

## Cela vaut-il la peine de faire une prépa ?

### Une réponse à partir de l'insertion des ingénieurs

#### **Noël Adangnikou**

Institut de recherche sur l'éducation (IREDU), CNRS-Université de Bourgogne,  
Pôle AAFE, Esplanade Erasme, BP 26513, 21065 Dijon cedex.  
E-mail : noel.adangnikou@u-bourgogne.fr  
Web : www.u-bourgogne.fr/IREDU.

#### **Jean-Jacques Paul**

Institut de recherche sur l'éducation (IREDU), CNRS-Université de Bourgogne,  
Pôle AAFE, Esplanade Erasme, BP 26513, 21065 Dijon cedex.  
E-mail : jean-jacques.paul@u-bourgogne.fr  
Web : www.u-bourgogne.fr/IREDU.

Document réalisé avec la collaboration de Bertrand Hounkannounon.

Cette recherche a été réalisée dans le cadre du  
« Groupe de travail sur l'enseignement supérieur » du Céreq.  
<http://matisse.univ-paris1.fr/gtes>.

Elle a bénéficié d'une subvention du Commissariat général du Plan dans le cadre de l'appel d'offres « Évaluation du système d'éducation et de formation ».

Ce document est présenté sur le site du Céreq afin de favoriser la diffusion et la discussion de résultats de travaux d'études et de recherches. Il propose un état d'avancement provisoire d'une réflexion pouvant déboucher sur une publication. Les hypothèses et points de vue qu'il expose, de même que sa présentation et son titre, n'engagent pas le Céreq et sont de la responsabilité des auteurs.

**Juin 2004**

## SOMMAIRE

<b>1. LE RECRUTEMENT DES ÉCOLES D'INGÉNIEURS EN FRANCE. ....</b>	<b>3</b>
1.1. Un recrutement différencié selon la formation bac+2 d'origine.....	3
1.2. Hiérarchie des écoles et modes de recrutement.....	6
<b>2. DESCRIPTION ET ANALYSE DES INGÉNIEURS ISSUS DE LA BASE</b>	
<b>GÉNÉRATION 98 DU CÉREQ .....</b>	<b>7</b>
2.1. Description rapide de la population d'ingénieurs étudiée	7
2.2. Analyse de la relation entre la formation bac+2 initiale et les conditions d'insertion des ingénieurs .....	10
<b>3. L'EFFET DES ÉCOLES SUR LES SALAIRES DES INGÉNIEURS.....</b>	<b>11</b>
3.1. La nécessité du recours à une analyse multiniveau .....	11
3.2. Les résultats de l'analyse multiniveau .....	12
<b>4. LA QUESTION DE LA COMPARABILITÉ ENTRE LES ÉTUDIANTS PASSÉS</b>	
<b>PAR UNE CLASSE PRÉPARATOIRE ET LES AUTRES .....</b>	<b>14</b>
4.1. La problématique de l'évaluation de l'impact d'un "traitement" appliquée aux classes préparatoires .....	14
4.2. Mise en œuvre de la méthode d'appariement par probabilité prédite ou <i>propensity</i> <i>score matching</i> .....	17
<b>5. LA RECHERCHE DE DIFFÉRENCES DE SALAIRES DUES À DES</b>	
<b>CARACTÉRISTIQUES INOBSERVABLES .....</b>	<b>23</b>
<b>CONCLUSION.....</b>	<b>27</b>
<b>RÉFÉRENCES BIBLIOGRAPHIQUES.....</b>	<b>28</b>
<b>ANNEXES.....</b>	<b>29</b>

La classe préparatoire aux grandes écoles (CPGE), communément appelée classe prépa ou prépa, est d'abord liée à l'histoire des écoles d'ingénieurs. L'École des Ponts et Chaussées, créée en 1747 sous l'Ancien Régime et considérée comme la première d'entre elles, présente d'emblée l'une des principales caractéristiques des grandes écoles, à savoir un nombre de places très inférieur aux candidatures ; cela va conférer aux préparations, futures classes préparatoires, une importance décisive<sup>1</sup>. Au fil du temps, la prépa s'est imposée comme la voie d'excellence pour intégrer les écoles d'ingénieurs dont le prestige, dans une société marquée par l'innovation et le développement technologique, n'a guère été désavoué depuis lors. Si la prépa constitue la voie la plus prestigieuse pour accéder aux grandes écoles, elle est aussi la plus coûteuse. Au niveau de l'État, la dépense annuelle moyenne pour un élève de prépa s'élève à 13 220 € contre 6 850 € pour un étudiant d'université quel que soit son cycle d'étude<sup>2</sup>. Au niveau individuel, la concentration exclusive sur des objectifs scolaires est régulièrement mise en avant et questionnée. Ainsi selon Bourdieu<sup>3</sup>, "Ce qui distingue les classes préparatoires de toutes les autres institutions d'enseignement supérieur, c'est avant tout le système des moyens institutionnels, incitations, contraintes et contrôles, qui concourent à réduire toute l'existence de ceux qu'on appelle encore ici des "élèves" (par opposition aux "étudiants") à une succession ininterrompue d'activités scolaires intensives, rigoureusement réglées et contrôlées tant dans leur moment que dans leur rythme." Pour certains, comme Cier<sup>4</sup>, le système de la classe prépa conduit à un résultat bénéfique à plus ou moins long terme et plus particulièrement sur le plan professionnel. Il rapporte que "de nombreux ex-préparationnaires considèrent que de par la diversité des matières abordées et le sérieux avec lequel, concours oblige, elles ont été étudiées, cette période de leur vie a été très formatrice". Dans une perspective communément admise dans l'opinion publique, il met lui aussi en avant l'acquisition forcée d'un grand nombre de connaissances, dans un temps très limité, qui obligerait au maniement de méthodes de travail efficaces. "Celles-ci constitueront un plus durant toute la vie professionnelle. Les entreprises recherchent souvent ces ex-préparationnaires ayant su développer ces solides capacités de travail, d'organisation et d'adaptation qui caractérisent la classe préparatoire". A partir de l'enquête de l'Observatoire de la vie étudiante (OVE), Lahire<sup>5</sup> confirme cette attitude de concentration sur le travail scolaire des CPGE de loin supérieure aux autres formations de même niveau (excepté pour les étudiants de médecine). Cependant, d'autres comme Danin-Boileau<sup>6</sup>, attirent l'attention sur le préjudice en terme d'estime de soi et de développement de la personnalité pouvant résulter du couplage entre d'une part une situation "d'ascétisme, de frustrations massives et d'hyperintellectualisation" à laquelle les élèves de prépa sont soumis et d'autre part le fait que bon nombre d'entre eux soient placés dans une situation de difficultés scolaires liées à des résultats irréguliers en courbe descendante. Qu'il s'agisse de l'État ou des élèves, le "jeu" en vaut-il la "chandelle" ? Dans la même perspective que Duru et Kieffer<sup>7</sup>, la question se pose de

---

<sup>1</sup> Yves Gaupeau, *La France à l'école*, Découverte Gallimard, n°147, 1997. Pour une présentation détaillée sur l'histoire des CPGE et la place qu'elles occupent aujourd'hui dans l'enseignement supérieur, le lecteur pourra se reporter à la partie introductive du colloque "Démocratie, classes préparatoires et grandes écoles" des 16 et 17 mai 2003. Actes disponibles sur internet à l'adresse suivante :

<http://www.int-evry.fr/LIESSE/Democratie/democratie-VersionFinale.htm>.

<sup>2</sup> Direction de l'Évaluation et de la Prospective, "Le coût de l'éducation en 2002", *Note d'information* 03.57, novembre 2003.

<sup>3</sup> Pierre Bourdieu, *La Noblesse d'État*, Éditions de Minuit, 1988.

<sup>4</sup> Bernard Cier, "Les admissions parallèles dans les écoles de commerce", Espace Études Éditions, 1999.

<sup>5</sup> Bernard Lahire, Les manières d'étudier, *Cahier de l'Observatoire de la Vie Étudiante* n°2, La Documentation Française. 1997.

<sup>6</sup> Henri Danin-Boileau, "Ruptures psychiques en classes préparatoires" in *Classes préparatoires, des étudiants pas comme les autres*, sous la direction de Claire-Marine François-Poncet et d'Alain Braconnier, Bayard Éditions-Fondation de France, 1998.

<sup>7</sup> En utilisant les enquêtes du CNISF (Conseil National des Ingénieurs et Scientifiques de France), Marie Duru et Annick Kieffer s'interrogent sur l'existence d'un effet prépa dans la carrière d'ingénieur. Duru-Bellat Marie,

savoir si cette différence de coûts, quelle que soit leur nature, conduit à une différence de gains.

Une première manière simple de répondre à cette question peut consister à s'intéresser aux salaires des ingénieurs selon qu'ils sont passés ou non par une prépa. Comme le rappelle le magazine *L'Étudiant*<sup>8</sup>, à partir des données de la Direction de la Programmation et du développement (DPD), plus de la moitié des effectifs des écoles d'ingénieurs ne proviennent pas des classes préparatoires. Il faut désormais compter avec les autres formations bac+2 que sont le DEUG, le DUT et le BTS<sup>9</sup>. Dès lors, il devient possible de regarder si, en moyenne, pour une même école d'ingénieurs, les diplômés préalablement issus de prépa gagnent significativement plus que leurs homologues issus de DEUG, DUT ou BTS. Les données mobilisées pour aborder cette question sont celles de l'enquête Génération 98 du Céreq (Centre d'études et de recherches sur les qualifications), réalisée en 2001 auprès de sortants du système éducatif en 1998, soit trois ans après la fin des études initiales.

La première partie de cette étude portera un regard sur la structure de recrutement des écoles d'ingénieurs en France. Connaissant l'existence d'un fort "effet école" sur les salaires des ingénieurs débutants<sup>10</sup>, il s'agira de voir si les élèves issus de prépa et ceux issus des autres formations bac+2 accèdent aux mêmes écoles. La seconde partie visera à observer, de manière descriptive puis analytique, si, par rapport à la base de données étudiée, des différences ressortent en termes de salaires entre les diplômés ingénieurs issus de prépas et ceux issus d'autres formations bac+2.

