

Une analyse des trajectoires scolaires au secondaire : Un modèle en chaîne de Markov

Guy Lacroix (Université Laval)

PRÉPARÉ POUR LE MINISTÈRE DE L'ÉDUCATION¹

18 juillet 2025

1. L'auteur remercie Marie-lin Hamel et Éliane Landry-Tremblay pour leurs judicieux conseils et leurs nombreux commentaires.

FAITS SAILLANTS

- L'analyse s'intéresse aux trajectoires scolaires au niveau secondaire de plus de 365 169 élèves entrés en première année du primaire au cours des années scolaires 2003-2004 à 2008-2009 ;
- Six états potentiels sont considérés, dont trois sont transitoires : formation générale des jeunes (FGJ), formation professionnelle (FP), formation générale des adultes (FGA), deux sont diplômants : diplôme d'études secondaires (DES), tout autre diplôme ou qualification (DEP), et un est dit « absorbant », hors étude (HE) ;
- Les trajectoires scolaires sont analysées à l'aide d'outils de séquençage et de modélisation économétrique ;
- L'analyse montre une très forte hétérogénéité dans les parcours scolaires ;
- Cette hétérogénéité est intimement liée aux caractéristiques des élèves et du milieu duquel ils sont issus ;
- La plupart des variables explicatives prises en compte dans l'analyse ont des effets statistiquement très significatifs et de signe attendu ;
- Parce que les trajectoires sont liées aux caractéristiques individuelles, elles sont relativement prévisibles.

Table des matières

1	Introduction	1
2	Données et analyse descriptive	1
2.1	Sélection de l'échantillon	2
2.2	Caractéristiques de l'échantillon	2
2.3	Trajectoires individuelles	5
3	Modèle en chaîne de Markov	10
4	Résultats	14
5	Conclusion	20
A	Paramètres de base (Matrice P)	26
B	Paramètres des variables explicatives (β)	27

Liste des tableaux

1	Transitions brutes des élèves des cohortes 2003-2004 à 2008-2009	2
2	Caractéristiques des élèves	4
3	Transitions des élèves de l'échantillon final	5
4	Caractéristiques des regroupements	10
5	Effets des variables explicatives sur la durée espérée	16
6	Distribution de l'âge à l'entrée dans l'échantillon	17
7	Distribution de l'âge à la sortie de l'échantillon	18
8	Caractéristiques des élèves selon la participation (ou non) à <i>Passe-Partout</i> et <i>Maternelle 4 ans</i>	20
A1	Paramètres de base (Matrice P)	26
B1	Caractéristiques des élèves	27
B2	Programmes	28
B3	Milieux socio-économique	29
B4	Effets de cohorte	30

Table des figures

1	Sélection de l'échantillon	3
2	Trajectoires individuelles des parcours scolaires	7
3	Distribution des types d'élèves selon le CSS	9
4	Paramètres : Caractéristiques des élèves	22
5	Paramètres : Programmes	23
6	Paramètres : Milieu socio-économique	24
7	Paramètres : Cohortes	25

1 Introduction

Ce rapport s'inscrit dans le cadre de l'Entente de services #350046177 entre le ministère de l'Éducation et l'Université Laval. Cette entente prévoit la réalisation d'une étude portant sur « Les coûts réels du décrochage scolaire au Québec : une analyse économétrique du décrochage et du raccrochage scolaires. Analyse de survie et de transition ». ²

2 Données et analyse descriptive

Les données mobilisées aux fins de l'analyse empirique proviennent de la base de données longitudinales du Québec (BDLEQ). L'ensemble des élèves des cohortes ayant débuté leur scolarisation primaire au cours des années 2003-2004 à 2008-2009 est pris en compte. Une analyse fine des données est essentielle à l'élaboration d'une stratégie de modélisation des trajectoires scolaires des élèves.

Les données documentent les parcours scolaires de plus de 447 807 élèves. Ces parcours contiennent plus de 7 121 704 observations (élèves \times années). Ils couvrent six états potentiels soit : la formation générale des jeunes (FGJ), la formation professionnelle (FP), la formation générale des adultes (FGA), l'obtention d'un diplôme d'études secondaires (DES), l'obtention de tout autre diplôme ou qualification (DEP) ainsi que l'état hors étude (HE). Ces parcours scolaires des élèves peuvent être représentés par une *matrice de transitions*, laquelle indique en ordonnées l'état de l'élève à l'année t (origine) et son état à l'année $t + 1$ en abscisse (destination).

Le Tableau 1 répertorie plus de 6 673 897 transitions entre dix états scolaires. ³ Il a été convenu que lorsqu'un élève transite dans plusieurs parcours au cours d'une même année (FGA \rightarrow FP, FGJ \rightarrow FGA, FGJ \rightarrow FGA \rightarrow FP, FGJ \rightarrow FP), seul le dernier état observé allait être retenu, soit FP ou FGA selon le cas. Plusieurs remarques s'imposent à la lecture du tableau. Tout d'abord, certaines transitions sont très peu nombreuses en dépit du très grand nombre d'élèves considérés. Par ailleurs, on n'observe aucune transition HE \rightarrow FGJ, HE \rightarrow FP, ou encore HE \rightarrow FGA. En conséquence, la présente étude ne pourra pas porter sur le « raccrochage » scolaire si tant est que celui-ci soit défini comme une de ces transitions. L'étude portera plutôt sur l'analyse des transitions pour lesquelles suffisamment d'observations peuvent être mobilisées.

2. Le titre du rapport est différent de celui qui était prévu dans l'entente pour deux raisons. Tout d'abord, la base de données ne permet pas d'identifier les situations de décrochage et de raccrochage sans faire des hypothèses fortes sur la situation des élèves entre deux présences parfois éloignées dans le temps. Deuxièmement, les données étant de nature discrète (annuelles), elles ne se prêtent pas bien à une analyse de survie. Il est nettement préférable de les analyser à l'aide de modèles de transitions discrètes, tels les modèles en chaîne de Markov.

3. La dernière année d'observation de chaque élève ne donne pas lieu à une transition. Il est donc normal d'avoir moins de transitions que d'observations.

Tableau 1 – Transitions brutes des élèves des cohortes 2003-2004 à 2008-2009

Orig (t) \diagdown	Dest ($t+1$)	Collège	FGA	FGA ↓ FP	FGJ	FGJ ↓ FGA	FGJ ↓ FGA ↓ FP	FGJ ↓ FP	FP	HE	Univ
		Collège	422 939	0	0	0	0	0	0	0	39 261
FGA	17 352	106 893	9 739	296	248	9	8	16 537	25 080	3 342	
FGA→FP	1496	5039	4399	23	15	6	4	14210	2036	94	
FGJ	154 101	86 779	12 219	4 717 580	11 104	600	4 415	54 094	9 702	70 466	
FGJ→FGA	979	6 265	980	265	121	15	23	15 12	89	212	
FGJ→FGA→FP	26	189	207	7	1	12	10	353	5	0	
FGJ→FP	261	755	633	602	58	258	1 247	1 456	17	33	
FP	26 532	5 059	3127	26	10	10	12	69 646	21 007	1 709	
HE	39 427	0	0	0	0	0	0	0	76 274	13 660	
Univ	195 193	0	0	0	0	0	0	0	13 480	204 968	

2.1 Sélection de l'échantillon

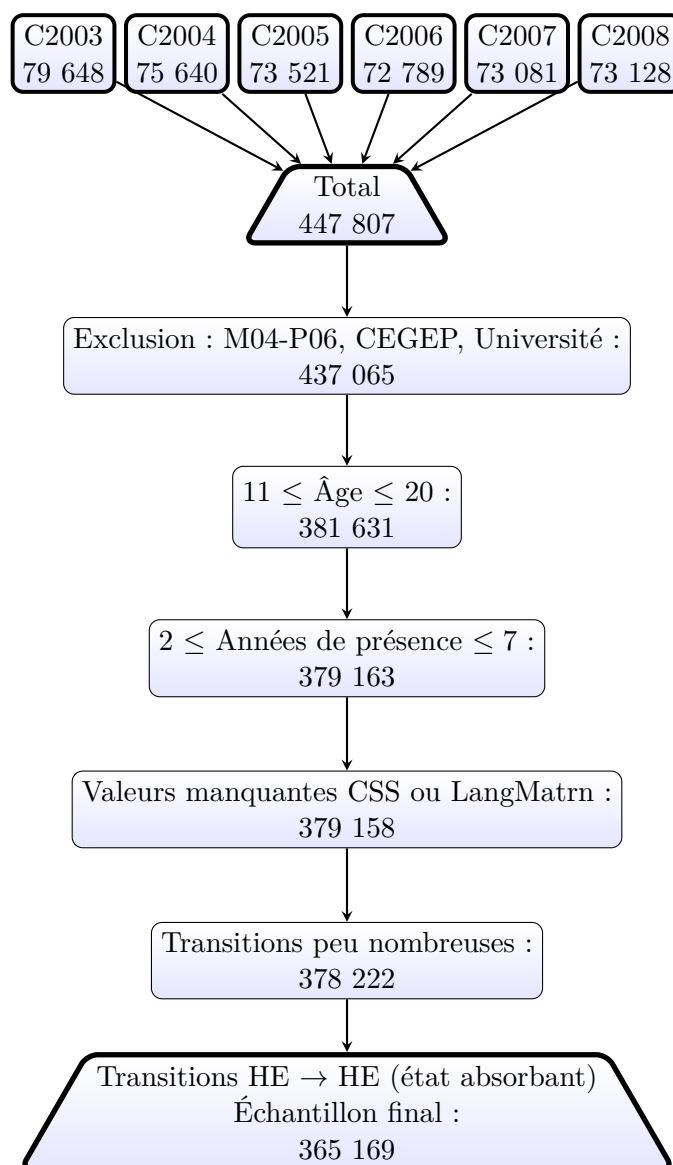
La présente étude porte sur les trajectoires scolaires au niveau secondaire. La population d'intérêt constitue un sous-ensemble des données présentes dans les fichiers BDLEQ. Ainsi, plusieurs filtres doivent être appliqués aux données brutes pour en tirer un échantillon pertinent pour notre objet d'étude. La Figure 1 illustre les différentes étapes de filtrage qui ont mené à l'échantillon final. La ligne du haut indique le nombre d'élèves présents dans chacune des cohortes. Les élèves fréquentant la maternelle, l'école primaire, le CÉGEP ou l'université ont tout d'abord été éliminés des données brutes. Par la suite, seuls les élèves ayant entre 11 et 20 ans ont été considérés.⁴ Un minimum de deux années de présence est nécessaire pour pouvoir calculer des transitions. Le nombre d'années d'observation a été limité à 7, bien que ce critère soit peu contraignant étant donné que la limite d'âge est fixée à 20 ans. Les observations pour lesquelles le code CSS ou le code de langue maternelle étaient manquants ont également été éliminées. Enfin, les transitions trop peu nombreuses pour être traitées à l'aide d'outils économétriques ont été supprimées. Enfin, les transitions HE → HE ont été omises car nous verrons plus loin que cet état est considéré comme « absorbant ». Au final, l'échantillon est constitué de plus de 365 169 élèves fréquentant le secondaire et de 1 928 425 observations, lesquelles ont généré plus de 1 563 256 transitions.

2.2 Caractéristiques de l'échantillon

Le Tableau 2 regroupe les principales caractéristiques des élèves qui sont prises en compte dans la modélisation économétrique des parcours scolaires. Les statistiques sont rapportées distinctement pour les filles et les garçons. La plupart des moyennes sont relativement similaires :

4. Ce choix est dicté par le fait qu'un candidat âgé de plus de 21 ans ayant un diplôme d'études secondaires ou diplôme d'études professionnelles, peut, sous certaines conditions, accéder à des études postsecondaires.

Figure 1 – Sélection de l'échantillon



78% des élèves ont le français comme langue maternelle et 10% d'entre eux habitent l'île de Montréal. Les filles, tout comme les garçons, ont fréquenté environ 1,13 CSS durant leur cheminement scolaire et plus de 26% ont fréquenté un établissement privé durant un minimum d'une année. La principale distinction entre les deux groupes concerne les besoins particuliers : les garçons sont deux fois plus nombreux que les filles à être handicapés ou à présenter des difficultés d'adaptation ou d'apprentissage (EHDAA). Les filles sont également proportionnellement plus nombreuses à participer à un programme agréé⁵, alors que le taux de participation à un programme enrichi, à Passe-Partout ou encore à la maternelle quatre ans est très semblable.

