère augmentation de la proportion de diplômés qui ne jugent plus leur emploi en adéquation avec le secteur disciplinaire de leur formation, i.e. au fil du temps, ces diplômés ont tendance à s'écarter du secteur disciplinaire de leur formation. Pour les promotions 2013, 2014 et 2015, interrogées ces 3 dernières années, cette proportion est parfois en hausse, parfois en baisse : aucune tendance nette ne se dégage.



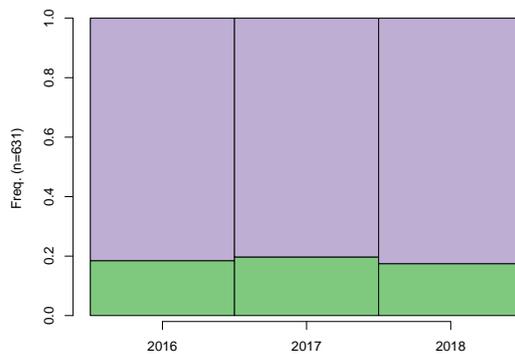
(a) Promotion 2012.



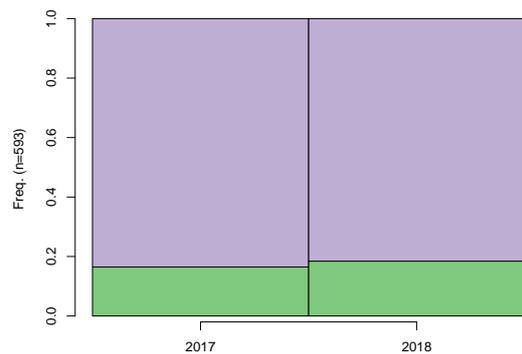
(b) Promotion 2013.



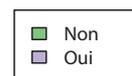
(c) Promotion 2014.



(d) Promotion 2015.



(e) Promotion 2016.



(f) Légende des états.

Figure 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%).

3.2 Analyse des indicateurs métriques

Pour la rémunération brute annuelle (exprimée en kEuros constants), la Figure 8 montre les trajectoires moyennes estimées, pour les femmes et les hommes des promotions 2012, 2013, 2014 et 2015, par le modèle de régression multi-niveaux, avec leur intervalle de confiance à 95%.