## **1. LE RECRUTEMENT DES ÉCOLES D'INGÉNIEURS EN FRANCE.**

### **1.1. Un recrutement différencié selon la formation bac+2 d'origine**

En 2000, la Direction de la Programmation et du Développement (DPD) recense 237 écoles d'ingénieurs accueillant 89 313 élèves. Est considérée comme école d'ingénieurs, "*tout établissement public ou privé, quel que soit le ministère de rattachement, habilité à délivrer un titre d'ingénieur diplômé par le Ministre de l'éducation nationale et par la Commission des Titres d'ingénieur (loi du 10 juillet 1934)*"<sup>11</sup>. Usuellement, la DPD opère une distinction des écoles selon le ministère de rattachement conduisant à une présentation en trois grandes catégories : les écoles publiques sous la tutelle du ministère de l'Éducation nationale (128 établissements), les écoles publiques sous la tutelle d'un autre ministère (45 établissements) et les écoles privées (64 établissements). Le parcours menant au diplôme d'ingénieur se réalise le plus souvent, et classiquement, par un accès à l'école au niveau bac+3 (après une formation bac+2) suivi de trois années d'étude au sein de l'école, ce qui est dénommé officiellement cycle ingénieur. C'est pourquoi dans le tableau suivant (Tableau 1) sont présentés les nouveaux inscrits dans ce cycle (en ne retenant pas les inscriptions en classes préparatoires intégrées ou en année de spécialisation) par origine scolaire.

---

Kieffer Annick, Adangnikou Noël, "Efficacité et équité de l'enseignement supérieur français dans la production des élites", Premier colloque scientifique du Réseau d'étude sur l'enseignement supérieur (RESUP), mai 2002.

<sup>8</sup> L'Étudiant, "Ingénieurs, le vrai recrutement des écoles", *La lettre de l'Étudiant*, n° 562, novembre 2000.

<sup>9</sup> Diplôme universitaire d'études générales (DEUG), diplôme universitaire de technologie (DUT), brevet de technicien supérieur (BTS).

<sup>10</sup> Loïc Le Pellec et Sébastien Roux, "Les salaires des ingénieurs diplômés en 2000", *Insee première*, Insee, n°812, novembre 2001.

<sup>11</sup> Source DPD, Tableaux statistiques 6787, juin, 2001.

Il apparaît que les écoles dépendant du ministère de l'Éducation nationale rassemblent la part la plus importante des nouveaux inscrits (presque 60 %). Les écoles dépendant d'un autre ministère en regroupent presque 15 % alors que les écoles privées en rassemblent 26 %. Les écoles publiques représentent donc 75 % des effectifs. Il est à noter la place considérable occupée aujourd'hui par les écoles universitaires qui représentent plus de 38 % de l'ensemble des nouveaux inscrits des écoles.

Les élèves issus de CPGE constituent presque 58 % des nouveaux inscrits en cycle ingénieur. Cela confirme la place prééminente que joue la classe préparatoire dans le cursus d'ingénieur. Le second groupe le plus important est constitué par les classes préparatoires intégrées (21,3 %). Ces derniers intègrent essentiellement trois types d'écoles : les INSA (instituts nationaux de sciences appliquées), les ENI (écoles nationales d'ingénieurs) et les écoles privées. Les étudiants issus d'une autre formation qu'une prépa représentent un peu plus de 20 % des nouveaux inscrits. Parmi eux, les élèves issus de DEUG sont un peu plus de 7 % et sont moins nombreux que leurs homologues universitaires venant de DUT (11 %). Les premiers comme les seconds se trouvent essentiellement dans les écoles dépendant du ministère de l'Éducation nationale et plus particulièrement dans les écoles universitaires. Enfin les élèves issus de BTS sont minoritaires (3 %). Ils intègrent plus particulièrement deux types d'écoles, celles dépendant du ministère de l'Éducation nationale et les écoles privées.

Tableau 1  
Nouveaux inscrits en cycle ingénieur dans les écoles d'ingénieurs en 2000/2001

		Nombre de nouveaux inscrits en cycle ingénieur 2000/2001					
		CPI	CPGE	BTS	DUT	DEUG	Ensemble
<b>Écoles d'ingénieurs dépendant du MEN</b>	Écoles universitaires (95)	1 245	5 274	231	1 631	1 294	9 675
	ENI (4)	229	39	41	161	10	480
	INSA (4)	1 101	110	12	172	105	1 500
	ENSAM (12)	47	1 160	68	125	1	1 401
	Autres écoles (13)	43	1 592	47	15	17	1 714
	<b>Sous Total (128)</b>	<b>2 665</b>	<b>8 175</b>	<b>399</b>	<b>2 104</b>	<b>1427</b>	<b>14 770</b>
	<b>%</b>	<b>50,4 %</b>	<b>57,0 %</b>	<b>53,6 %</b>	<b>76,7 %</b>	<b>80,8 %</b>	<b>59,4 %</b>
<b>Écoles d'ingénieurs dépendant d'autres ministères</b>	Agriculture (15)	50	853	13	17	96	1 029
	Ville de Paris (2)		86			4	90
	Industrie (8)	348	255		10	15	628
	Équipement transport (5)		358		1	13	372
	Santé (1)						0
	Télécommunications (3)		424	4			428
	Défense (11)		1 060		1	16	1 077
	<b>Sous Total (45)</b>	<b>398</b>	<b>3 036</b>	<b>17</b>	<b>29</b>	<b>144</b>	<b>3 624</b>
<b>%</b>	<b>7,5 %</b>	<b>21,2 %</b>	<b>2,3 %</b>	<b>1,1 %</b>	<b>8,2 %</b>	<b>14,6 %</b>	
<b>Écoles privées (64)</b>		<b>2 226</b>	<b>3 127</b>	<b>329</b>	<b>609</b>	<b>195</b>	<b>6 486</b>
<b>%</b>		<b>42,1 %</b>	<b>21,8 %</b>	<b>44,2 %</b>	<b>22,2 %</b>	<b>11,0 %</b>	<b>26,1 %</b>
<b>Total écoles d'ingénieurs (237)</b>		<b>5 289</b>	<b>14 338</b>	<b>745</b>	<b>2 742</b>	<b>1 766</b>	<b>24 880</b>
<b>%</b>		<b>100 %</b>	<b>100 %</b>	<b>100 %</b>	<b>100 %</b>	<b>100 %</b>	<b>100 %</b>
<b>Part de chaque formation dans l'ensemble</b>		<b>21,3 %</b>	<b>57,6 %</b>	<b>3,0 %</b>	<b>11 %</b>	<b>7,1 %</b>	<b>100 %</b>

Note : • Les chiffres indiquant le nombre d'élèves venant des classes préparatoires intégrées constituent une approximation. En l'état, les chiffres des Tableaux Statistiques ne permettent pas de déterminer parmi les nouveaux inscrits en cycle ingénieur ceux qui viennent des classes préparatoires intégrées. Le nombre de ces derniers a donc été approximé en considérant les effectifs de la première année du cycle ingénieur auxquels ont été soustraits les redoublants ainsi que les nouveaux inscrits qui a priori n'intégraient pas la première année du cycle ingénieur (il s'agit des individus issus de maîtrise, d'ingénieurs diplômés ou d'individus classés dans la rubrique "autre"). Enfin les individus non déclarés ont été pris en considération au prorata des différents flux d'entrée.

• Parmi les CPGE figurent les élèves issus de classes préparatoires ATS (adaptation pour les techniciens supérieurs) destinées aux élèves issus de BTS et de DUT. Leur nombre est assez limité (voir Annexe 1).

Sources : Direction de la Programmation et du Développement, *Tableaux Statistiques*, n°6787, "Les écoles d'ingénieurs, effectifs des élèves en 2000-2001", juin 2001.

Un regard sur les pourcentages d'élèves de classes préparatoires parmi les nouveaux inscrits en cycle ingénieur révèle des variations selon le statut des écoles (Tableau 2). Ils sont hégémoniques dans les écoles dépendant d'autres ministères que celui de l'Éducation nationale (où ils représentent 94,8 % des inscrits). Et ils sont encore fortement présents dans les écoles dépendant du ministère de l'Éducation nationale (73,4 %). Mais cette moyenne recouvre de fortes disparités puisque s'ils constituent plus de 95 % dans les "autres écoles", qui comprennent notamment les Écoles Centrales et quelques instituts spécialisés, ils ne sont en moyenne que 67 % dans le principal groupe de cet ensemble que sont les écoles universitaires.

Tableau 2

**Proportion d'élèves issus de classes préparatoires parmi les nouveaux inscrits en écoles d'ingénieurs.**

Pourcentage d'élèves issus de classes préparatoires parmi les nouveaux inscrits		(I)	(II)
Écoles d'ingénieurs dépendant du MEN	Écoles universitaires (95)	54,5 %	67,4 %
	ENI (4)	8,1 %	55,8 %
	INSA (4)	7,3 %	80,7 %
	ENSAM (12)	82,8 %	86,2 %
	Autres écoles (13)	92,9 %	95,4 %
	<b>Sous Total (128)</b>	<b>55,3 %</b>	<b>73,4 %</b>
Écoles d'ingénieurs dépendant d'autres ministères	Agriculture (15)	82,9 %	87,8 %
	Ville de Paris (2)	95,6 %	95,6 %
	Industrie (8)	40,6 %	96,0 %
	Équipement transport (5)	96,2 %	96,2 %
	Santé (1)	0,0 %	0,0 %
	Télécommunications (3)	99,1 %	99,1 %
	Défense (11)	98,4 %	98,4 %
	<b>Sous Total (45)</b>	<b>83,8 %</b>	<b>94,8 %</b>
Écoles privées (64)		<b>48,2 %</b>	<b>82,5 %</b>
<b>Total écoles d'ingénieurs (237)</b>		<b>57,6 %</b>	<b>78,9 %</b>

Note : • La colonne (I) indique le pourcentage de CPGE par rapport à l'ensemble des inscrits en cycle ingénieur (voir Tableau 1).