Tableau 2 – Caractéristiques des élèves

	Filles	Garçons
Caractéristiques		
Français	0,788	0,782
Montréal	0,103	0,111
# CSS	1,123	1,146
École privée	0,273	0,256
EHDAA	0,138	0,265
Programmes		
ProgAgréé	0,144	0,091
ProgEnrichi	0,088	0,097
PassePartout	0,117	0,116
Mater4anDT	0,056	0,062
Milieu socio-économique		
Revenu moyen/AD ($\times 100\ 000$)	0,966	0,965
Revenu median/AD ($\times 100\ 000$)	0,867	0,868
IMSE	9,713	10,054
Nombre d'élèves	183 819	181 350

Le Tableau 3 est semblable au Tableau 1 mais porte sur l'échantillon final. L'état « DES » correspond à l'obtention d'un diplôme d'études secondaires alors que l'état « DEP » correspond à l'obtention de tout diplôme autre que DES.⁶ L'état « HE » ne figure pas parmi les origines. En revanche, on observe de nombreuses transitions des autres états vers « HE ». C'est pourquoi cet état est considéré comme un état « absorbant », c'est-à-dire qu'il s'agit d'une destination finale : une fois qu'on y entre, on ne peut en sortir.⁷ Les transitions marquées d'un "X" ont été retirées des

5. Les programmes agréés représentent les programmes reconnus, soit le PEI, Arts-Études et Sports-Études.

6. Ces états n'étaient pas présents dans le Tableau 1.

7. Dans la littérature médicale, la mort est un état absorbant naturel lorsqu'il s'agit d'étudier la progression de différentes maladies. Les modèles en chaîne de Markov ne requièrent pas l'existence d'un état absorbant. Toutefois, la notion de durée attendue dans un état quelconque doit être définie par rapport à un état de destination. Autrement

données, car elles sont trop peu nombreuses pour en permettre un traitement statistique. Enfin, la cellule « DES-DES » ne contient aucune observation car on ne peut évidemment transiter plus d’une fois dans cet état, bien que des transitions « DEP-DEP » soient possibles, quoique peu nombreuses.

Tableau 3 – Transitions des élèves de l’échantillon final

Orig \ Dest	FGJ	FP	FGA	DES	DEP	HE
FGJ	1 064 066	8 906	30 822	281 515	17 181	1 021
FP	X	6 419	1 454	1 961	15 599	3 653
FGA	X	2 853	30 523	3 408	4 882	6 791
DES	X	23 673	13 027	–	6919	19 461
DEP	1 695	4 667	6 545	X	3 588	2 627

2.3 Trajectoires individuelles

Le Tableau 3 présente les transitions issues de l’échantillon sélectionné selon les critères énumérés à la Figure 1. Les états « Collège » et « Université » ont été supprimés de l’échantillon. Nous avons par ailleurs défini les états « DES » (FGJ, FGA) et « DEP » (FP, FGA) pour rendre compte de l’obtention d’un diplôme, ce qui n’est pas représenté dans le Tableau 1.

Le Tableau 3 rapporte l’ensemble des transitions observées dans les données sans prendre en compte leur chronologie. Il peut être utile de considérer les trajectoires individuelles des élèves de façon chronologique afin de rendre compte de l’hétérogénéité des parcours scolaires. Pour ce faire, un tirage aléatoire correspondant à 10% de l’échantillon final (36 603 élèves) a été sélectionné à des fins d’analyse. Les 36 603 trajectoires individuelles sont analysées et une mesure de la distance entre chacune d’elles est ensuite calculée.

Calcul de la distance entre les trajectoires individuelles

La méthode de mesure de dissimilarité la plus couramment utilisée en analyse de séquences, notamment dans les sciences sociales, est l’algorithme dit de *Optimal Matching (OM)*. Cet algorithme est utilisé dans le contexte des parcours professionnels, les biographies familiales, les séquences d’activités, ou encore les parcours scolaires. L’algorithme *OM* calcule le coût minimal pour transformer une séquence en une autre à l’aide de trois types d’opérations à *coût* défini :

dit, on doit calculer la durée espérée d’un épisode avant de faire une transition dans un état précis. Il y a autant de durées espérées qu’il y a d’états vers lesquels on peut transiter. Les calculs sont d’autant plus complexes que les états sont récurrents.

-
1. Substitution (S) : remplacer un état par un autre. Coût = 2 (deux opérations).
 2. Insertion (I) : insérer un état. Coût = 1 (une opération).
 3. Suppression (D) : supprimer un état. Coût = 1 (une opération).

Exemple Soit les deux séquences suivantes :

Séquence 1 : A B C D

Séquence 2 : A C D E

Il existe deux manières possibles de rendre ces séquences identiques :

1. Substituer B en C (2), D en E (2) → Coût = 4
2. Supprimer B (1), insérer E (1) → Coût = 2

Le coût minimal est donc deux, car c'est la transformation la moins chère qui permet d'obtenir deux séquences identiques.

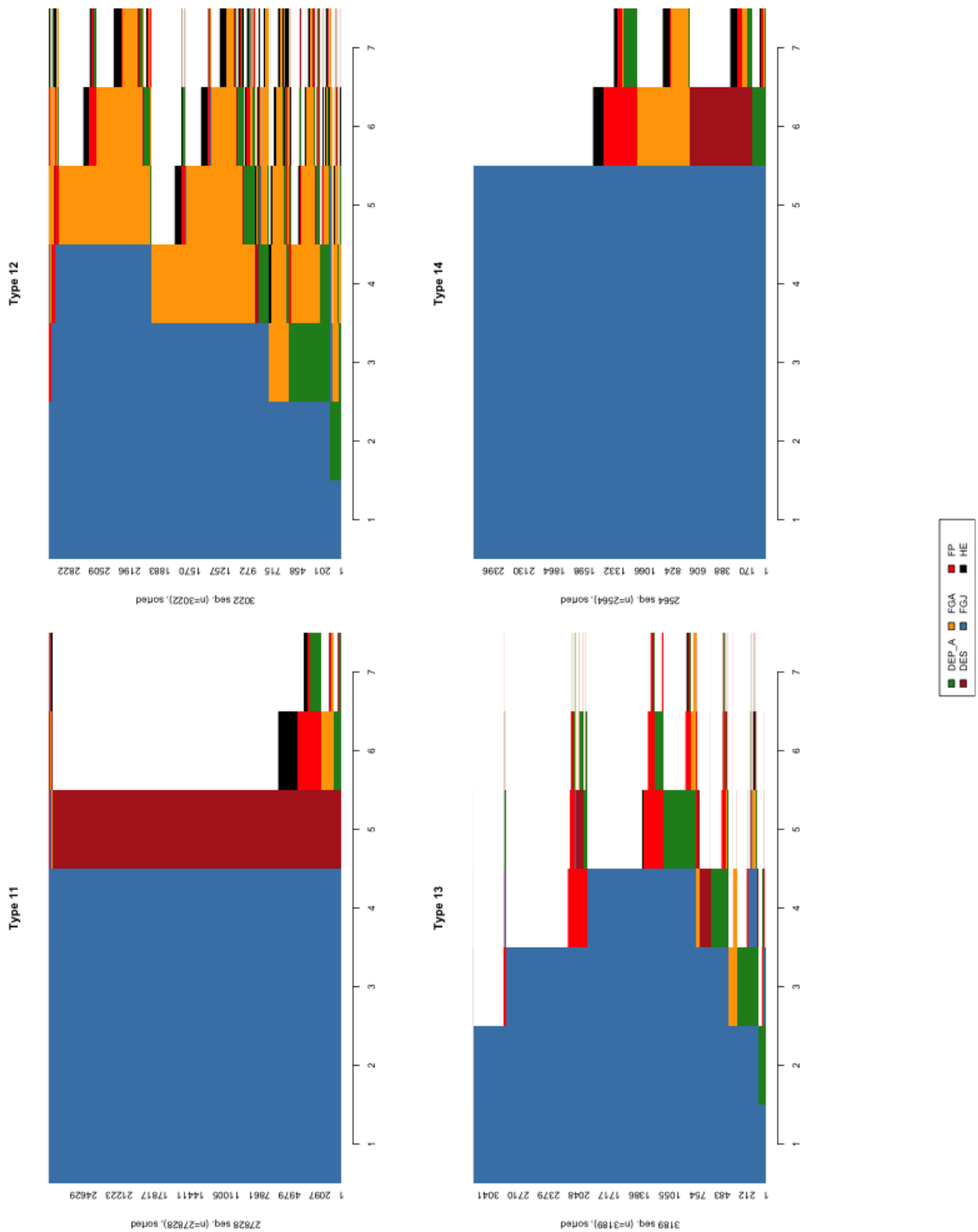
Dans le contexte de la présente étude, l'algorithme doit donc comparer chacune des **36 603** trajectoires définies sur 6 états et 7 périodes avec les **36 602** autres trajectoires. On comprend dès lors que le nombre de calculs à effectuer augmente de façon exponentielle avec le nombre de trajectoires, le nombre d'états considérés et le nombre de périodes d'observation.⁸ Au final, chaque trajectoire est associée à 36 602 coûts qui correspondent à ce qu'il en coûte minimalement de la transformer à l'identique à chacune des autres.

Les parcours individuels sont ensuite regroupés en un nombre arbitraire de sous-groupes, ou « clusters », sur la base de la proximité des coûts.⁹ Cette agrégation en sous-groupes fait en sorte que les parcours sont relativement homogènes à l'intérieur de groupes, mais relativement hétérogènes entre les groupes. Aux fins de la présente étude, nous avons arbitrairement choisi de regrouper les parcours individuels en quatre sous-groupes qui apparaissent à la Figure 2, soit les groupes *Type 11*, *Type 12*, *Type 13* et *Type 14*. Dans chacune des figures, on retrouve l'année civile correspondant à chaque élève le long de l'abscisse. Chaque élève est représenté par une ligne différente le long de cet axe. En ordonnée, on retrouve le nombre cumulatif d'élèves. Ainsi, la strate « Type 11 » regroupe plus de 27 828 élèves. Le premier nombre le long de l'ordonnée à l'abscisse, 2 097, indique que les parcours de plus de 2 097 élèves sont représentés dans la surface située sous ce nombre. Le deuxième chiffre, 4 979, s'interprète de la même façon. C'est donc dire que la surface comprise entre les deux nombres représente les parcours de plus de 2 882 élèves.

8. L'analyse des 36 603 trajectoires prend plus de 20 heures de calcul sur un ordinateur très performant et nécessite plus de 100Mo de mémoire vive. Au final, l'algorithme génère une matrice symétrique de dimension 36 603 × 36 603 dont la diagonale est 0.

9. Plusieurs algorithmes peuvent être utilisés à cette fin, bien que les résultats soient souvent insensibles au choix d'un algorithme particulier.

Figure 2 – Trajectoires individuelles des parcours scolaires



La strate « Type 11 » est de loin la plus nombreuse. Les élèves de cette strate occupent l'état FGJ durant leurs quatre premières années de secondaire (surface peinte en bleu) puis obtiennent pour la plupart un DES (surface peinte en bourgogne) à leur cinquième année. Très peu d'entre eux occupent toujours l'état FGJ à la cinquième année, et un certain nombre d'entre eux effectuent soit une transition vers la formation professionnelle (rouge) soit à la formation générale des adultes (orange). Les 3 022 élèves regroupés dans la strate « Type 12 » ont majoritairement fait des études en FP ou en FGA et très peu d'entre eux ont obtenu un DES. Certains ont obtenu un diplôme autre qu'un DES, et ce, dès la deuxième ou la troisième année du secondaire. Les 3 189 élèves de la strate « Type 13 » sont peu nombreux à avoir terminé cinq années de scolarité. Certains d'entre eux ont néanmoins obtenu un DES ou un DEP. Enfin, pratiquement aucun des 2 564 élèves du groupe « Type 14 » n'a obtenu un DES dans les délais normaux. L'obtention d'un DES ou d'un DEP survient surtout à la sixième ou à la septième année suivant l'entrée au secondaire.