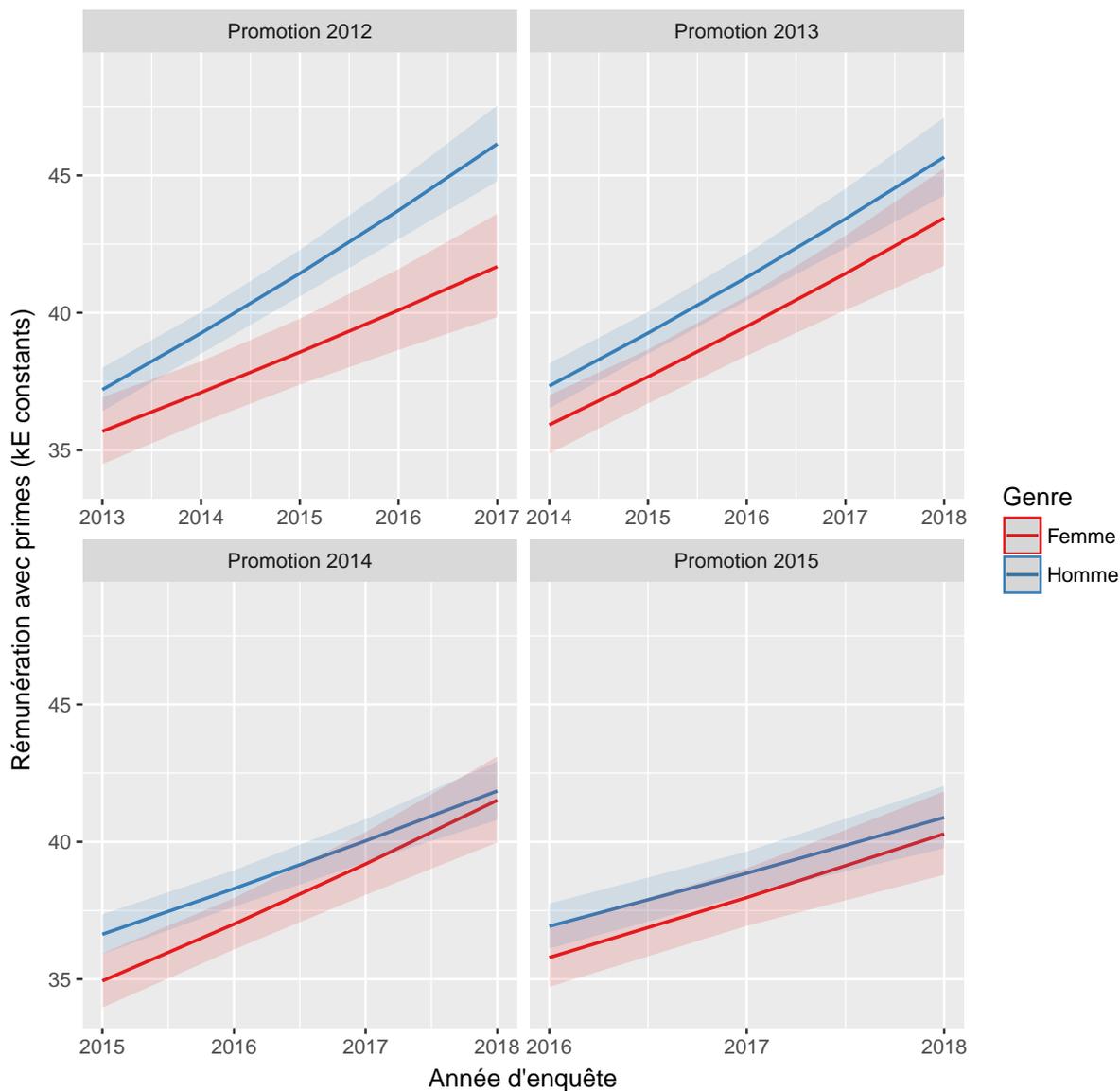


Figure 8: Rémunération brute annuelle (avec primes) - Trajectoires moyennes estimées pour les femmes et les hommes des promotions 2012, 2013, 2014 et 2015 (avec leur intervalle de confiance à 95%).

On note que :

- i) Sans surprise, la rémunération brute annuelle augmente avec le nombre d'années après diplomation, 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) L'augmentation de la rémunération des femmes de la promotion 2014 en fonction du nombre d'années après diplomation est plus importante, relativement aux femmes des autres promotions. 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).
- iv) *A contrario*, l'augmentation de la rémunération des hommes des promotions 2014 et 2015 en fonction du nombre d'années après diplomation 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 iii) et iv), 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 des promotions 2014 et 2015 convergent, i.e. seules les femmes des promotions 2014 et 2015 semblent être en mesure de rattraper leur retard sur les hommes diplômés la même année. Concernant la promotion 2014, cette tendance était déjà observée l'année dernière ; elle est donc confirmée.

Annexes

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 diplômés des promotions 2012 à 2016 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 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] :

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

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

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

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 diplomation, 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 :

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

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

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

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 $\sigma_{u_{12}}$).

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.

Les *séquences d’états*³ sont étudiées. Seules les réponses des diplômés qui ont participé à l’ensemble des enquêtes les concernant (et qui ont répondu, à chaque fois, à la question donnée) sont prises en compte, d’où l’absence de valeurs manquantes. 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

³Dans le cas présent, la *séquence d’états* d’un individu est composée de plusieurs états, dont le nombre dépend de sa promotion :

- Promotions 2012 et 2015 : 3 états, i.e. un premier état pour l’année de la première enquête (2013 et 2016, respectivement), un second pour l’année de la seconde enquête (2016 et 2017, respectivement), et un troisième pour l’année de la troisième enquête (2017 et 2018, respectivement).
- Promotions 2013 et 2014 : 4 états, i.e. un premier état pour l’année de la première enquête (2014 et 2015, respectivement), un second pour l’année de la seconde enquête (2016), un troisième pour l’année de la troisième enquête (2017), et un quatrième pour l’année de la quatrième enquête (2018).
- Promotion 2016 : 2 états, i.e. un premier état pour l’année de la première enquête (2017), et un second pour l’année de la seconde enquête (2018).

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*⁴ sont aussi examinées. Ici, les réponses des diplômés qui ont répondu au moins à une enquête (et à la question donnée) sont prises en compte, d'où la présence possible de valeurs manquantes. 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 (le nombre de valeurs manquantes n'est pas pris en compte dans le calcul des proportions). 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 (de rémunération) est inférentielle. Elle concerne les diplômés 2012 à 2015 qui ont participé au moins à une enquête (d'où la présence possible de valeurs manquantes).

Les données individuelles sont modélisées 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 à 2 niveaux (Cf. Section A.2) est réalisée à l'aide du logiciel R et du package `lme4`. Le modèle de régression testé est *linéaire*, i.e. on cherche à déterminer une équation qui relie linéairement la rémunération brute annuelle au nombre d'années après diplomation. Notons que la modélisation ne prend pas en compte les réponses des diplômés 2016, puisque, pour cette promotion, nous disposons cette année seulement de 2 mesures répétées dans le temps (et donc, la régression linéaire serait parfaite). Le modèle 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 :

$$Y_{ti} = \pi_{0i} + \pi_{1i} T_{ti} + e_{ti} \quad (5)$$

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

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

$$\pi_{0i} = \beta_{00} + \beta_{01} G_i + \beta_{02} P_i + \beta_{03} G_i \times P_i + u_{0i} \quad (6)$$

$$\pi_{1i} = \beta_{10} + \beta_{11} G_i + \beta_{12} P_i + \beta_{13} G_i \times P_i + u_{1i} \quad (7)$$

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 diplomation ; 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

⁴La *distribution d'états* correspond à une distribution *transversale* des états, à chaque position temporelle.

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.

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_mcmamus_panel_slides.pdf. 18
- [2] J. J. Hox (2010). *Multilevel Analysis. Techniques and Applications. Second Edition*. New York and Hove: Routledge, 382 pages. 18
- [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>. 19