• La colonne (II) indique le pourcentage de CPGE et de CPI (classes préparatoires intégrées) par rapport à l'ensemble des inscrits en cycle ingénieur.

Sources : Direction de la Programmation et du Développement, *Tableaux Statistiques*, n°6787, "Les écoles d'ingénieurs, effectifs des élèves en 2000-2001", juin 2001.

Cette distribution des formations bac+2 observée dans le cycle ingénieur des écoles s'inverse quand on considère l'ensemble des effectifs en formation scientifique bac+2 en France (Tableau 3). Ainsi les premiers cycles universitaires qui ne représentent que 7 % des effectifs en écoles constituent 37,6 % des effectifs de l'ensemble des formations bac+2. La part des STS s'élève à 26,2 % des formations bac+2 au niveau national alors que leurs élèves ne constituent pas plus de 3 % des effectifs d'écoles. Les DUT représentent 20 % des formations bac+2 au niveau national et leurs élèves constituent 11 % des effectifs d'écoles. Enfin les CPGE qui représentent 57,6 % des effectifs d'écoles ne constituent que 14,2 % des effectifs de l'ensemble des formations bac+2 et les CPI seulement 1,9 % contre 21,3 % en écoles.

Tableau 3

**Effectifs des formations scientifiques bac+2 en France.**

Effectifs des formations scientifiques bac+2 en France							
2001		BTS	DUT	DEUG	CPI	CPGE	TOTAL
	Effectifs	80 730	63 228	115 980	5 855	43 437	309 230
	%	26,1 %	20,4 %	37,5 %	1,9 %	14,0 %	100,0 %

Sources : Direction de la Programmation et du Développement (voir le mode de calcul en Annexe 2), *Repères et Références Statistiques*, édition 2001 ; *Tableaux Statistiques*, n°6787, "Les écoles d'ingénieurs, effectifs des élèves en 2000-2001", juin 2001 ; *Tableaux Statistiques*, n°6746, "Classes préparatoires aux grandes écoles 2000-2001", décembre 2000 ; *Tableaux Statistiques*, n°6784, "Effectifs dans les instituts universitaires de technologie 2000-2001", mai 2001 ; *Tableaux Statistiques*, n°6749, "Sections de techniciens supérieurs, préparations diverses post-baccalauréat 2000-2001", janvier 2001.

En résumé, on constate une certaine diversification de l'accès en cycle ingénieur, un élève sur cinq n'étant pas originaire d'une classe prépa, ce phénomène étant plus accentué pour les écoles universitaires où c'est le cas d'un élève sur trois. Dans l'ensemble, il apparaît donc que la classe préparatoire permet d'intégrer toutes les écoles quel que soit leur type alors que les autres formations bac+2 permettent plutôt l'accès aux écoles dépendant du ministère de l'Éducation nationale et plus particulièrement aux écoles universitaires. La principale question que nous allons traiter maintenant est de savoir si la formation suivie avant d'entrer en école d'ingénieur influencera le niveau de rémunération après que les étudiants auront obtenu leur diplôme d'ingénieur.

## 1.2. Hiérarchie des écoles et modes de recrutement

Dans leur étude, Le Pellec et Roux<sup>12</sup> hiérarchisent les écoles eu égard au salaire de leurs diplômés et fournissent en particulier un classement des 19 écoles les plus rémunératrices. Cette classification contient les écoles les plus prestigieuses telles que Polytechnique, les "grandes" Mines, Ponts et Chaussées et les Ecoles Centrales. Parmi ces dix-neuf, dix opèrent un recrutement exclusif à base d'élèves issus des classes préparatoires<sup>13</sup> (les sept premières d'entre elles sont aussi les sept écoles les plus rémunératrices) (Tableau 4). Les autres écoles sont à recrutement mixte comme le sont aujourd'hui la quasi-totalité des écoles d'ingénieurs.

Tableau 4  
Part des écoles recrutant quasi exclusivement des prépa au niveau bac+3.

	Écoles	Effectifs	Pourcentages
1	Polytechnique	2 054	8 %
2	Mines de Paris		
3	Centrale Paris		
4	Ponts et chaussées		
5	École nationale supérieure des techniques avancées ENSTA Paris		
6	Mines de Saint-Étienne		
7	Sup Aéro Toulouse		
8	École catholique des arts et métiers - ECAM Lyon		
9	Centrale Lille		
10	Centrale Lyon		

Lecture : les écoles sont présentées dans l'ordre de rémunération.

Ces 10 écoles rassemblent 2 054 élèves au niveau bac+3 lesquels représentent 8 % des élèves de même niveau de l'ensemble des écoles en France (soit 26 824 élèves).

Au bout du compte, ces dix écoles permettent à leurs élèves issus de classes préparatoires de devenir les ingénieurs débutants les mieux payés de France. Pour ces élèves, faire une prépa fut un "jeu" qui en valut vraiment la "chandelle". Mais quid des 17 570 autres prépas (soit 89 %) qui se sont vus côtoyer sur les bancs de l'école d'ingénieurs, des élèves issus de DEUG, de DUT et de BTS...?

<sup>12</sup> Loïc Le Pellec et Sébastien Roux, "Les salaires des ingénieurs diplômés en 2000", *Insee Première*, n°812, novembre 2001.

<sup>13</sup> Le type de recrutement des écoles a été déterminé à partir du guide élaboré par l'ONISEP (Office national d'information sur les enseignements et les professions) en 2000 (voir Annexe 4). ONISEP, "Ecoles d'ingénieurs, niveaux d'admission", in *Après le bac... réussir ses études*, 2000.

## 2. DESCRIPTION ET ANALYSE DES INGÉNIEURS ISSUS DE LA BASE GÉNÉRATION 98 DU CÉREQ

### 2.1. Description rapide de la population d'ingénieurs étudiée

La base Génération 98 du Céreq permet d'appréhender la scolarité et l'insertion professionnelle de 1098 diplômés, issus directement du cycle ingénieur. Ces derniers ont été enquêtés en 2001 soit 3 ans après leur sortie. L'objectif est ici d'expliquer le salaire en fonction du parcours avant l'entrée en école d'ingénieurs. Sur le plan technique, il est très courant face à un tel objectif d'effectuer les analyses à partir d'un modèle multivarié. Bien que cela conduise à un certain nombre d'obstacles, comme nous le verrons par la suite, cette perspective a constitué notre point de départ avec pour conséquence immédiate la délimitation de l'échantillon utilisable pour notre étude. Ainsi, sur les 1098 diplômés d'une école d'ingénieurs, le parcours initial est renseigné pour 1084 individus. De plus, les ingénieurs ayant suivi leur formation en apprentissage ont été retirés en raison de l'effet spécifique de leur expérience professionnelle antérieure à l'obtention du diplôme d'ingénieur. Notre base de départ est donc constituée de 972 individus.

Cet échantillon fait apparaître la nette prédominance des diplômés issus de prépa (53,8 %). Viennent ensuite les diplômés issus de DEUG (14,7 %) et de DUT (14,3 %) puis de classes préparatoires intégrées (12 %) et enfin les diplômés issus de BTS (5,3 %). En terme d'origine l'échantillon ne diffère pas sensiblement de la population mère même si quelques différences sont à observer. La comparaison des effectifs du présent échantillon avec les données nationales révèle une moindre représentation des CPGE dans l'échantillon (ils constituent 58,7 % des effectifs au niveau national) (Tableau 5). Les DEUG et les DUT apparaissent plus présents dans l'échantillon qu'au niveau national (14,7 % contre 10,2 % pour les DEUG et 14,3 % contre 11,8 % pour les DUT) alors que le contraire s'observe pour les diplômés issus des classes préparatoires intégrées (11,8 % contre 15,4 %). Enfin la représentation des BTS est aussi supérieure dans l'échantillon (5,3 % contre 3,8 % au niveau national).

Tableau 5

Comparaison de l'échantillon extrait de la base Génération 98 avec l'ensemble des écoles d'ingénieurs.

Formations	BTS	DUT	DEUG	CPI	CPGE	Total
Effectifs de l'échantillon Céreq	52	139	143	115	523	972
Structure en pourcentage de l'échantillon Céreq	5,3 %	14,3 %	14,7 %	11,8 %	53,8 %	100 %
Effectifs pour l'ensemble des écoles *	746	2 297	2 000	3 018	11 480	19 541
Structure en pourcentage pour l'ensemble des écoles*	3,8 %	11,8 %	10,2 %	15,4 %	58,7 %	100 %

Note : CPI : classes préparatoires intégrées.

\*Comparaison avec les nouveaux inscrits en 1995/1996 et diplômés en 1998. Sources : Direction de l'évaluation et de la prospective (DEP), tableaux statistiques n°6436, 1995/1996

Les ingénieurs de l'échantillon proviennent de 118 écoles d'ingénieurs. En termes d'effectifs, les ingénieurs issus d'écoles publiques dépendant du ministère de l'Éducation nationale représentent 70,9 % des effectifs (contre 58 % au niveau national). Les ingénieurs issus d'écoles publiques dépendant d'autres ministères représentent eux 9,3 % des effectifs (contre 19,9 % au niveau national) et ceux issus d'écoles privées 19,1 % (contre 25,1 % au niveau national) (voir Tableau 6).



*Tableau 6*  
**Effectifs des écoles dans l'échantillon et au niveau national.**

STATUT	ÉDUCATION	AUTRES MINISTÈRES	PRIVE	ENSEMBLE
<b>ÉCHANTILLON</b>	70,9 %	9,3 %	19,8 %	100 % (n=972)
<b>NATIONAL</b>	58 %	16,9 %	25,1 %	100 % (n=31 048)

Source niveau national : DPD, *Tableaux Statistiques*, n°6787, "les effectifs des écoles d'ingénieurs en 2000/2001", juin 2001.