La Figure 3 illustre la distribution des élèves des quatre strates à travers les différents CSS. Ainsi, la strate « Type 11 » regroupe environ 76% de la totalité des élèves de l'échantillon aléatoire. Les strates « Type 12 », « Type 13 » et « Type 14 » regroupent 8,2%, 8,7% et 7,0% des observations, respectivement. Les CSS qui regroupent proportionnellement le plus d'élèves de la strate « Type 11 » sont CSS des Découvreurs, CS Central Québec, CS English-Montréal, et CS Sir-Wilfrid-Laurier. De la même façon, les CSS des Îles, du Lac-Abitibi, des Hauts-Bois-de-l'Outaouais, et de la Baie-James regroupent plus d'élèves de la strate « Type 12 ». Enfin, c'est dans les CS Kativik et CS Crie que l'on retrouve la plus grande proportion d'élèves des strates « Type 13 » et « Type 14 ».

Il est instructif d'analyser les caractéristiques des élèves de chacune des strates. À cette fin, le Tableau 4 rapporte les caractéristiques moyennes des élèves de chacune d'elles. Ainsi, on constate que la strate « Type 11 » est composée majoritairement de filles et présente le plus faible taux de EHDAA. Les élèves ont vécu moins de changements de CSS au cours de leur scolarité et sont proportionnellement beaucoup plus nombreux à fréquenter l'école privée. Ils sont également plus nombreux à participer à un programme agréé ou un programme enrichi. Enfin, ils sont issus d'un milieu socio-économique avantageux, ce qui se reflète dans le revenu médian de leur aire de diffusion, tout comme dans un score IMSE plus faible.¹⁰ Les autres strates sont composées davantage de garçons, présentant un taux élevé de EHDAA. Très peu d'entre eux ont participé à un programme agréé ou à un programme enrichi. Enfin, ils sont tous issus d'un milieu nettement moins favorable que ceux de la strate « Type 11 ».

10. La valeur moyenne de l'IMSE est vraisemblablement surestimée car elle exclut les élèves fréquentant des écoles privées.

Figure 3 – Distribution des types d'élèves selon le CSS

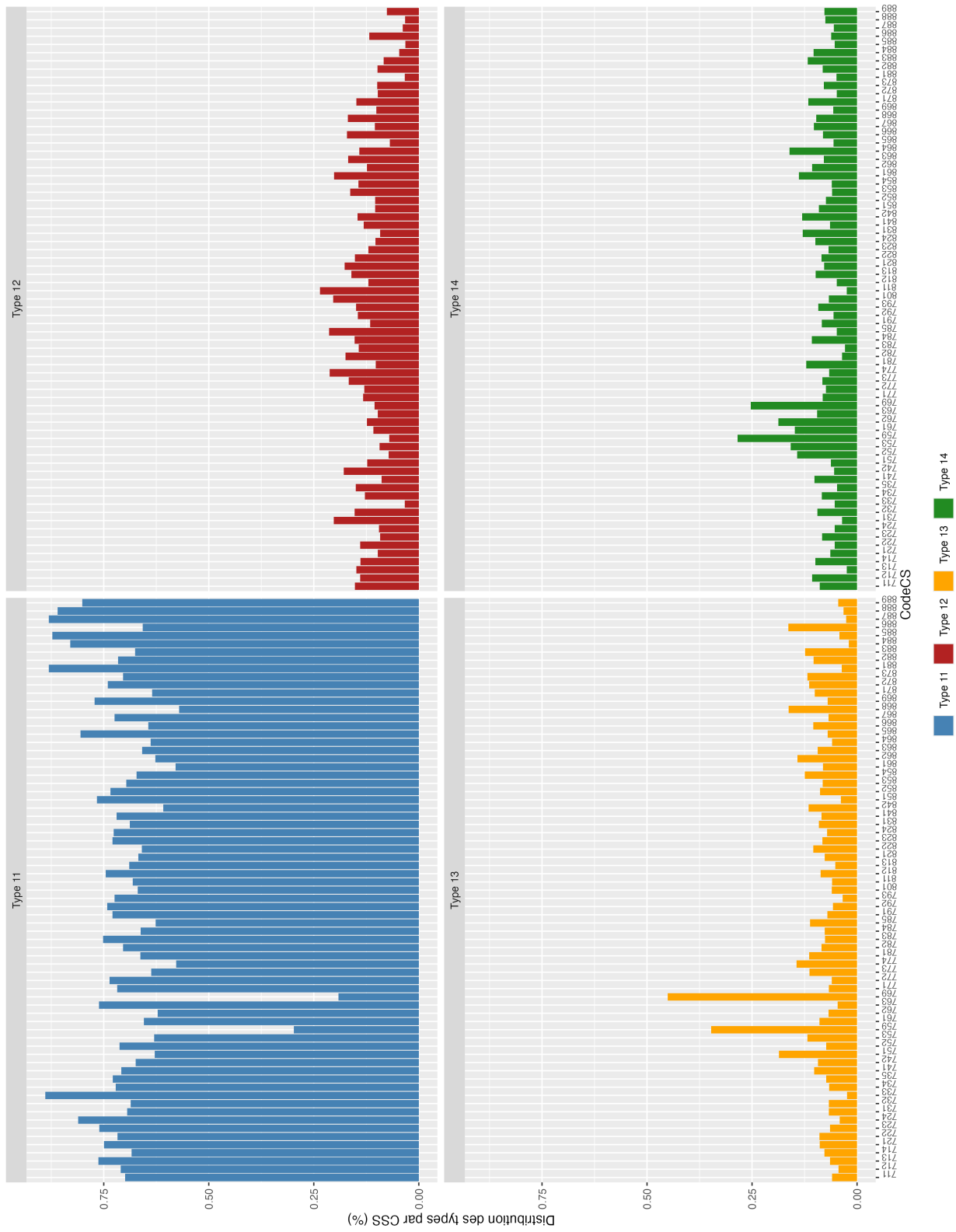


Tableau 4 – Caractéristiques des regroupements

	Type 11	Type 12	Type 13	Type 14
Caractéristiques des élèves				
Fille	0,550	0,365	0,300	0,398
Français	0,784	0,861	0,814	0,740
Montréal	0,090	0,138	0,109	0,207
EHDAA	0,103	0,397	0,541	0,361
$\Delta\text{CSS} > 0^\dagger$	0,108	0,162	0,152	0,165
École privée	0,316	0,065	0,120	0,105
Programmes				
ProgAgrée	0,143	0,007	0,014	0,023
ProgEnrichi	0,093	0,024	0,025	0,041
PassePartout	0,112	0,155	0,135	0,107
Maternelle 4 ans	0,046	0,083	0,095	0,105
Milieu socio-économique				
Revenu moyen/AD ($\times \$100,000$)	1,069	0,871	0,886	0,882
Revenu médian/AD ($\times \$100,000$)	0,967	0,831	0,816	0,831
IMSE [‡]	8,706	12,902	11,926	12,175

[†] $\Delta\text{CSS} > 0$: changement de CSS. [‡]Exclut les écoles privées.

3 Modèle en chaîne de Markov

L'analyse économique repose souvent sur la modélisation de dynamiques temporelles : cycles économiques, comportements de consommation, évolution du marché du travail, *etc.* Dans un contexte d'incertitude et d'évolution aléatoire, les chaînes de Markov d'ordre 1 constituent un outil puissant permettant de modéliser ces phénomènes de manière rigoureuse.¹¹

Considérons un modèle de Markov en temps discret avec les éléments suivants :

- Un ensemble fini d'états de dimension K : $\mathcal{S} = \{1, 2, \dots, K\}$ (FJG, FGA, FP, ...)
- Le temps est indexé de manière discrète : $t = 1, 2, \dots, T$. Il ne s'agit pas du temps calendrier, mais de la période d'observation de chaque individu.
- Pour un individu i , le processus d'états est : $Y_{it} \in \mathcal{S}$, *i.e.* l'état observé à la période t .
- Soit X_{it} un vecteur de covariables (potentiellement dépendant du temps)
- Un état absorbant noté $a \in \mathcal{S}$, tel que :

$$\mathbb{P}(Y_{it} = a \mid Y_{i,t-1} = a) = 1 \quad \text{et} \quad \mathbb{P}(Y_{it} \neq a \mid Y_{i,t-1} = a) = 0$$

11. On parle d'évolution aléatoire en ce que les trajectoires individuelles ne sont pas parfaitement prévisibles, du moins pas du point de vue de l'analyste.

Le modèle en chaîne de Markov d'ordre 1 est fondé sur l'hypothèse suivante :

$$\mathbb{P}(Y_{it} = j \mid Y_{i,t-1} = k, X_{i,t-1}).$$

Ceci implique que l'état à la période présente (t) ne dépend que de l'état et des covariables à la période précédente ($t - 1$). Les probabilités de transition sont définies par un modèle logit multinomial :

$$\mathbb{P}(Y_{it} = j \mid Y_{i,t-1} = k, X_{i,t-1}) = p_{ijk}(X_{i,t-1}) = \frac{\exp(\beta_{jk}^\top X_{i,t-1})}{\sum_{\ell=1}^K \exp(\beta_{\ell k}^\top X_{i,t-1})},$$

où β_{jk} sont les paramètres associés à la transition de l'état $k \rightarrow j$. Il y a donc autant de vecteurs de paramètres qu'il y a de transitions possibles. Cette équation fait ressortir que chaque individu est susceptible de transiter de l'état k vers l'état j à chaque période (d'où la nature aléatoire des transitions). Il est important de noter que les paramètres de la transition j vers k ne sont pas nécessairement les mêmes que ceux de la transition de k vers j .

Fonction de Vraisemblance

En supposant que l'on dispose de données pour N individus observés pendant T_i périodes, la vraisemblance s'écrit :¹²

$$L(\beta) = \prod_{i=1}^N \mathbb{P}(Y_{i1}) \prod_{t=2}^{T_i} \mathbb{P}(Y_{it} \mid Y_{i,t-1}, X_{i,t-1}).$$

Puisque la probabilité d'occuper l'état Y_{it} ne dépend que de l'état occupé à la période précédente, $Y_{i,t-1}$, les transitions successives sont indépendantes les unes des autres. La vraisemblance se limite alors au produit de leurs probabilités. Si l'état initial Y_{i1} est fixé ou indépendant des paramètres β , alors $\mathbb{P}(Y_{i1}) = 1$.¹³ On peut alors se concentrer sur les seules transitions qui dépendent des paramètres :

$$L(\beta) \propto \prod_{i=1}^N \left(\prod_{t=2}^{T_i} p_{i,Y_{it},Y_{i,t-1}}(X_{i,t-1}) \right),$$

où \propto signifie « est proportionnel à ... ». Cette équation montre que l'on doit multiplier entre elles l'ensemble des probabilités de transitions individuelles, puis de prendre le produit sur l'ensemble

12. En écrivant T_i plutôt que T on admet que l'échantillon peut ne pas être cylindré.

13. Dans notre cas, l'état initial est FGJ pour tous les élèves. Cet état ne résulte pas d'un choix, et est donc considéré comme exogène. Autrement dit, il n'y a pas de problème de **condition initiale**.

de l'échantillon. En utilisant des fonctions indicatrices, on obtient :

$$L(\beta) = \prod_{i=1}^N \prod_{t=2}^{T_i} \prod_{j=1}^K \prod_{k=1}^K [p_{ijk}(X_{i,t-1})]^{\mathbb{I}(Y_{it}=j, Y_{i,t-1}=k)}$$

En prenant le logarithme de la vraisemblance, on obtient la fonction qui est effectivement optimisée à l'aide de divers algorithmes :

$$\begin{aligned} \log L(\beta) &= \sum_{i=1}^N \sum_{t=2}^{T_i} \sum_{j=1}^K \sum_{k=1}^K \mathbb{I}(Y_{it} = j, Y_{i,t-1} = k) \cdot \log p_{ijk}(X_{i,t-1}) \\ &= \sum_{i=1}^N \sum_{t=2}^{T_i} \sum_{j=1}^K \sum_{k=1}^K \mathbb{I}(Y_{it} = j, Y_{i,t-1} = k) \left[\beta_{jk}^\top X_{i,t-1} - \log \left(\sum_{\ell=1}^K \exp(\beta_{\ell k}^\top X_{i,t-1}) \right) \right] \end{aligned}$$

Les termes associés aux transitions depuis l'état absorbant a sont exclus puisque leurs probabilités sont constantes (égales à 1 pour $a \rightarrow a$ et 0 pour tout $a \rightarrow j \neq a$).