Si l'on se focalise sur les ingénieurs issus des écoles universitaires ou formations universitaires, qui ont le recrutement le plus mixte, ils apparaissent surreprésentés dans l'échantillon où ils constituent 53,4 % des effectifs contre 36,9 % au niveau national (voir Annexe 5).

La sous représentation des élèves issus de classe prépa repérée globalement se retrouve au niveau des écoles dépendant du ministère de l'Éducation nationale comme des autres ministères. Par contre la part des prépas dans les écoles privées est légèrement supérieure par rapport au niveau national (74 % contre 71,6 %) (tableau ci-dessous).

*Tableau 7*  
**Part des classes préparatoires dans l'échantillon et au niveau national.**

STATUT	ENSEMBLE ÉDUCATION	AUTRES MINISTÈRES	PRIVE	ENSEMBLE
<b>ÉCHANTILLON</b>	62 %	75,8 %	74 %	65,6 %
<b>NATIONAL</b>	67 %	94 %	71,6 %	72,9 %

Source niveau national : DPD, *Tableaux Statistiques*, n°6787, "les effectifs des écoles d'ingénieurs en 2000/2001", juin 2001

Concernant les caractéristiques sociodémographiques, les hommes constituent les trois quarts de la population (75,4 %). Les femmes sont moins présentes au niveau des formations professionnelles, BTS et DUT (respectivement 15,4 et 17,3 %). Elles sont plus particulièrement représentées au niveau des classes préparatoires intégrées (34,8 %) et des DEUG (28,7 %). Viennent ensuite les CPGE où leur part représente 24,1 % des effectifs.

Les données confirment le recrutement socialement biaisé des classes préparatoires, puisque 54 % de leurs étudiants ont un père cadre, ingénieur, profession libérale, professeur. Cette proportion est nettement plus importante que chez les jeunes issus de STS et de DEUG. On pourra remarquer au passage que la part des enfants de cette catégorie, plus importante chez les jeunes titulaires de DUT que chez les jeunes titulaires de DEUG, renforce l'idée que les IUT constituent un élément de la stratégie scolaire des catégories les plus informées sur le fonctionnement du système.

Plus de 63 % des diplômés ont eu leur bac "à l'heure". Ce sont plutôt les CPGE qui l'ont eu avec un an d'avance (28 %) et les BTS avec un an de retard (30 %). Plus d'un tiers (32 %) des ingénieurs ont été diplômés dans le domaine des technologies industrielles. Les autres spécialités oscillent entre 12 % en mécanique et 22,1 % en informatique et électronique.

Tableau 8

## Caractéristiques sociodémographiques et scolaires des diplômés ingénieurs de l'échantillon (en %).

FORMATION INITIALE	BTS	DUT	DEUG	CPI	CPGE	Ensemble
<b>Sexe</b>						
Femme	15,4	17,3	28,7	34,8	24,1	24,6
<b>Catégorie socioprofessionnelle</b>						
Agriculteur	10,6	5,9	4,6	11,9	5,8	6,6
Artisan, commerçant, chef d'entreprise	12,8	14,1	12,4	8,3	7,8	9,6
Cadre, ingénieur, profession libérale, professeur	29,8	43,0	36,4	55,0	53,3	48,4
Technicien, agent de maîtrise, VRP, profession intermédiaire	19,2	14,1	10,9	11,9	12,5	12,8
Employé	17,0	11,1	23,3	10,1	14,7	15,0
Ouvrier	10,6	11,1	10,9	2,8	4,2	6,3
NSP	0	0,7	1,5	0	1,7	1,3
<b>Ensemble</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Age au bac</b>						
17 ans	9,6	15,8	19,6	20,9	28,3	23,3
18 ans	53,8	61,9	68,5	66,1	63,1	63,6
19 ans	30,8	15,8	9,1	12,1	7,5	10,7
20 ans	5,8	6,5	2,8	0,9	1,1	2,4
<b>Ensemble</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>
<b>Spécialités de formation</b>						
Technologies industrielles	40	31,6	31,4	16,2	35,8	32,4
Agriculture, agroalimentaire et connexe	20	9,8	19,0	32,5	17,3	18,4
Transformations matériaux et énergie	4,4	15,8	21,9	7,2	15,5	15,0
Mécanique	11,1	12,0	10,2	21,6	10,7	12,1
Informatique, électronique	24,5	30,8	17,5	22,5	20,7	22,1
<b>Ensemble</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

Au niveau de quelques caractéristiques professionnelles, il apparaît que le travail à temps partiel est très peu présent (1,5 %). L'Île-de-France accueille plus de 39 % de l'ensemble. Il est à remarquer que les ingénieurs issus de CPGE travaillent plus souvent en Ile de France que ceux issus des autres formations. Enfin, les ingénieurs sont peu nombreux à travailler dans le secteur public (8,3 %).

Tableau 9

## Caractéristiques professionnelles des diplômés ingénieurs de l'échantillon.

Caractéristiques professionnelles	BTS	DUT	DEUG	CPI	CPGE	Total
Travail à temps partiel	0	0	0	1,8	2,4	1,5
Travail en Ile de France	21,15	30,9	35,0	40,0	45,1	39,7
<b>Nombre de salariés dans l'entreprise</b>						
Moins de 50	37,2	23,1	25,0	17,6	18,3	20,9
Entre 50 et 199	17,7	20,0	14,8	19,3	21,8	20,0
Entre 200 et 499	15,7	20,0	16,4	21,9	17,4	18,1
Plus de 500	21,6	30,0	34,4	36,8	33,1	32,7
Secteur public	7,8	6,9	9,4	4,4	9,4	8,3
<b>Ensemble</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>	<b>100</b>

La question se pose désormais de savoir si les ingénieurs issus de classes préparatoires retirent des gains significativement supérieurs aux autres du fait du passage par la classe prépa.

## 2.2. Analyse de la relation entre la formation bac+2 initiale et les conditions d'insertion des ingénieurs

Avant de commencer l'analyse des salaires proprement dite, nous allons porter un premier regard général sur l'insertion des ingénieurs en fonction de leur formation bac+2 à partir de quelques indicateurs.

Tableau 10  
Indicateurs d'insertion selon la formation bac+2.

Indicateurs d'insertion					
Formation bac+2	Durée moyenne d'accès à l'emploi	Part du temps au chômage	Taux de chômage en mars 2001	Part des emplois précaires	Taux de cadres en mars 2001
Prépa	2,5 mois	6,6 %	2,1 %	7 %	93,1 %
Autre	2,5 mois	6,3 %	1,8 %	5,6 %	90,4 %
ENSEMBLE	2,5 mois	6,5 %	2 %	6,6 %	92,2 %

La durée moyenne d'accès à l'emploi pour l'ensemble des diplômés a été de 2,5 mois. Les ingénieurs issus préalablement des classes préparatoires semblent s'être insérés dans le même laps de temps que les autres. La part du temps de chômage apparaît équivalente pour les deux groupes d'ingénieurs. Si le taux de chômage des ingénieurs issus de classe préparatoire se révèle un peu plus élevé que celui des autres, il apparaît également qu'ils sont plus nombreux que les autres à avoir un emploi précaire. Concernant l'accès au statut de cadre, les ingénieurs prépas sont un peu plus nombreux à en bénéficier trois ans après l'entrée sur le marché du travail. Ces premiers indicateurs ne permettent pas de conclure sur les différences de qualité d'insertion entre les ingénieurs passés par les classes préparatoires et les autres.

Tableau 11  
Salaires des ingénieurs selon leur formation bac+2.

Salaires annuels (en €)	TOTAL	PRÉPA	AUTRES	Différence entre les prépas et les autres
Effectifs	849	560	289	-
Salaires moyen	32 254	32 703	31 383	1 320
Ecart-type	74,3	74,4	73,6	-
Salaires minimum	13 095	13 095	14 879	-
Salaires maximum	73 404	62 946	73 404	-

Considérons ce qu'il en est des niveaux de rémunérations. Une première approche comparative des salaires (Tableau 11) montre que, trois ans après la sortie de l'école, les ingénieurs prépas gagnent en moyenne 1 320 euros de plus par an que les autres. Un test de comparaison de moyennes indique que cette différence de salaire est significative à 5 %. Cependant une telle comparaison n'est pas satisfaisante et cela à deux titres au moins. Comme nous l'avons vu précédemment les classes prépa peuvent être surreprésentées dans les écoles les plus rémunératrices. Or c'est la formation prépa elle-même dont nous visons à évaluer l'impact sur le marché du travail et non son effet médiatisé à travers l'école d'ingénieurs. Deuxièmement, comme nous allons l'expliquer, on ne sait pas dans quelle mesure les deux populations, ingénieurs passés ou non par une prépa, sont directement comparables. Nous allons rendre compte de la comparaison des salaires en suivant l'ordre des deux obstacles évoqués ci-dessus. Il s'agira donc dans un premier temps d'aborder la question de l'effet des

écoles sur les salaires puis de voir dans quelle mesure les deux groupes d'ingénieurs, ceux issus des classes prépas et les autres, peuvent être comparés de manière aussi juste que possible.

### 3. L'EFFET DES ÉCOLES SUR LES SALAIRES DES INGÉNIEURS

#### 3.1. La nécessité du recours à une analyse multiniveau

Comme cela a été vu plus haut (Le Pellec, Roux, 2001), l'école de sortie apparaît comme l'un des facteurs déterminants pour expliquer les différences de salaires entre les ingénieurs. Il importe donc de le prendre en considération.

Dit simplement, l'intérêt de l'analyse multivariée, souvent mise en œuvre à travers la méthode des moindres carrés ordinaires, est d'étudier l'effet d'une variable en maintenant constantes toutes les autres (incluses dans le modèle). Que peut-on dire ainsi de l'effet sur le salaire du passage par une classe préparatoire à origine sociale égale, à niveau de réussite dans le secondaire égal, à spécialité de l'école de sortie égale, etc. ? Cependant, les données que nous utilisons ici ont une particularité qui les rend a priori impossibles à être traitées par la méthode des moindres carrés ordinaires (MCO). En effet, l'une des conditions de cette méthode est que les erreurs soient indépendantes les unes par rapport aux autres. La population analysée est composée d'anciens élèves d'un ensemble d'écoles d'ingénieurs. Or, le fait d'être issu d'une même école peut constituer pour des anciens élèves un facteur commun d'accès à l'emploi. Certaines écoles peuvent être plus prestigieuses que d'autres, certaines peuvent offrir des réseaux plus actifs, etc. Dans ce cadre, omettre l'école du modèle peut conduire à produire des erreurs de même amplitude pour les élèves d'une même école. Ainsi, les erreurs ne sont pas indépendantes. Une contrainte imposée par la méthode implique un nombre minimal d'observations par établissement. La limite fixée ici a été de huit individus par école. En conséquence, l'échantillon est restreint 690 individus et 46 établissements.

#### Limites des moindres carrés ordinaires et apports de l'analyse multiniveau

On peut formaliser le problème de la façon suivante :

Soit  $y_{ij}$  le revenu  $y$  de l'individu  $i$  de l'établissement  $j$

$\beta_0$  la valeur de la constante

$x_{ij}$  une variable explicative du salaire

$\beta_1$  la pente de la droite de régression

$e_{ij}$  représente une erreur aléatoire (résidu) associée à chaque ancien étudiant  $i$  de l'établissement  $j$  de moyenne nulle et de variance  $\sigma^2$ .

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + e_{ij} \quad [1]$$

Les résidus n'étant pas indépendants les uns des autres, l'utilisation des MCO peut entraîner un biais dans les erreurs-types des coefficients des variables, ce qui peut affecter leur significativité ; on peut être ainsi conduit à considérer significative une variable qui ne le serait pas.

Dans les modèles multiniveau, on modélise les variances estimées sur l'échantillon considéré. On considère ainsi que les établissements proviennent d'une population mère. La variance des

paramètres est composée d'une partie « vraie » (explicable) et d'une partie aléatoire. La méthode multiniveau va "purger" la variance des coefficients de sa partie aléatoire.

Les équations suivantes du modèle sont estimées simultanément par la méthode.

Au niveau des anciens élèves

$$y_{ij} = \beta_{0j} + \beta_{1j}x_{ij} + e_{ij} \quad [2]$$

Au niveau des établissements :

$$\beta_{0j} = \gamma_{00} + \gamma_{01}z_j + u_{0j} \quad [3]$$

$$\beta_{1j} = \gamma_{10} + \gamma_{11}z_j + u_{1j} \quad [4]$$

Dans l'équation [3], l'erreur aléatoire  $u_{0j}$  indique que les constantes varient d'un établissement à l'autre (d'où l'indice  $j$  qui leur est attribué dans [2]). La constante moyenne dans l'échantillon est représentée par  $\gamma_{00}$  ;  $u_{0j}$  représente l'écart de chaque établissement par rapport à cette moyenne. Plus la constante d'un établissement est élevée, plus élevé est le niveau de salaire de ses anciens élèves.

Dans l'équation [4], l'erreur aléatoire  $u_{1j}$  indique que les pentes  $\beta_{1j}$  varient d'un établissement à l'autre. La pente moyenne dans l'échantillon est représentée par  $\gamma_{10}$  ;  $u_{1j}$  représente l'écart de chaque établissement par rapport à cette pente moyenne.

Dans les équations [3] et [4],  $z_j$  représente une variable de niveau 2, caractérisant l'établissement  $j$ , comme son ancienneté, son type (par exemple école universitaire ou autre), son champ disciplinaire. Elle est introduite comme une variable susceptible d'expliquer les variations de constantes et de pentes entre les établissements.

Si l'on intègre [3] et [4] dans [2], on obtient l'équation suivante :

$$y_{ij} = \gamma_{00} + \gamma_{01}z_j + \gamma_{10}x_{ij} + \gamma_{11}z_jx_{ij} + (u_{0j} + u_{1j}x_{ij} + e_{ij})$$

L'équation comporte une partie fixe et une partie aléatoire (représentée entre parenthèses). La partie aléatoire constitue la caractéristique des modèles multiniveau par rapport aux MCO. La méthode va estimer les variances et covariances des termes résiduels. Elle permettra de juger dans quelle mesure l'écart est significatif entre les éléments d'un même niveau (entre les individus ou entre les établissements) et de conclure à l'intérêt ou non de tenir compte en particulier de l'effet de l'établissement.

### 3.2. Les résultats de l'analyse multiniveau

Une régression linéaire multiniveau du logarithme des salaires a donc été effectuée en deux temps. Dans une première phase, un modèle, communément appelé modèle vide parce qu'il ne contient aucune variable explicative, a été estimé. Celui-ci permet d'appréhender la variance résultant des différents niveaux d'analyse. Ces niveaux sont ici au nombre de deux : le niveau établissement (niveau 2) et le niveau individu (niveau 1). Ce premier modèle révèle une variance résiduelle significative tant au niveau établissement qu'au niveau individuel. Il apparaît ici que 20 % des différences de salaire sont liées au fait de sortir de telle ou telle école. Dans une seconde phase, un autre modèle a été estimé en prenant en considération cinq types de variables : le sexe, la catégorie socio professionnelle du père, la spécialité du

diplôme d'ingénieur obtenu, la région de travail du diplômé et enfin sa formation bac+2 d'origine. La spécialité du diplôme ainsi que la région du lieu de travail sont les variables qui ont le plus d'impact sur le salaire. Viennent ensuite l'origine sociale et le sexe. Deux résultats importants apparaissent ici. Le premier résulte du constat selon lequel l'introduction de variables supplémentaires annule la variance liée au niveau établissement. Ceci provient sans doute du fait que l'effet établissement tenait fortement à la spécialité de formation. Dans le cas présent il sera légitime de recourir à une analyse par la méthode des MCO. Il est à noter eu égard à notre problématique initiale qu'il n'apparaît pas de différences significatives de salaires (au moins à 10 % de risque de première espèce) entre les ingénieurs préalablement passés par une classe préparatoire et ceux issus d'une autre formation bac+2.

*Tableau 12*  
**Régression linéaire multiniveau sur le logarithme des salaires.**

Régression multiniveau du log du salaire	Coef.	Sigma	Sig.
Constante	7,98	0,042	***
Femme	-0,034	0,038	ns
<b>Cadres (Référence)</b>			
Agriculteurs	-0,70	0,066	ns
Artisans et chefs d'entreprise	0,02	0,055	ns
Professions intermédiaires	0,006	0,048	ns
Employés	-0,04	0,046	ns
Ouvriers	-0,026	0,065	ns
<b>Informatique, électronique (Référence)</b>			
Technologies industrielles	-0,002	0,045	ns
Agriculture, agro-alimentaire	-0,426	0,056	***
Mécanique	0,014	0,056	ns
Matériaux et énergie	-0,161	0,059	***
<b>Travail en Ile de France (Référence)</b>			
Travail en province	-0,159	0,032	***
<b>Classe préparatoire (Référence.)</b>			
Formation bac+2 autre que prépa	-0,009	0,034	ns
<b>n</b>		<b>603</b>	
<b>Part de variance au niveau établissement (%)</b>		<b>0</b>	
<b>Part de variance au niveau individu (%)</b>		<b>100</b>	

Notes : • "Coef." : valeur du coefficient de régression. "Sigma" : valeur de l'écart-type.

• "Sig." : significativité du coefficient. "ns" : non significatif.

Lecture : \* : significatif à moins 10 % ; \*\* : significatif à moins 5 % ; \*\*\* : significatif à au moins 1 %.

Dans ce modèle la personne de référence est un ingénieur homme dont l'origine socio professionnelle du père est cadre. Il est diplômé en informatique, électronique. Il travaille en Ile de France et est passé par une classe préparatoire pour intégrer l'école d'ingénieurs.

L'intérêt du recours au logarithme du salaire est de pouvoir assimiler les coefficients de régression à un pourcentage de salaire.

Ainsi toutes les variables incluses dans le modèle étant égales par ailleurs, un ingénieur travaillant en province a ici un salaire inférieur de 15,9 % à celui d'un ingénieur travaillant en Ile de France.

Cependant, ce résultat ne peut être considéré définitif dans la mesure où nous ne nous sommes pas assurés des conditions de comparabilité entre les diplômés issus d'une prépa et les autres. En effet le processus par lequel les étudiants, en fonction de certaines de leurs caractéristiques, peuvent préférentiellement s'orienter vers les classes préparatoires n'a pas été pris en compte. On peut dès lors être conduit à mésestimer l'effet de la classe préparatoire.

#### **4. LA QUESTION DE LA COMPARABILITÉ ENTRE LES ÉTUDIANTS PASSÉS PAR UNE CLASSE PRÉPARATOIRE ET LES AUTRES**

Dans quelle mesure la différence de salaire entre les anciens de prépa et les autres peut-elle être attribuée au fait d'avoir fait une classe préparatoire ? Cette question suppose la résolution d'un certain nombre de problèmes, liés à la comparabilité des deux groupes d'ingénieurs, que l'on retrouve beaucoup en évaluation bio-médicale par exemple, ainsi que dans l'évaluation des programmes d'aide en faveur de certains publics<sup>14</sup>. L'objet de l'évaluation revient dans ces domaines à déterminer l'impact d'un "traitement" (ou du bénéfice d'un dispositif d'aide) en comparant deux groupes, l'un ayant subi le traitement et l'autre pas. Nous allons dans cette partie d'abord expliquer la problématique générale que recouvre ce type d'évaluation. Puis, nous présenterons les méthodes qui peuvent être utilisées et plus particulièrement la méthode dite d'appariement en insistant sur le déroulement de sa mise en œuvre. Enfin, nous présenterons les résultats auxquels nous parvenons à partir de la méthode d'appariement.