Durée espérée dans chaque état

La durée espérée (ou espérance de séjour) dans un état transitoire $s \in \mathcal{S} \setminus \{a\}$, notée d_s , est le nombre moyen de périodes qu'un individu reste dans l'état s avant d'atteindre l'état absorbant a . On la calcule à partir de la sous-matrice de transition Q correspondant aux états transitoires :

$$Q_{jk} = \mathbb{P}(Y_{t+1} = k \mid Y_t = j), \quad j, k \in \mathcal{S} \setminus \{a\}.$$

On définit ensuite la matrice fondamentale :

$$N = (I - Q)^{-1}$$

La durée espérée dans l'état s est alors donnée par la somme de la ligne s de N :

$$d_s = \sum_k N_{sk}$$

Cette durée peut être calculée en remplaçant les probabilités de transition Q_{jk} par leurs valeurs conditionnelles moyennes ou en les évaluant pour des valeurs données des covariables X .¹⁴

14. C'est ce qui est fait dans le Tableau 5 ci-dessous.

Les transitions scolaires empiriques

Les données de notre échantillon peuvent être représentées à l'aide de matrices empiriques \hat{P} pour chacune des années suivant l'entrée au secondaire. Soit \hat{P}^t la matrice des transitions **observées** à l'année t , avec $t = 2 \dots 5$.¹⁵ À des fins d'**illustration**, nous utilisons deux échantillons différents pour calculer la matrice P . Le premier, « *échantillon filtré* », contient toutes les transitions des élèves entrés en première année de secondaire et qui sont présents dans la base de données pendant cinq années consécutives sans interruption. Cet échantillon contient 1 083 475 observations portant sur 216 695 élèves distincts. Le second échantillon, « *non filtré* », contient l'ensemble des observations portant sur les 365 169 élèves répertoriés à la Figure 1. Certains ont quitté l'échantillon prématurément, d'autres ont des années manquantes (AnCivill), *etc.* Ces élèves ont généré 1 928 425 observations. Tout comme pour l'échantillon filtré, nous nous attardons aux cinq premières années d'observation. Les transitions (exprimées en pourcentage) de la première à la deuxième année de chaque échantillon sont :

$$P^2 = \left(\begin{array}{c|cccccc} & \text{Échantillon filtré} & & & & & \\ & \hline & FGJ & FP & FGA & DES & DEP & HE \\ \hline FGJ & 100.0 & 0 & 0 & 0 & 0 & 0 \\ FP & 0 & 0 & 0 & 0 & 0 & 0 \\ FGA & 0 & 0 & 0 & 0 & 0 & 0 \\ DES & 0 & 0 & 0 & 0 & 0 & 0 \\ DEP & 0 & 0 & 0 & 0 & 0 & 0 \\ HE & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \left(\begin{array}{c|cccccc} & \text{Échantillon non filtré} & & & & & \\ & \hline & FGJ & FP & FGA & DES & DEP & HE \\ \hline FGJ & 99.42 & 0.002 & 0.045 & 0 & 0.53 & 0 \\ FP & 0 & 0 & 0 & 0 & 0 & 0 \\ FGA & 0 & 0 & 0 & 0 & 0 & 0 \\ DES & 0 & 0 & 0 & 0 & 0 & 0 \\ DEP & 0 & 0 & 0 & 0 & 0 & 0 \\ HE & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

Comme il se doit, tous les élèves de l'échantillon filtré font une transition $FGJ \rightarrow FGJ$. Il en est de même pour l'échantillon non filtré, bien que quelques transitions inattendues soient observées à destination des états FP, FGA et DEP. Des différences s'observent également entre les deuxièmes et troisièmes années :

$$P^3 = \left(\begin{array}{c|cccccc} & \text{Échantillon filtré} & & & & & \\ & \hline & FGJ & FP & FGA & DES & DEP & HE \\ \hline FGJ & 99.9 & 0.01 & 0.001 & 0 & 0.06 & 0 \\ FP & 0 & 0 & 0 & 0 & 0 & 0 \\ FGA & 0 & 0 & 0 & 0 & 0 & 0 \\ DES & 0 & 0 & 0 & 0 & 0 & 0 \\ DEP & 0 & 0 & 0 & 0 & 0 & 0 \\ HE & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \left(\begin{array}{c|cccccc} & \text{Échantillon non filtré} & & & & & \\ & \hline & FGJ & FP & FGA & DES & DEP & HE \\ \hline FGJ & 97.20 & 0.07 & 0.89 & 0 & 0.01 & 1.83 \\ FP & 0 & 60.0 & 20.0 & 0 & 0 & 20.0 \\ FGA & 0 & 1.45 & 74.64 & 1.45 & 6.52 & 15.94 \\ DES & 0 & 0 & 0 & 0 & 0 & 0 \\ DEP & 25.41 & 2.84 & 45.26 & 0 & 26.30 & 0.19 \\ HE & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

Il ne faut pas interpréter certaines transitions trop rigoureusement. Par exemple, les transitions de la matrice $P^{2 \rightarrow 3}$ hors de l'état FP calculées à partir de l'échantillon non filtré ne contiennent que 5 observations, d'où les pourcentages exacts.¹⁶ Pour les années suivantes, nous observons les

15. L'année 1 correspond à l'entrée au niveau secondaire.

16. Les transitions peu nombreuses sont exclues de l'estimation du modèle, comme on le verra plus tard.

transitions suivantes :

$$P^4 = \left(\begin{array}{c|cccccc} & \text{Échantillon filtré} & & & & & \\ \hline & FGJ & FP & FGA & DES & DEP & HE \\ \hline FGJ & 99.7 & 0.16 & 0.009 & 0.0005 & 0.001 & 0 \\ FP & 0 & 86.67 & 6.67 & 0 & 6.67 & 0 \\ FGA & 0 & 0 & 100.0 & 0 & 0 & 0 \\ DES & 0 & 0 & 0 & 0 & 0 & 0 \\ DEP & 17.3 & 0.72 & 56.12 & 0 & 25.9 & 0 \\ HE & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \left(\begin{array}{c|cccccc} & \text{Échantillon non filtré} & & & & & \\ \hline & FGJ & FP & FGA & DES & DEP & HE \\ \hline FGJ & 94.55 & 0.75 & 3.29 & 0.51 & 0.89 & 0.01 \\ FP & 0 & 30.83 & 37.97 & 1.50 & 26.32 & 3.38 \\ FGA & 0 & 1.65 & 78.68 & 1.24 & 10.12 & 8.31 \\ DES & 0 & 57.14 & 42.86 & 0 & 0 & 0 \\ DEP & 21.44 & 7.18 & 51.06 & 0 & 19.89 & 0.43 \\ HE & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

$$P^5 = \left(\begin{array}{c|cccccc} & \text{Échantillon filtré} & & & & & \\ \hline & FGJ & FP & FGA & DES & DEP & HE \\ \hline FGJ & 2.84 & 0.05 & 0.30 & 96.24 & 0.57 & 0 \\ FP & 0 & 6.20 & 2.43 & 81.40 & 9.97 & 0 \\ FGA & 0 & 2.845 & 79.00 & 4.98 & 13.17 & 0 \\ DES & 0 & 0 & 100.0 & 0 & 0 & 0 \\ DEP & 3.28 & 11.48 & 59.02 & 0 & 26.23 & 0 \\ HE & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right) \left(\begin{array}{c|cccccc} & \text{Échantillon non filtré} & & & & & \\ \hline & FGJ & FP & FGA & DES & DEP & HE \\ \hline FGJ & 8.37 & 0.88 & 3.36 & 85.99 & 1.37 & 0.04 \\ FP & 0 & 24.56 & 12.13 & 37.88 & 23.22 & 2.21 \\ FGA & 0 & 4.51 & 71.87 & 3.13 & 13.54 & 6.95 \\ DES & 0 & 47.84 & 22.26 & 0 & 7.81 & 22.09 \\ DEP & 1.19 & 34.0 & 42.82 & 0 & 17.63 & 4.36 \\ HE & 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right)$$

À partir des matrices non filtrées, on peut inférer les paramètres de base ainsi que les paramètres des variables explicatives.

4 Résultats

Les paramètres estimés sont présentés aux Annexes A (matrice des constantes de régression) et B (paramètres des variables explicatives, Tableaux B1–B4). Comme il y a de nombreux paramètres, il est préférable de les représenter graphiquement pour en faciliter la lecture.¹⁷ Les Figures 4–7 illustrent les paramètres associés aux variables explicatives ainsi que leurs intervalles de confiance pour chacune des transitions considérées dans l’analyse.¹⁸

Il y a plus de 416 paramètres de pente dans le modèle. Il serait fastidieux de les analyser tous séparément. On peut toutefois en dégager de grandes tendances. Tout d’abord, la plupart des variables exogènes ont un effet sur une ou plusieurs transitions. Seul le fait d’être né au troisième trimestre (« automne ») ou être issus des cohortes 2004–2007 ont des effets relativement mitigés sur les transitions. Les filles se distinguent nettement des garçons dans leurs trajectoires scolaires, tout comme le fait d’avoir le français comme langue maternelle. Les élèves ayant des besoins particuliers (« EHDAA ») et ceux ayant connu plus d’un Centre de services scolaire

17. Les constantes de régression ne sont pas particulièrement intéressantes puisqu’elles n’ont pas d’interprétation particulière. On se limite donc à analyser les paramètres des variables explicatives.

18. Pour rappel, un paramètre égal à un (statistiquement) n’a pas d’effet sur la transition à laquelle il se rapporte. Un paramètre égal à 1.2 implique que la variable à laquelle il est rattaché augmente la transition de 20 % par rapport à la moyenne, *ceteris paribus*. À l’inverse, un paramètre égal à 0.8 diminue la transition de 20 %.

(« NCS ») au cours de leur cheminement scolaire ont également des parcours distinctifs. Comme on doit s’y attendre, les élèves ayant participé à un programme agréé (« ProgAgréé ») ou un programme enrichi (« ProgEnrichi ») ont des probabilités nettement plus faibles de transiter vers la formation professionnelle (« FP ») ou la formation générale des adultes (« FGA »). Peu de paramètres associés à la maternelle 4 ans (« Matern4Ans ») ont des effets statistiquement significatifs. Lorsqu’ils le sont, la participation est associée à davantage de transitions vers les états FGA et DEPA (diplômes autres) et diminue quelque peu les transitions vers l’état DES. D’un autre côté, davantage de paramètres associés au programme Passe-Partout sont statistiquement significatifs. La participation à ce programme favorise les transitions vers les états FP, DES et DEP, et diminuent surtout celles vers l’état HE. Enfin, les élèves issus des secteurs caractérisés par des IMSE élevés (IMSE2–IMSE4) ont des transitions nettement plus marquées vers les états FP, FGA et DEP et moindres vers l’état DES.¹⁹

Simulations

Les modèles en chaîne de Markov peuvent être utilisés pour calculer la durée espérée de séjour dans chaque état selon différentes configurations de variables. Cela permet de quantifier l’effet des variables explicatives sur les durées à travers leurs interactions sur les différentes transitions. Pour les besoins du présent rapport, nous allons simuler *quatre scénarios différents*²⁰ :

1. Toutes les variables explicatives sont mesurées à leurs moyennes ;
2. Élèves issus de milieux aisés : RevMed=1, NCS=0, dProgAgréé=1, dProgEnrichi=0 ;
3. Élèves issus de milieux aisés : RevMed=1, NCS=0, dProgAgréé=0, dProgEnrichi=1 ;
4. Élèves issus de milieux défavorisés : dPassePartout=1, NCS=1, dEHDA=1, IMSE4=1 ;
5. Élèves issus de milieux défavorisés : dMater4anDT=1, NCS=1, dEHDA=1, IMSE4=1.