##### **4.1. La problématique de l'évaluation de l'impact d'un "traitement" appliquée aux classes préparatoires**

Pour répondre à cette question de savoir si la différence de salaire entre les anciens de prépa et les autres peut être attribuée au fait d'avoir fait une classe préparatoire, il faudrait savoir ce qui serait arrivé aux anciens de prépas s'ils avaient intégré une école d'ingénieurs sans avoir fait prépa. La première réaction serait d'utiliser l'échantillon qui n'a pas fait de prépa comme groupe de comparaison, pour inférer ce qui serait arrivé au premier groupe s'il n'avait pas suivi une prépa. Cette comparaison, telle que nous l'avons présentée plus haut (32 703 euros annuels pour les ingénieurs prépas significativement supérieurs à 31 383 euros annuels pour les ingénieurs non prépas), contourne cependant une question fondamentale : les ingénieurs comparés sont-ils directement comparables ? De la réponse à cette question dépend la qualité du résultat de l'évaluation.

Dans un cadre comme celui-ci où l'on souhaite évaluer l'impact d'une formation sur des individus, l'idéal consisterait à comparer le résultat sur deux groupes d'individus qui à l'origine seraient en tout point identiques. Les différences observées ne pourraient dès lors être imputables qu'au seul fait d'avoir suivi ou non la formation. Dans la réalité sociale, une telle situation s'avère inaccessible. La formalisation théorique, telle qu'on la trouve dans la littérature en évaluation, va encore plus loin. Ici, évaluer l'impact d'une formation revient à déterminer pour chaque individu ce qui se serait produit si celui-ci n'avait pas suivi la formation. En d'autres termes quel aurait été le salaire de chacun des ingénieurs qui ont fait prépa s'ils n'avaient pas fait prépa. L'estimation de cette dernière situation appelée ici contrefait (counterfactual en anglais) est au cœur de l'évaluation de l'impact de la formation. Voici, présenté dans l'encadré ci-dessous, comment celle-ci s'articule en terme de formalisation lorsqu'on l'applique au cas des classes préparatoires.

---

<sup>14</sup> Voir toute la littérature liée, entre autres, au travail de James Heckman sur la question des biais de sélection.

### Détermination du contrefait ou counterfactual

Soit  $D_i$  la variable binaire indiquant si oui ou non l'individu a suivi une classe préparatoire.  $D_i$  peut donc prendre deux valeurs possibles, soit  $D_i = 1$  si l'ingénieur est passé par une prépa, ou  $D_i = 0$ , si ce n'est pas le cas. Si le ième ancien d'école d'ingénieur n'est pas passé par une classe prépa, alors son niveau de salaire est noté  $S_{0i}$ , qui représente le salaire du ième ingénieur, quand  $D_i = 0$ . Si l'ingénieur est passé par une prépa, alors son niveau de salaire est  $S_{1i}$ . Le gain provenant du fait d'être passé par une prépa est alors  $S_{1i} - S_{0i}$ .

Le gain pour le ième ingénieur qui est passé par classe préparatoire ( $D_i = 1$ ) est alors :

$$G_i = (S_{1i} - S_{0i}) / D_i = 1.$$

La barre signifie "sachant que" et est nécessaire pour bien indiquer que le calcul du gain s'applique à un individu qui est passé par une classe prépa. Si l'on veut connaître le gain moyen, il suffit de calculer la moyenne de tous les  $G_i$ , qui donne le gain moyen dans l'échantillon pour tous ceux qui sont passés par une prépa. Sous réserve que l'échantillon soit bien construit, cette moyenne peut être considérée comme une estimation sans biais de la vraie moyenne des gains. Cette dernière est l'espérance de gains, et peut s'écrire

$$G = E[(S_{1i} - S_{0i}) / (D_i = 1)].$$

Dans la littérature, notamment médicale et pharmaceutique,  $E[(S_{1i} - S_{0i}) / (D_i = 1)]$  s'appelle quelque fois l'effet du traitement, ou l'effet moyen du traitement sur le traité (average treatment effect, en abréviation ATT). Dans ce cas, le passage en classe prépa est considéré comme le traitement.

Mais en réalité, on n'a pas calculé  $G$ , mais la différence de gains entre les ingénieurs qui sont passés par une prépa et ceux qui n'y sont pas passés. C'est l'estimation sur l'échantillon de :

$$D = E[S_{1i} / (D_i = 1)] - E[S_{0i} / (D_i = 0)].$$

L'identité suivante relie  $D$  et  $G$  :

$$D = G + B.$$

Le terme  $B$  est le biais dans l'estimation, et est donné par :

$$B = E[S_{0i} / (D_i = 1)] - E[S_{0i} / (D_i = 0)].$$

En d'autres termes, le biais est l'espérance de la différence de gains liée au non passage par une prépa entre les ingénieurs qui sont effectivement passés par une prépa et ceux qui n'y sont pas passés. Ce biais pourrait être corrigé si  $E[S_{0i} / (D_i = 1)]$  était connu, mais ce n'est pas possible. On ne peut observer quel salaire aurait perçu un ingénieur qui est passé par une prépa s'il n'y était pas passé. Cet élément manquant est aussi appelé moyenne contre-factuelle.

Il faut se demander si, en l'absence de classes prépas, ceux qui sont passés par une telle classe, auraient eu un salaire plus élevé et de combien aurait été cette différence. S'il n'y avait pas de biais, le salaire supplémentaire dû au passage en classe prépa est bien égal à la différence de salaire entre les deux groupes. Ainsi, le biais existe s'il y a une différence entre



les moyennes de gains de ceux qui ont intégré une prépa et les autres, en l'absence de classes prépas.

Pour éliminer ce biais, la meilleure approche consisterait à affecter les bacheliers au hasard entre classes prépas et autres filières. Ainsi, les individus du groupe "traité" et les individus du groupe de "contrôle" auraient la même espérance de gains en l'absence de classes préparatoires. On pourrait écrire que

$$E[S_{0i} / (D = 1)] = E[S_{0i} / (D_i = 0)].$$

Le salaire de ceux qui ne sont pas passés par une prépa serait alors la mesure de la variable contre-factuelle, c'est-à-dire une mesure du salaire qu'auraient perçu ceux qui n'ont pas fait prépa. L'affectation aléatoire égaliserait d'ailleurs toute la distribution et pas seulement les moyennes.

Dans le système actuel, on peut légitimement penser que ceux qui entrent dans une classe préparatoire ont des caractéristiques spécifiques, en terme de brio académique notamment. Et que ces caractéristiques pourraient leur permettre d'accéder à des emplois plus rémunérateurs, même sans être passés par une classe préparatoire. Mais si les classes préparatoires sont une formation efficace, on doit s'attendre à ce que les ingénieurs issus des classes prépas perçoivent des salaires plus élevés que ceux qu'ils auraient perçus s'ils n'étaient pas passés par ces classes. Ainsi,  $E[S_{1i} / (D_i = 1)] > E[S_{0i} / (D_i = 1)]$  et donc le calcul originel présenté plus haut (32 703 euros annuels pour les ingénieurs prépas significativement supérieurs à 31 383 euros annuels pour les ingénieurs non prépas), surestime l'accroissement de gains dû au passage par une classe prépa. Il s'agit maintenant de voir comment concrètement le calcul du vrai gain de la classe préparatoire peut être déterminé.

Dans la pratique, deux types de méthodes permettent cette estimation du contrefait : d'une part les méthodes dites expérimentales et, d'autre part, les méthodes dites quasi-expérimentales ou non expérimentales. Pour les premières comme pour les secondes, déterminer le contrefait se réalise par le recours à un groupe de contrôle ou de comparaison (ensemble d'individus qui n'ont pas suivi la formation mais ayant autant que possible les mêmes caractéristiques que ceux qui l'ont suivi). Les méthodes expérimentales sont considérées comme les plus robustes. Elles consistent, à partir d'une population cible, à répartir de manière aléatoire les individus en deux groupes, l'un suivant la formation (le groupe "traité") et l'autre pas (le groupe de contrôle). En théorie, les deux groupes ainsi créés sont statistiquement équivalents et l'impact de la formation peut simplement se mesurer en comparant le résultat moyen des deux groupes. Dans la pratique en sciences sociales, un tel dispositif peut s'avérer difficile à mettre en place voire impossible. On perçoit aisément ce type de difficulté dans le cas de l'orientation des bacheliers en classe préparatoire qui dépend du "libre" choix du bachelier, de sa famille et de l'établissement d'accueil. Dans un tel cas, il est alors possible de recourir aux méthodes quasi-expérimentales. Ici, le groupe traité et le groupe de contrôle résultent des choix des individus et s'imposent à l'évaluateur. Dès lors, la tâche de ce dernier consistera à reconstruire un groupe de contrôle à partir du groupe de contrôle socialement constitué en sélectionnant les individus ressemblant autant que possible à ceux du groupe traité (au moins en termes de caractéristiques observées). Si les méthodes quasi expérimentales apparaissent plus simples en s'appliquant directement au cadre social construit, elles sont cependant statistiquement complexes et moins robustes que les méthodes expérimentales en raison principalement de la question des biais de sélection auxquelles elles sont confrontées.

Dans le cas présent, la question des biais de sélection repose sur l'idée que l'orientation en classe prépa ne se fait pas par hasard mais résulte d'un certain nombre de facteurs impliquant que les élèves s'orientant en prépa ont vraisemblablement des caractéristiques qui les distinguent des autres. Par exemple, selon leur origine sociale, leur niveau de réussite dans l'enseignement secondaire, leur établissement d'accueil dans le secondaire, les élèves peuvent plus ou moins vouloir s'orienter vers une classe prépa. Dès lors, dans l'analyse des résultats, il importe de pouvoir prendre en compte ces caractéristiques afin qu'elles ne viennent pas biaiser l'estimation de l'impact propre à la classe préparatoire.

Il existe deux grands types de biais. Il y a ceux liés à des différences observables à travers les données dont on dispose, il s'agira par exemple de caractéristiques comme celles citées ci-dessus telles que l'origine sociale, le niveau de réussite dans le secondaire, la présence de classes prépa dans l'établissement secondaire d'accueil. Le second type de biais correspond à des différences non observables qui, comme leur nom l'indique, renvoient à l'influence de facteurs dont on ne peut avoir connaissance à travers les données dont on dispose (c'est généralement l'influence de ces dernières que l'on nomme biais de sélection). Il pourrait, par exemple, s'agir de compétences particulières comme nous aurons l'occasion d'en rediscuter ultérieurement. Ces deux sources de biais peuvent affecter de manière déterminante les résultats. On voit ici l'intérêt des méthodes expérimentales qui, si elles n'éliminent pas les biais, permettent de les contrôler en les répartissant aléatoirement entre le groupe traité et le groupe de contrôle. Le défi des méthodes quasi-expérimentales est d'essayer de modéliser au mieux le processus de sélection en jeu pour le maintenir constant dans la comparaison entre les deux groupes. Parmi ces dernières, les méthodes d'appariement (*matching*) apparaissent comme les plus utilisées.

Appliquée à notre cas, l'idée du *matching* est de constituer deux groupes en trouvant pour chaque ingénieur prépa un ingénieur non prépa ayant exactement les mêmes caractéristiques et ensuite de comparer le résultat moyen du groupe des ingénieurs prépas à celui des ingénieurs non prépas. Dans la pratique, le recours à ce type d'appariement direct dépend du nombre de caractéristiques que l'on prend en compte pour faire l'appariement. Plus elles sont nombreuses et plus il devient difficile de trouver dans le groupe de contrôle des individus ayant exactement les mêmes caractéristiques de chacun des individus du groupe traité. Dans une telle situation, il est possible de recourir à une méthode d'appariement par probabilité prédite, dans notre cas de passer par une classe préparatoire (*propensity score matching*).

#### **4.2. Mise en œuvre de la méthode d'appariement par probabilité prédite ou *propensity score matching***

D'une manière générale il s'agit d'un type d'appariement qui tend à être de plus en plus fréquemment utilisé. Ici le groupe de contrôle est appareillé au groupe traité sur la base de la probabilité prédite de faire prépa (*propensity score*), qui a l'avantage d'être le résumé d'un ensemble de caractéristiques des individus. Nous allons utiliser l'exemple de cette méthode pour présenter concrètement la démarche que nous avons suivie pour évaluer l'impact du passage en classes préparatoires sur les salaires des ingénieurs.

La première étape consiste à modéliser, en fonction des variables disponibles dans la base de données, le processus par lequel certains individus s'orientent préférentiellement en prépa. On peut regretter que la variable de mention au bac ne figure pas dans la base dans la mesure où elle conditionne fortement l'entrée dans une classe préparatoire. Par contre, nous disposons de l'âge au bac qui peut faire figure de variable traduisant le brio académique. Il aurait également

été intéressant de savoir si le lycée où le jeune a terminé ses études secondaires, offrait ou non une classe préparatoire. La seule variable permettant de caractériser l'établissement est son département d'implantation. Deux variables individuelles sont disponibles dans la base, le genre et la profession du père. A donc été testé un modèle régressant la probabilité d'être en classe préparatoire sur l'âge au bac, la localisation de l'établissement (Ile de France versus les autres départements), le genre et la profession du père. Seules les variables, profession du père et âge au bac se sont révélées significatives. Ce sont les résultats relatifs au modèle recourant à ces variables qui sont présentés dans le Tableau 13.

Tableau 13  
Régression logistique sur la probabilité d'être en classe préparatoire.

Régression logistique sur la probabilité d'être en classe préparatoire	Coef.	Sig.
Constante	9,85	***
<b>Cadres, professions intermédiaires et agriculteurs (Référence)</b>		
Artisans	-0,722	***
Employés	-0,447	**
Ouvriers	-1,151	***
Age au bac	-0,499	***
n	911	
D de Sommers	28 %	

Note : "Coef." : valeur du coefficient de régression ; "Sig." : significativité de celui-ci.

Lecture : \* : significatif à moins 10 % ; \*\* : significatif à moins 5 % ; \*\*\* : significatif à au moins 1 %.

Dans ce modèle la personne de référence est un ingénieur dont l'origine socioprofessionnelle du père est soit cadre, soit profession intermédiaire ou agriculteur.

Dans ce modèle, avoir un père artisan, employé ou ouvrier conduit significativement moins à s'orienter dans une classe préparatoire comparé aux PCS citées précédemment. Par ailleurs, plus l'âge au bac est élevé moins les bacheliers s'orientent vers la prépa.

A partir du modèle présenté dans le Tableau 13, on peut estimer les probabilités individuelles d'entrée dans une classe prépa pour chacun des individus composant notre population qu'il soit passé ou non par une classe prépa.

#### Calcul de la probabilité prédite d'être en classe préparatoire

Soit  $P(\text{prépa}=1)$  la probabilité prédite d'être en classe préparatoire. P s'exprime sous la forme suivante :

$$P(\text{prépa} = 1) = \frac{e^{\beta X}}{1 + e^{\beta X}}$$

$P(\text{prépa}=1)$  peut se calculer à partir de la formule suivante :

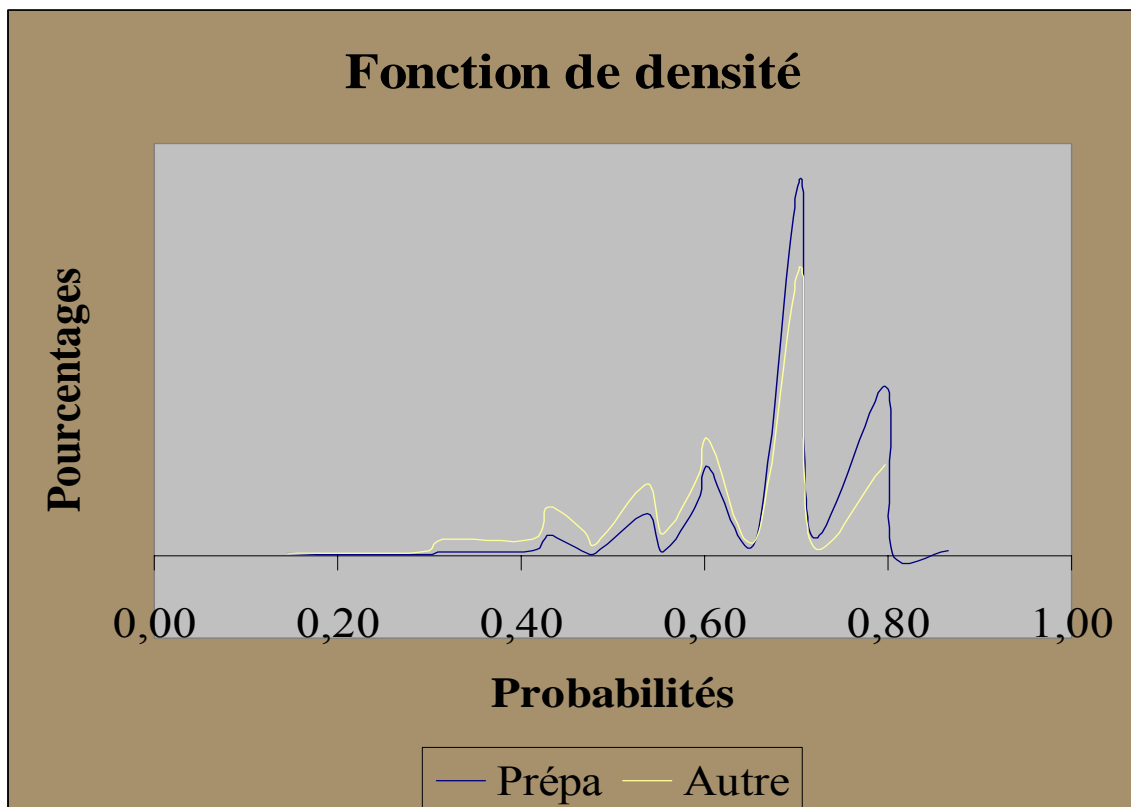
$$P(\text{prépa} = 1) = \frac{e^{(9,85 - 0,722 \times \text{artisan} - 0,447 \times \text{employé} - 1,151 \times \text{ouvrier} - 0,499 \times \text{âge au bac})}}{1 + e^{(9,85 - 0,722 \times \text{artisan} - 0,447 \times \text{employé} - 1,151 \times \text{ouvrier} - 0,499 \times \text{âge au bac})}}$$

Le tableau ci-dessous révèle l'étendue et la fréquence des probabilités pour chacun des deux groupes.

*Tableau 14*  
**Probabilités prédites de faire une classe prépa.**

Prépa			Autre		
Probabilité	%	%cumulés	Probabilité	%	%cumulés
0,14	0,33	0,33	0,14	0,32	0,32
0,22	0,17	0,50	0,22	0,32	0,65
0,30	0,17	0,66	0,25	0,32	0,97
0,31	0,50	1,16	0,30	0,65	1,62
0,41	0,66	1,82	0,31	1,95	3,57
0,43	2,49	4,31	0,41	1,95	5,52
0,47	0,50	4,81	0,43	5,84	11,36
0,48	0,33	5,14	0,47	2,92	14,29
0,54	5,14	10,28	0,48	1,30	15,58
0,55	0,50	10,78	0,54	8,77	24,35
0,59	6,47	17,25	0,55	2,60	26,95
0,60	10,78	28,03	0,59	9,42	36,36
0,66	1,99	30,02	0,60	13,96	50,32
0,70	45,77	75,79	0,66	1,95	52,27
0,72	2,65	78,44	0,70	35,06	87,34
0,80	20,56	99,00	0,72	1,62	88,96
0,81	0,33	99,34	0,80	11,04	100
0,87	0,66	100			

*Figure 1*  
**Distribution des probabilités prédites.**



L'intérêt du Tableau 14 et de la Figure 1 est de bien montrer que les bornes du spectre des probabilités prédites sont quasiment les mêmes pour les deux groupes. Cela nous confirme que nous comparons bien des individus ayant les mêmes caractéristiques au regard des variables observables. C'est là une condition indispensable à la qualité de la comparaison. Si des individus avaient eu des probabilités prédites en dehors de ces bornes communes (question du "*common support*" dans la littérature), il aurait fallu les retirer de la comparaison.