La seule différence entre les scénarios 3 et 4 tient à la participation à Passe-Partout (Scénario 3) ou à la Maternelle 4 ans (Scénario 4). Le Tableau 5 montre la durée **totale** attendue pour chaque état du modèle selon les valeurs prises par les variables explicatives des quatre scénarios énumérés ci-dessus. On constate que la durée attendue dans l’état FGJ varie considérablement d’un scénario à l’autre. Lorsque toutes les variables sont évaluées à leur moyenne (Scénario 1), la durée attendue est de 4,74 années. La durée est inférieure à 5 car un certain nombre d’élèves quittent la FGJ pour transiter vers FP ou FGA. La durée en FGJ des élèves issus des milieux aisés est inférieure à celle du Scénario 1. Comme toutes les durées attendues du Scénario 2 sont

19. Ceux qui habitent des régions caractérisées par un revenu supérieur à la médiane ont des parcours davantage orientés vers l’état HE, ce qui peut être interprété comme favorisant la poursuite du parcours scolaire secondaire à l’extérieur du Québec. Ce résultat est *a priori* étonnant. Voir note de bas de page # 21 ci-après.

20. Les variables non mentionnées sont évaluées à zéro, sauf Français qui vaut 1 dans tous les scénarios et dSexeF=0,5012, la moyenne échantillonnale.

Tableau 5 – Effets des variables explicatives sur la durée espérée

	Scénario 1	Scénario 2	Scénario 3	Scénario 4	Scénario 5
FGJ	4,737 (0,009)	2,87 (0,026)	2,365 (0,020)	5,239 (0,061)	5,838 (0,056)
FP	1,392 (0,015)	1,212 (0,036)	1,122 (0,036)	1,291 (0,061)	1,243 (0,056)
FGA	3,346 (0,057)	1,931 (0,079)	1,930 (0,079)	2,113 (0,142)	2,597 (0,166)
DES	1,080 (0,005)	0,903 (0,020)	0,797 (0,017)	1,076 (0,022)	1,071 (0,028)
DEP	1,463 (0,022)	0,860 (0,047)	0,860 (0,047)	1,816 (0,130)	1,677 (0,126)

Écarts-types entre parenthèses

inférieures à celles du Scénario 1, on peut conclure que ces élèves diplôment plus rapidement et transitent vers l'état absorbant (HE).²¹ Enfin, les élèves défavorisés des scénarios 3 et 4 ont des durées attendues plus grandes que celles des élèves du Scénario 2 dans tous les états. Cela indique vraisemblablement que ces élèves redoublent au moins une année de FGJ.

Discussion

Le modèle est estimé à partir de l'échantillon non filtré tel que décrit dans la Figure 1. Or, cet échantillon contient de nombreuses entrées et sorties en cours de parcours dont il faut tenir compte dans l'interprétation des résultats du modèle. Parmi les 365 169 élèves considérés dans l'analyse, plus de 269 219 ont des observations complètes, c'est-à-dire que les informations sont renseignées à chacune des années sans interruption. En revanche, 59 310 élèves présentent une année de données manquante, et plus de 30 955 en compte deux. Enfin, plus de 13 605 (3 024) élèves ont trois (quatre) années manquantes entre leur entrée dans l'échantillon et leur sortie définitive.

Pour illustrer la dynamique particulière de l'échantillon non filtré, le tableau 6 ci-après rapporte l'âge des élèves à leur première entrée dans l'échantillon. On voit clairement que la grande majorité des élèves sont observés pour la première fois à l'âge de douze ans (325 048). Un nombre

21. Ces élèves ont présumément des caractéristiques semblables à celles des élèves du Type 11 dans le Tableau 4. Par ailleurs, il ne faut perdre de vue que certains élèves des cohortes à l'étude ont pu quitter le système scolaire québécois pour une période, ce qui pourrait expliquer leur réintégration aux observations plus tardivement qu'en secondaire 1. Enfin, les élèves des scénarios 2 et 3 ont des profils relativement avantageux. Il est possible que la courte durée dans l'état FGJ soit due à une sortie du réseau scolaire québécois à la faveur d'une scolarité ailleurs au Canada ou à l'étranger.

Tableau 6 – Distribution de l’âge à l’entrée dans l’échantillon

Année \ Âge	11	12	13	14	15	16	17	18	19
2007	3	0	0	0	0	0	0	0	0
2008	280	210	0	0	0	0	0	0	0
2009	844	56943	544	0	0	0	0	0	0
2010	841	54689	6092	56	0	0	0	0	0
2011	760	49201	5836	151	2	0	0	0	0
2012	817	53766	10558	784	492	5	0	0	0
2013	855	54198	6414	362	145	70	1	0	0
2014	550	55989	6713	178	99	25	18	0	0
2015	0	52	6391	195	90	19	4	0	0
2016	0	0	14	786	660	161	36	9	1
2017	0	0	0	0	84	55	27	14	0
2018	0	0	0	0	0	11	16	5	13
2019	0	0	0	0	0	0	8	6	3
2020	0	0	0	0	0	0	0	4	1
2021	0	0	0	0	0	0	0	0	5
Total	4 950	325 048	42 562	2 512	1 572	346	110	38	23

non négligeable d’élèves sont observés à l’âge de 13 ans (42 562) et dans une moindre mesure à l’âge de 11 ans (4 950). Par ailleurs, plus de 4 601 élèves ont fait leur entrée entre l’âge de 14 et 19 ans. Ces entrées tardives dans l’échantillon affectent principalement les transitions au cours de périodes plus avancées dans les parcours scolaires, vraisemblablement vers les états FP et FGA.

L’analyse de l’âge à la sortie définitive de l’échantillon est également instructive. Le Tableau 7 est structuré de façon similaire au tableau précédent. Comme on doit s’y attendre, la majorité des élèves sortent de l’échantillon à l’âge de 16 ans. Les nombreuses sorties à l’âge de 17 ans sont cohérentes avec le nombre élevé d’entrées à l’âge de 13 ans (42 562 entrées *vs* 39 930 sorties). Le tableau montre également que la dynamique des parcours scolaires est caractérisée par de nombreuses sorties hâtives et tardives. En effet, de nombreux élèves quittent l’échantillon à 12, 13 ou 14 ans, alors que d’autres sont présents jusqu’aux âges de 18-20 ans. Cette dynamique complexe des entrées et sorties tardives ou hâtives a comme conséquence que le nombre d’années, pas nécessairement consécutives, durant lesquelles on observe les étudiants est très variable.²²

Les échantillons dans lesquels tous les individus ne sont pas observés un même nombre de périodes sont dits « non cylindrés (“unbalanced”) ». Heureusement, de nombreux modèles économétriques peuvent traiter de telles données, dont le modèle de transitions en chaîne de Markov. En revanche, aucun estimateur ne permet de prendre en compte de façon satisfaisante les périodes

22. Nombre d’années d’observation : 2 ans \rightarrow 3 559 ; 3 ans \rightarrow 9 130 ; 4 ans \rightarrow 13 877 ; 5 ans \rightarrow 237 791 ; 6 ans \rightarrow 56 230 ; 7 ans 2 \rightarrow 44 582.

Tableau 7 – Distribution de l'âge à la sortie de l'échantillon

Année \ Âge	Âge									
	12	13	14	15	16	17	18	19	20	
2008	1	0	0	0	0	0	0	0	0	0
2009	2	6	0	0	0	0	0	0	0	0
2010	3	349	19	0	0	0	0	0	0	0
2011	16	727	1109	39	0	0	0	0	0	0
2012	5	211	506	1527	126	0	0	0	0	0
2013	11	214	488	1970	38859	260	0	0	0	0
2014	14	225	473	1885	37873	6569	126	0	0	0
2015	4	216	520	1944	37881	6558	5928	82	0	0
2016	0	5	532	2008	38703	6744	5612	4501	162	0
2017	0	0	19	2165	39724	6698	5609	4195	5203	0
2018	0	0	0	476	40959	6733	5350	3937	4892	0
2019	0	0	0	0	47	6351	5061	3441	4372	0
2020	0	0	0	0	0	17	6070	4052	4576	0
2021	0	0	0	0	0	0	27	5336	4327	0
2022	0	0	0	0	0	0	0	21	2490	0
Total	56	1 953	3 666	12 014	234 172	39 930	33 783	25 565	26 022	0

durant lesquelles un individu s'absente d'un échantillon. Ces absences correspondent *stricto sensu* à des valeurs manquantes.²³ Plusieurs alternatives peuvent être envisagées quant au choix du type de modélisation :

1. Modéliser uniquement les périodes contigües et omettre celles suivant la première sortie de l'échantillon ;
2. Omettre les individus pour lesquels des périodes d'absence sont observées ;
3. Dans le contexte du modèle en chaîne de Markov, considérer l'absence comme un état en soi et supposer que les variables observées à la période précédant l'absence sont valides durant les périodes d'absence ;
4. Ignorer les années d'absence et considérer uniquement les périodes durant lesquelles les individus sont observés comme étant contigües.

Aucune des alternatives ci-dessus n'est entièrement satisfaisante. Elles sont toutes susceptibles de générer des biais d'estimation plus ou moins importants. Dans le contexte du présent travail,

23. Bien que la très grande majorité des périodes d'observation (« AnCivill ») se succèdent normalement, on retrouve tout de même plus de 90 508 périodes séparées de deux années, 19 819 séparées de trois années et 6 709 séparées de quatre années. On dénombre également un petit nombre d'observations séparées par six, sept et jusqu'à huit années. Le Ministère n'est toutefois pas en mesure d'identifier avec certitude les cas de décrochage scolaire, certains élèves pouvant quitter le système scolaire pour des raisons autres telles qu'un déménagement, une maladie grave ou un décès. Ce travail n'a pas été fait. C'est pourquoi l'analyse ne prend pas en compte ce phénomène.

l'alternative #4 a été retenue pour deux raisons. Tout d'abord, elle est la plus simple à implémenter. Deuxièmement, elle est susceptible d'être la moins biaisée des quatre alternatives car elle n'élimine pas d'observations possiblement non représentatives de l'ensemble des élèves (alternatives #1 et #2) et ne fait pas d'hypothèses vraisemblablement fortes à propos des périodes d'absence parfois relativement longues (alternative #3. Voir note de bas de page #23).

Les résultats présentés dans cette étude doivent être interprétés à la lumière de ce qui précède. Les critères de sélection de l'échantillon retenu aux fins de l'analyse sont minimalistes.²⁴ On peut penser que l'échantillon est représentatif de la dynamique d'entrée et de sortie de la population d'élèves du secondaire. Toutefois, il est possible d'envisager d'autres critères de sélection, lesquels vont générer un échantillon correspondant à un sous-ensemble différent de cette même population. Le choix d'un échantillon plutôt qu'un autre va nécessairement impacter les paramètres estimés, à moins, bien sûr, que les critères soient entièrement exogènes aux trajectoires scolaires, c'est-à-dire non corrélés aux caractéristiques observables et inobservables des élèves. Ce choix relève des analystes du Ministère et non de l'auteur de la présente étude.

Enfin, il faut prendre garde d'interpréter les paramètres estimés comme des effets causaux des variables auxquelles ils sont associés sur les transitions entre les différents états scolaires. Les paramètres captent davantage l'hétérogénéité des parcours qui peut être imputée aux caractéristiques des élèves et à leur statut de participation à des programmes tels *Passe-Partout* ou *Maternelle 4 ans* (possiblement endogène). Le Tableau 8 ci-après est identique au Tableau 2 mais distingue les participants aux deux programmes des élèves non participants. De toute évidence, les participants aux programmes ont des caractéristiques individuelles et sont issus de milieux moins favorables à la réussite scolaire. Ainsi, attribuer à l'un ou l'autre des programmes des effets néfastes sur la réussite scolaire (DES ou DEP) est inexact. Les paramètres associés à ces programmes captent principalement l'hétérogénéité dans les parcours scolaires qui n'est pas prise en compte par les autres variables incluses dans le modèle. Par ailleurs, les parents qui inscrivent leurs enfants à l'un ou l'autre des programmes ont peut-être des caractéristiques inobservables qui favorisent leur réussite scolaire. Pour s'en convaincre, il suffit de simuler les trajectoires scolaires des enfants ayant les caractéristiques des participants aux deux programmes (colonne (2) et (3), respectivement) et de comparer leur situation selon qu'ils participent ou non au programme. Dans le cas de *Passe-Partout*, la participation est associée à un léger accroissement du temps passé en FGJ. Toutefois, on n'observe pas vraiment de changement dans les trajectoires des participants au programme *Maternelle 4 ans*. Cela est vraisemblablement dû au fait que peu de paramètres associés au programme sont statistiquement significatifs. Ainsi, on peut présumer que si les élèves qui n'ont pas participé à l'un ou l'autre des deux programmes avaient les mêmes caractéristiques que les participants, ils auraient sensiblement les mêmes trajectoires scolaires.

24. Voir Figure 1.

Tableau 8 – Caractéristiques des élèves selon la participation (ou non) à *Passe-Partout* et *Maternelle 4 ans*

	Aucun (1)	Passe-Partout (2)	Maternelle (3)
Caractéristiques			
Fille	0,505	0,504	0,766
Français	0,776	0,970	0,540
Montréal	0,104	0,006	0,345
# CSS	1,138	1,110	1,141
École privée	0,292	0,137	0,130
EHDAA	0,191	0,209	0,333
Programmes			
ProgAgréé	0,125	0,082	0,087
ProgEnrichi	0,097	0,087	0,042
Milieu socio-économique			
Revenu moyen AD ($\times 100\ 000$)	1,002	0,810	0,766
IMSE	9,110	13,270	13,958
Nombre d'élèves	300 949	42 574	21 646

5 Conclusion

Ce rapport s'intéresse aux trajectoires scolaires au niveau secondaire de tous les élèves entrés en première année du cycle primaire au cours des années 2003 à 2008. Six états potentiels sont considérés, dont trois sont transitoires ("Transient") : FGJ, FP, FGA, deux sont diplômants : DES, DEP, et un est dit « absorbant », HE.

Une analyse sommaire des trajectoires individuelles à l'aide d'outils de séquençage montre que, bien que celles-ci soient très hétérogènes, elles peuvent néanmoins être regroupées en un certain nombre de « clusters » relativement homogènes. Une analyse des trajectoires propres à chaque cluster montre qu'il existe tout de même une forte hétérogénéité entre les centres de services scolaires. En particulier, cette hétérogénéité semble associée aux caractéristiques des élèves qui les fréquentent.

Afin d'isoler les facteurs associés aux transitions scolaires, nous faisons appel à un modèle dit de « chaîne de Markov ». Ce type de modèle est tout à fait approprié pour étudier les transitions entre différents états à des moments discrets (année scolaire).

Il ressort de l'estimation du modèle que les trajectoires des élèves sont intimement liées à leurs caractéristiques observables, mais également au milieu socio-économique d'où ils sont issus. L'avantage d'un tel modèle est de pouvoir quantifier la contribution de chaque caractéristique à l'évolution du parcours scolaire des élèves.

Comme toujours, les résultats d'un tel exercice reposent sur un échantillon sélectionné à partir de critères spécifiques et de choix méthodologiques particuliers. Les limites et les conséquences de notre approche ont été évoquées dans la section précédente. D'autres critères et d'autres méthodologies pourraient éventuellement être envisagés. Mais, compte tenu de l'ampleur d'un tel travail, cela devra être fait dans le cadre d'une nouvelle étude.

Figure 4 – Paramètres : Caractéristiques des élèves

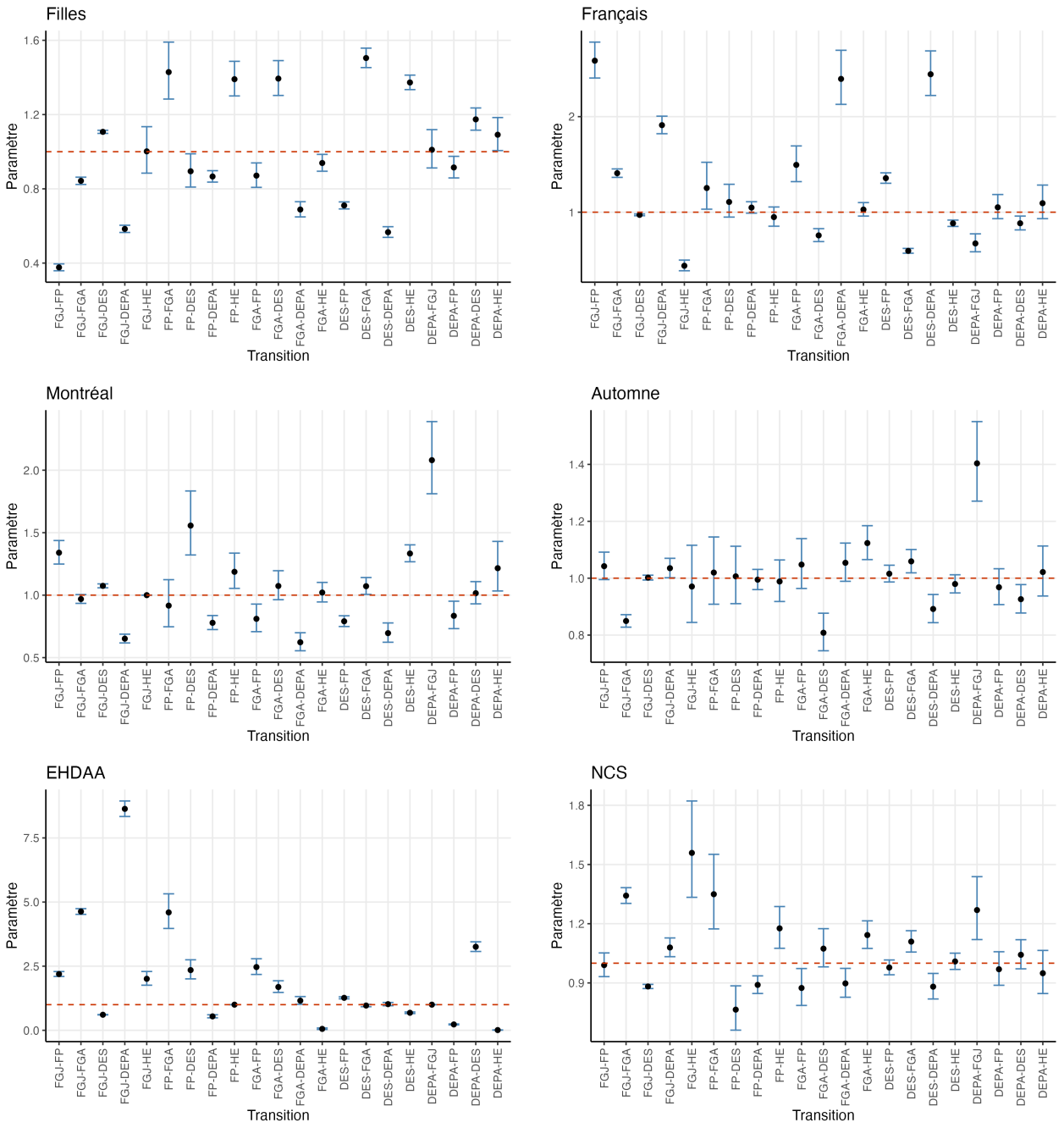


Figure 5 – Paramètres : Programmes

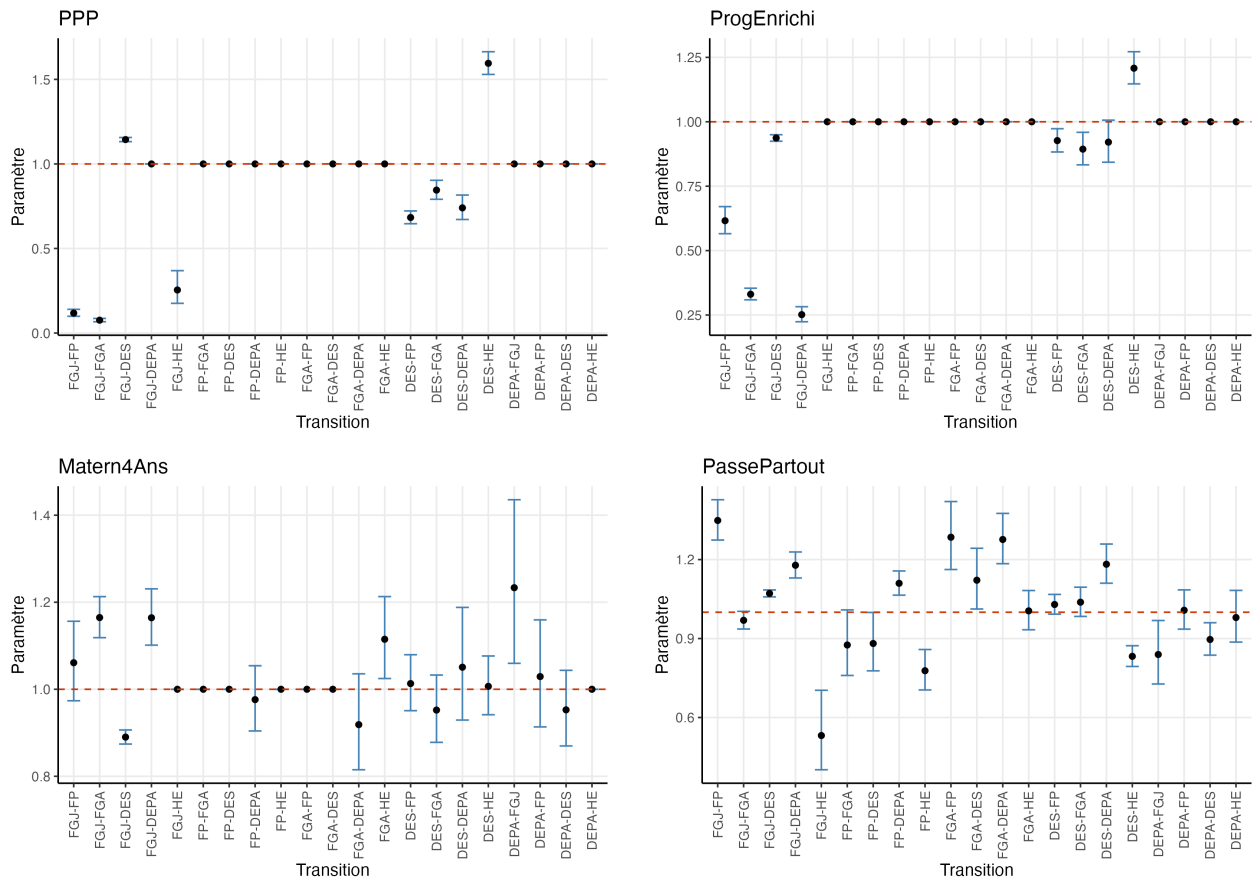


Figure 6 – Paramètres : Milieu socio-économique

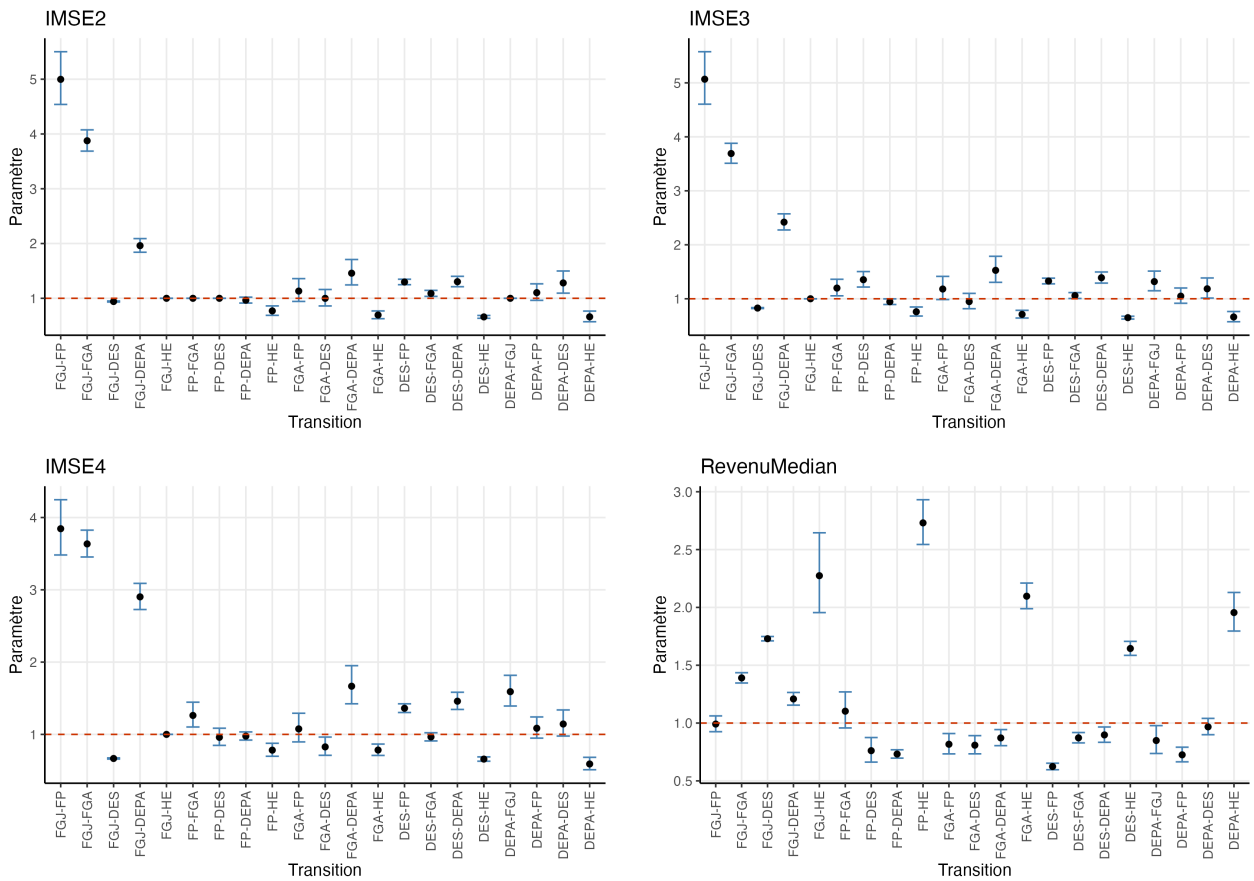
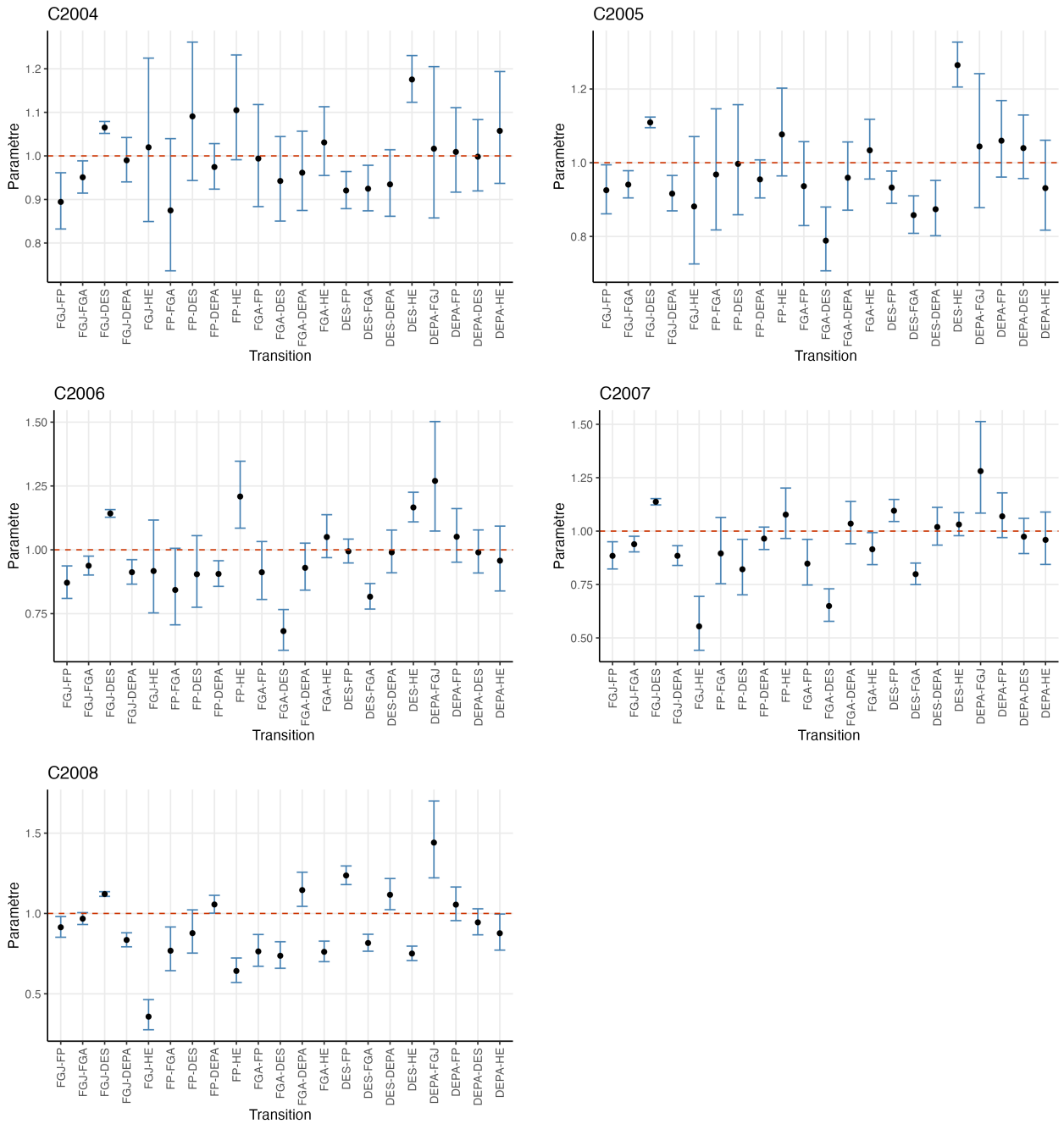


Figure 7 – Paramètres : Cohortes



A Paramètres de base (Matrice P)

Tableau A1 – Paramètres de base (Matrice P)

	FGJ	FP	FGA	DES	DEP	HE
FGJ	-0,2111 (4,10E-04)	0,0035 (6,32E-05)	0,0109 (1,20E-04)	0,1908 (3,79E-04)	0,0054 (7,27E-05)	0,0005 (2,17E-05)
FP		-0,7183 (7,60E-03)	0,0584 (1,88E-03)	0,0716 (2,12E-03)	0,4680 (6,60E-03)	0,1204 (2,49E-03)
FGA		0,0614 (1,68E-03)	-0,2989 (5,06E-03)	0,0770 (1,82E-03)	0,0816 (2,09E-03)	0,0789 (3,88E-03)
DES		0,3482 (2,58E-03)	0,2025 (1,91E-03)	-0,9256 (4,20E-03)	0,0904 (1,39E-03)	0,2846 (2,31E-03)
DEP	0,0780 (2,58E-03)	0,2890 (6,38E-03)	0,2055 (5,48E-03)		-0,6833 (1,02E-02)	0,1108 (5,22E-03)

Écart-types entre parenthèses

B Paramètres des variables explicatives (β)

Tableau B1 – Caractéristiques des élèves

	Fille		Français		Montréal	
FGJ-FP	0,377 *	(0,359 , 0,396)	2,585 *	(2,404 , 2,780)	1,340 *	(1,249 , 1,437)
FGJ-FGA	0,843 *	(0,824 , 0,863)	1,408 *	(1,365 , 1,453)	0,969	(0,934 , 1,005)
FGJ-DES	1,107 *	(1,099 , 1,115)	0,972 *	(0,963 , 0,981)	1,074 *	(1,059 , 1,090)
FGJ-DEP	0,584 *	(0,565 , 0,604)	1,911 *	(1,821 , 2,005)	0,652 *	(0,618 , 0,688)
FGJ-HE	1,002	(0,885 , 1,135)	0,439 *	(0,387 , 0,499)	1,000 *	(1,000 , 1,000)
FP-FGA	1,429 *	(1,284 , 1,590)	1,253 *	(1,031 , 1,522)	0,916	(0,747 , 1,124)
FP-DES	0,895 *	(0,810 , 0,989)	1,107	(0,949 , 1,292)	1,557 *	(1,322 , 1,834)
FP-DEP	0,866 *	(0,836 , 0,898)	1,048	(0,990 , 1,110)	0,779 *	(0,725 , 0,837)
FP-HE	1,391 *	(1,300 , 1,487)	0,949	(0,853 , 1,055)	1,187 *	(1,054 , 1,336)
FGA-FP	0,871 *	(0,808 , 0,940)	1,495 *	(1,320 , 1,693)	0,811 *	(0,708 , 0,928)
FGA-DES	1,394 *	(1,303 , 1,491)	0,757 *	(0,692 , 0,828)	1,073	(0,964 , 1,196)
FGA-DEP	0,689 *	(0,649 , 0,731)	2,395 *	(2,129 , 2,695)	0,623 *	(0,555 , 0,699)
FGA-HE	0,939 *	(0,895 , 0,986)	1,028	(0,959 , 1,101)	1,021	(0,946 , 1,102)
DES-FP	0,710 *	(0,692 , 0,730)	1,356 *	(1,303 , 1,412)	0,791 *	(0,749 , 0,835)
DES-FGA	1,505 *	(1,453 , 1,558)	0,595 *	(0,571 , 0,621)	1,072 *	(1,006 , 1,141)
DES-DEP	0,567 *	(0,539 , 0,596)	2,444 *	(2,222 , 2,688)	0,696 *	(0,623 , 0,778)
DES-HE	1,373 *	(1,334 , 1,413)	0,884 *	(0,851 , 0,918)	1,333 *	(1,267 , 1,403)
DEP-FGJ	1,010	(0,912 , 1,119)	0,673 *	(0,586 , 0,774)	2,080 *	(1,812 , 2,389)
DEP-FP	0,915 *	(0,859 , 0,975)	1,051	(0,932 , 1,185)	0,835 *	(0,732 , 0,952)
DEP-DES	1,174 *	(1,116 , 1,236)	0,884 *	(0,815 , 0,959)	1,016	(0,930 , 1,108)
DEP-HE	1,091 *	(1,006 , 1,184)	1,094	(0,932 , 1,284)	1,216 *	(1,034 , 1,430)
	Troisième trimestre		EHDAA		Δ CSS	
FGJ-FP	1,042	(0,995 , 1,092)	2,193 *	(2,100 , 2,290)	0,990	(0,932 , 1,052)
FGJ-FGA	0,850 *	(0,828 , 0,872)	4,631 *	(4,521 , 4,744)	1,342 *	(1,303 , 1,383)
FGJ-DES	1,002	(0,994 , 1,011)	0,607 *	(0,600 , 0,614)	0,882 *	(0,872 , 0,893)
FGJ-DEP	1,035 *	(1,002 , 1,070)	8,633 *	(8,336 , 8,941)	1,079 *	(1,033 , 1,128)
FGJ-HE	0,971	(0,845 , 1,116)	2,008 *	(1,758 , 2,293)	1,559 *	(1,334 , 1,822)
FP-FGA	1,020	(0,909 , 1,145)	4,596 *	(3,971 , 5,321)	1,350 *	(1,174 , 1,552)
FP-DES	1,006	(0,911 , 1,112)	2,348 *	(2,005 , 2,749)	0,765 *	(0,661 , 0,885)
FP-DEP	0,995	(0,960 , 1,031)	0,543 *	(0,488 , 0,606)	0,890 *	(0,846 , 0,936)
FP-HE	0,988	(0,918 , 1,064)	1,000 *	(1,000 , 1,000)	1,177 *	(1,076 , 1,287)
FGA-FP	1,048	(0,964 , 1,139)	2,466 *	(2,178 , 2,791)	0,874 *	(0,786 , 0,973)
FGA-DES	0,808 *	(0,745 , 0,877)	1,689 *	(1,477 , 1,932)	1,074	(0,981 , 1,175)
FGA-DEP	1,054	(0,989 , 1,123)	1,155 *	(1,014 , 1,316)	0,898 *	(0,828 , 0,974)
FGA-HE	1,123 *	(1,065 , 1,185)	0,055 *	(0,034 , 0,090)	1,143 *	(1,075 , 1,215)
DES-FP	1,016	(0,987 , 1,046)	1,266 *	(1,227 , 1,306)	0,978	(0,941 , 1,016)
DES-FGA	1,059 *	(1,019 , 1,101)	0,965	(0,921 , 1,012)	1,109 *	(1,056 , 1,165)
DES-DEP	0,892 *	(0,844 , 0,943)	1,020	(0,959 , 1,085)	0,881 *	(0,819 , 0,948)
DES-HE	0,980	(0,948 , 1,012)	0,683 *	(0,653 , 0,715)	1,009	(0,968 , 1,051)
DEP-FGJ	1,404 *	(1,271 , 1,551)	1,000 *	(1,000 , 1,000)	1,269 *	(1,120 , 1,439)
DEP-FP	0,968	(0,907 , 1,033)	0,229 *	(0,212 , 0,246)	0,969	(0,888 , 1,058)
DEP-DES	0,926 *	(0,878 , 0,978)	3,257 *	(3,074 , 3,450)	1,043	(0,971 , 1,119)
DEP-HE	1,022	(0,937 , 1,113)	0,009 *	(0,006 , 0,014)	0,949	(0,846 , 1,065)

Tableau B2 – Programmes

	PPP	ProgEnrichi	Mater4anDT	PassePartout
FGJ-FP	0,118 * (0,099 , 0,140)	0,616 * (0,566 , 0,671)	1,061 (0,974 , 1,156)	1,349 * (1,274 , 1,427)
FGJ-FGA	0,076 * (0,067 , 0,086)	0,331 * (0,309 , 0,354)	1,165 * (1,119 , 1,213)	0,969 (0,936 , 1,003)
FGJ-DES	1,144 * (1,132 , 1,156)	0,937 * (0,925 , 0,950)	0,890 * (0,874 , 0,907)	1,071 * (1,058 , 1,085)
FGJ-DEP		0,252 * (0,224 , 0,282)	1,164 * (1,101 , 1,231)	1,178 * (1,130 , 1,229)
FGJ-HE	0,255 * (0,176 , 0,369)			0,532 * (0,402 , 0,703)
FP-FGA				0,876 (0,760 , 1,009)
FP-DES				0,881 * (0,777 , 0,999)
FP-DEP			0,976 (0,904 , 1,054)	1,110 * (1,065 , 1,157)
FP-HE				0,778 * (0,705 , 0,858)
FGA-FP				1,285 * (1,162 , 1,421)
FGA-DES				1,122 * (1,012 , 1,243)
FGA-DEP			0,919 (0,815 , 1,036)	1,276 * (1,184 , 1,375)
FGA-HE			1,115 * (1,025 , 1,213)	1,005 (0,933 , 1,082)
DES-FP	0,683 * (0,646 , 0,722)	0,927 * (0,883 , 0,973)	1,013 (0,951 , 1,079)	1,029 (0,993 , 1,068)
DES-FGA	0,845 * (0,791 , 0,903)	0,894 * (0,833 , 0,959)	0,952 (0,878 , 1,033)	1,038 (0,984 , 1,095)
DES-DEP	0,740 * (0,671 , 0,816)	0,921 (0,843 , 1,006)	1,051 (0,929 , 1,188)	1,182 * (1,110 , 1,259)
DES-HE	1,595 * (1,529 , 1,664)	1,208 * (1,147 , 1,272)	1,007 (0,942 , 1,077)	0,832 * (0,794 , 0,873)
DEP-FGJ			1,233 * (1,060 , 1,436)	0,839 * (0,727 , 0,969)
DEP-FP			1,029 (0,914 , 1,160)	1,008 (0,936 , 1,085)
DEP-DES			0,953 (0,870 , 1,044)	0,896 * (0,837 , 0,960)
DEP-HE				0,980 (0,887 , 1,083)

Tableau B3 – Milieux socio-économique

	IMSE2		IMSE3		IMSE4		Revenu Median	
FGJ-FP	4,998 *	(4,539 , 5,502)	5,068 *	(4,605 , 5,577)	3,845 *	(3,482 , 4,246)	0,991	(0,925 , 1,062)
FGJ-FGA	3,876 *	(3,687 , 4,076)	3,691 *	(3,510 , 3,880)	3,635 *	(3,454 , 3,825)	1,390 *	(1,346 , 1,435)
FGJ-DES	0,941 *	(0,932 , 0,951)	0,828 *	(0,819 , 0,836)	0,667 *	(0,659 , 0,675)	1,729 *	(1,711 , 1,748)
FGJ-DEP	1,962 *	(1,841 , 2,090)	2,419 *	(2,275 , 2,573)	2,903 *	(2,727 , 3,089)	1,209 *	(1,155 , 1,265)
FGJ-HE	1,000 *	(1,000 , 1,000)	1,000 *	(1,000 , 1,000)	1,000 *	(1,000 , 1,000)	2,274 *	(1,955 , 2,645)
FP-FGA	1,000 *	(1,000 , 1,000)	1,198 *	(1,055 , 1,361)	1,262 *	(1,102 , 1,445)	1,103	(0,958 , 1,269)
FP-DES	1,000 *	(1,000 , 1,000)	1,353 *	(1,217 , 1,504)	0,959	(0,848 , 1,086)	0,761 *	(0,662 , 0,875)
FP-DEP	0,966	(0,913 , 1,021)	0,945 *	(0,894 , 0,998)	0,976	(0,921 , 1,034)	0,732 *	(0,697 , 0,769)
FP-HE	0,770 *	(0,689 , 0,861)	0,759 *	(0,680 , 0,848)	0,782 *	(0,697 , 0,877)	2,731 *	(2,544 , 2,931)
FGA-FP	1,132	(0,943 , 1,359)	1,181	(0,985 , 1,415)	1,076	(0,896 , 1,293)	0,817 *	(0,734 , 0,909)
FGA-DES	0,999	(0,860 , 1,161)	0,947	(0,816 , 1,100)	0,827 *	(0,711 , 0,963)	0,809 *	(0,734 , 0,891)
FGA-DEP	1,458 *	(1,245 , 1,709)	1,527 *	(1,304 , 1,787)	1,666 *	(1,424 , 1,951)	0,871 *	(0,805 , 0,944)
FGA-HE	0,695 *	(0,628 , 0,769)	0,711 *	(0,643 , 0,786)	0,784 *	(0,709 , 0,866)	2,097 *	(1,989 , 2,211)
DES-FP	1,297 *	(1,249 , 1,348)	1,328 *	(1,276 , 1,382)	1,362 *	(1,303 , 1,423)	0,624 *	(0,597 , 0,653)
DES-FGA	1,089 *	(1,035 , 1,145)	1,058 *	(1,005 , 1,115)	0,965	(0,910 , 1,023)	0,872 *	(0,829 , 0,918)
DES-DEP	1,303 *	(1,212 , 1,401)	1,390 *	(1,290 , 1,496)	1,459 *	(1,345 , 1,583)	0,897 *	(0,834 , 0,966)
DES-HE	0,660 *	(0,635 , 0,687)	0,650 *	(0,624 , 0,676)	0,658 *	(0,629 , 0,689)	1,645 *	(1,585 , 1,706)
DEP-FGJ	1,000 *	(1,000 , 1,000)	1,318 *	(1,148 , 1,513)	1,591 *	(1,393 , 1,817)	0,849 *	(0,737 , 0,978)
DEP-FP	1,103	(0,963 , 1,264)	1,048	(0,917 , 1,199)	1,085	(0,948 , 1,241)	0,725 *	(0,665 , 0,791)
DEP-DES	1,281 *	(1,094 , 1,499)	1,185 *	(1,013 , 1,385)	1,143	(0,977 , 1,338)	0,967	(0,899 , 1,040)
DEP-HE	0,662 *	(0,572 , 0,767)	0,663 *	(0,575 , 0,764)	0,591 *	(0,511 , 0,683)	1,955 *	(1,796 , 2,129)

Tableau B4 – Effets de cohorte

	C2004		C2005		C2006		C2007		C2008	
FGJ-FP	0,894 *	(0,832, 0,961)	0,925 *	(0,861, 0,994)	0,871 *	(0,810, 0,937)	0,884 *	(0,823, 0,950)	0,914 *	(0,852, 0,981)
FGJ-FGA	0,951 *	(0,915, 0,989)	0,941 *	(0,905, 0,978)	0,938 *	(0,901, 0,975)	0,939 *	(0,903, 0,976)	0,968	(0,931, 1,005)
FGJ-DES	1,065 *	(1,052, 1,079)	1,109 *	(1,095, 1,124)	1,142 *	(1,128, 1,158)	1,137 *	(1,123, 1,152)	1,121 *	(1,107, 1,136)
FGJ-DEP	0,990	(0,940, 1,042)	0,916 *	(0,869, 0,965)	0,912 *	(0,866, 0,961)	0,884 *	(0,839, 0,932)	0,835 *	(0,792, 0,880)
FGJ-HE	1,020	(0,849, 1,225)	0,881	(0,725, 1,071)	0,917	(0,753, 1,117)	0,554 *	(0,442, 0,694)	0,358 *	(0,276, 0,464)
FP-FGA	0,875	(0,736, 1,040)	0,968	(0,818, 1,146)	0,843	(0,706, 1,006)	0,895	(0,754, 1,064)	0,768 *	(0,644, 0,916)
FP-DES	1,091	(0,944, 1,261)	0,997	(0,859, 1,158)	0,904	(0,775, 1,056)	0,821 *	(0,702, 0,961)	0,878	(0,753, 1,022)
FP-DEP	0,975	(0,924, 1,028)	0,955	(0,904, 1,008)	0,906 *	(0,857, 0,957)	0,965	(0,914, 1,019)	1,056 *	(1,002, 1,113)
FP-HE	1,105	(0,991, 1,232)	1,077	(0,964, 1,202)	1,209 *	(1,085, 1,347)	1,077	(0,966, 1,202)	0,642 *	(0,570, 0,723)
FGA-FP	0,994	(0,884, 1,118)	0,936	(0,829, 1,057)	0,912	(0,806, 1,032)	0,848 *	(0,748, 0,961)	0,764 *	(0,671, 0,869)
FGA-DES	0,942	(0,850, 1,044)	0,789 *	(0,707, 0,880)	0,682 *	(0,607, 0,766)	0,649 *	(0,578, 0,730)	0,737 *	(0,659, 0,824)
FGA-DEP	0,961	(0,875, 1,057)	0,959	(0,871, 1,056)	0,930	(0,842, 1,026)	1,035	(0,941, 1,139)	1,146 *	(1,044, 1,257)
FGA-HE	1,031	(0,955, 1,113)	1,034	(0,956, 1,118)	1,050	(0,969, 1,138)	0,915 *	(0,843, 0,993)	0,761 *	(0,700, 0,828)
DES-FP	0,921 *	(0,879, 0,964)	0,933 *	(0,890, 0,977)	0,994	(0,949, 1,042)	1,095 *	(1,045, 1,148)	1,237 *	(1,180, 1,296)
DES-FGA	0,925 *	(0,874, 0,979)	0,858 *	(0,808, 0,910)	0,816 *	(0,768, 0,868)	0,798 *	(0,750, 0,850)	0,816 *	(0,765, 0,871)
DES-DEP	0,935	(0,861, 1,014)	0,874 *	(0,802, 0,952)	0,990	(0,910, 1,077)	1,019	(0,934, 1,111)	1,116 *	(1,024, 1,218)
DES-HE	1,175 *	(1,123, 1,230)	1,265 *	(1,205, 1,327)	1,166 *	(1,109, 1,226)	1,031	(0,979, 1,086)	0,750 *	(0,707, 0,797)
DEP-FGJ	1,017	(0,858, 1,205)	1,044	(0,878, 1,242)	1,270 *	(1,074, 1,502)	1,281 *	(1,084, 1,513)	1,441 *	(1,222, 1,700)
DEP-FP	1,009	(0,917, 1,111)	1,060	(0,961, 1,169)	1,051	(0,952, 1,161)	1,069	(0,970, 1,179)	1,055	(0,956, 1,165)
DEP-DES	0,998	(0,920, 1,083)	1,040	(0,957, 1,129)	0,990	(0,909, 1,078)	0,974	(0,895, 1,060)	0,945	(0,867, 1,029)
DEP-HE	1,058	(0,937, 1,194)	0,931	(0,817, 1,061)	0,957	(0,839, 1,093)	0,959	(0,844, 1,089)	0,877 *	(0,772, 0,996)