A ce niveau de la démarche, l'appariement à partir des probabilités prédites consiste à trouver pour chaque ingénieur prépa, un ingénieur non prépa ayant la même probabilité prédite. On parle alors d'appariement "un à un". Cette procédure présente cependant l'inconvénient d'éliminer de la comparaison chaque ingénieur prépa n'ayant pas son homologue non prépa en termes de probabilité prédite. Pour pallier cette perte d'effectifs potentiellement importante, il est possible d'opérer un appariement basé sur l'estimateur kernel. L'avantage de ce dernier est qu'il utilise tous les individus du groupe de contrôle en leur affectant un poids en fonction de la distance de leur probabilité par rapport à chaque individu du groupe traité.

### **Procédures de *matching***

Chaque individu a deux possibilités de résultats,  $Y_0$  et  $Y_1$ , dans le groupe traité et dans le groupe de contrôle. Notons  $D = 1$  le fait de recevoir le traitement et  $D = 0$  l'absence de traitement.

Le problème de l'évaluation du programme provient du fait que nous n'observons que  $Y_0$  ou  $Y_1$  pour chaque individu, mais jamais les deux. On observe ainsi  $Y$  tel que  $Y = DY_1 + (1 - D)Y_0$ . Nous ne pouvons donc établir le gain net  $\Delta = Y_1 - Y_0$  pour quiconque. Dans les problèmes d'évaluation traditionnels, on a accès aux résultats des participants et à des données relatives à un groupe de contrôle. Ainsi, on peut construire la distribution conditionnelle des  $Y_1$ , étant donné un vecteur de variables conditionnelles  $X$  et  $D = 1$ , ainsi que la distribution conditionnelle des  $Y_0$  étant donné  $X$  et  $D = 0$ .

L'effet du traitement sur les traités peut être défini comme :

$$(1) \Delta(X) = E(\Delta / X, D = 1) = E(Y_1 / X, D = 1) - E(Y_0 / X, D = 1).$$

Pour que le paramètre (1) soit clairement interprétable, il faut que la distribution conditionnelle de  $X$  satisfasse  $F(X / Y_0, Y_1, D) = F(X / Y_0, Y_1)$ , c'est-à-dire que les conditions relatives aux résultats potentiels, étant donné  $D$ , ne « causent » pas ni ne prédisent  $X$ . Ceci permet d'éviter le problème de poser une condition sur les variables qui sont déterminées par  $D$  et qui risquent ainsi de masquer l'effet total de  $D$ .

Les données sur les participants identifient  $E(Y_1 / X, D = 1)$ . L'information manquante est celle nécessaire pour identifier  $E(Y_0 / X, D = 1)$ . La méthode de comparaison des groupes utilise les données du groupe de contrôle pour l'estimer. La méthode suppose que, étant donné  $X$ , les résultats des individus témoins approximent ce que le résultat des participants, s'ils n'avaient pas participé ; ainsi, elle suppose que  $E(Y_0 / X, D = 0) \cong E(Y_0 / X, D = 1)$ . Le biais de sélection,  $B(X)$ , associé à l'impact du programme,  $E(\Delta / X, D = 1)$  qui apparaît quand cette dernière hypothèse n'est pas vérifiée, est

$$B(X) = E(Y_0 / X, D = 1) - E(Y_0 / X, D = 0).$$

Si l'affectation aux deux groupes peut être réalisée aléatoirement, le biais est nul. Mais en général, une telle méthode ne peut être utilisée, pour des raisons éthiques ou pour des raisons pratiques. Pour résoudre le problème de comparaison quand les deux populations ne sont pas déterminées aléatoirement, on utilise différentes méthodes, dont celle de l'appariement.

La méthode est fondée sur l'hypothèse d'identification, selon laquelle, étant donné un ensemble de caractéristiques  $X$ ,  $Y_0$  est indépendant de  $D$ . On peut noter :

(A-1)  $Y_0 \perp D / X$ ,  $X \in \mathcal{X}_c$ , pour un ensemble  $\mathcal{X}_c$ , où " $\perp$ " symbolise l'indépendance et les variables à la droite de " $/$ " sont les variables conditionnelles. Cette hypothèse fournit un groupe de comparaison qui ressemble au groupe de contrôle d'une expérimentation selon un aspect clé : sous condition de  $X$ , la distribution des  $Y_0$ , pour  $D=1$ , est la même que la distribution des  $Y_0$  pour  $D=0$ . En particulier, si la moyenne existe :

$$E(Y_0 / X, D = 1) = E(Y_0 / X, D = 0), \text{ si bien que pour tout } X, \text{ le biais } B(X) = 0.$$

De nombreux estimateurs d'appariement ont été proposés qui utilisent la relation (A-1) ou la condition précédente sur les moyennes en toute valeur de  $X$ . Les méthodes traditionnelles d'appariement reposent sur la mise en correspondance d'individus qui ont des caractéristiques  $X$  identiques ou proches, en utilisant différentes métriques (voir Heckman et alii, 1998). En reprenant les notations de ces auteurs, pour chaque observation  $i$  dans l'échantillon du groupe traité, une moyenne pondérée des observations dans le groupe de contrôle est calculée pour estimer l'effet du traitement sur  $i$  :

$$Y_{li} - \sum_{j \in \{D=0\}} W_{N_0 N_1}(i, j) Y_{0j}$$

où  $\{D=0\}$  est l'ensemble des indices pour les individus du groupe de contrôle et  $\{D=1\}$  est l'ensemble des indices pour les individus du groupe traité.  $N_0$  est le nombre d'observation dans le groupe de contrôle,  $\{D=0\}$ ,  $N_1$  est le nombre d'observations dans le groupe traité,  $\{D=1\}$ , et  $\sum_{j \in \{D=0\}} W_{N_0 N_1}(i, j) = 1$  pour tout  $i$ .

Les estimateurs d'appariement varient selon la pondération attachée à chacun des membres du groupe de contrôle. On peut définir un voisinage  $C(X_i)$  pour chaque participant  $i$ . Les personnes appariées avec  $i$  sont dans  $A_i$  tel que  $A_i = \{j \in \{D = 0\} / X_j \in C(X_i)\}$ . Différentes méthodes d'appariement utilisent différents voisinages. L'appariement selon le voisin le plus proche établit que  $C(X_i) = \{X_j / X_i = \min_i \|X_i - X_j\|, j \in \{D = 0\}\}$ , où  $\| \cdot \|$  est une norme,  $W_{N_0 N_1}(i, j) = 1$ ,  $j \in A_i$  et  $W_{N_0 N_1}(i, j) = 0$  dans le cas contraire. En réalité, les voisins les plus proches peuvent être assez éloignés. Pour cette raison, on peut établir un critère garantissant que l'appariement ait un sens. L'appariement de caliper (*caliper matching*) définit  $C(X_i) = \{X_j / \|X_i - X_j\| < \varepsilon\}$ , où  $\varepsilon$  est déterminé arbitrairement. Si aucun  $X_j$  ne correspond à la spécification, l'observation  $i$  n'est appariée à aucune observation du groupe de contrôle. Si plus d'une personne est en  $A_i$ , le voisin le plus proche en terme de norme  $\| \cdot \|$  est utilisé pour réaliser l'appariement.

L'appariement de kernel définit

$$W_{N_0, N_1}(i, j) = \frac{G_{ij}}{\sum_{k \in \{D=0\}} G_{ik}}, \text{ où } G_{ik} = G\left(\frac{X_i - X_k}{a_{N_0}}\right)$$

est un kernel qui diminue le coefficient de pondération des observations distantes de  $X_i$  et  $a_{n_0}$  une séquence de paramètres d'amortissement ayant pour propriété  $\lim_{N_0 \rightarrow \infty} a_{N_0} = 0$ . Les valeurs non nulles de cette pondération définissent implicitement  $C(X_i)$  pour ce type d'appariement.

Au lieu de travailler sur des distances entre caractéristiques, on peut utiliser des techniques d'appariement qui comparent des individus à partir de leur probabilité de participation. On définit la probabilité de participation ou « *propensity score* » comme  $P(X) = P(D = 1 / X)$ .

Un théorème de Rosembaum et Rubin (1983) affirme que si (A.1) est satisfaite, alors (A-2)  $Y_0 \perp D / P(X)$  pour  $X \in \chi_c$ . Poser la condition sur  $P(X)$  plutôt que sur  $X$  produit une indépendance conditionnelle. L'implication suivant, issue de (A-2), indique ce qui est requis pour construire la moyenne conditionnelle contre-factuelle :

$$E(Y_0 / P(X), D = 1) - E(Y_0 / P(X), D = 0) = B(P(X)) = 0.$$

Cette implication peut être utilisée à la place de (A-1) ou de (A-2). Mettre la condition sur  $P(X)$  établit que  $B(P(X)) = 0$  et réduit la dimension du problème d'appariement à un appariement à partir du scalaire  $P(X)$ .

Le Tableau 15 indique les résultats obtenus en utilisant le logiciel Stata<sup>15</sup>. Ces derniers se lisent de la façon suivante. Sur la première ligne (non appariés) figure le salaire moyen annuel exprimé en euros de chacun des deux groupes. Il apparaît que les ingénieurs prépa gagnaient en 2001 en moyenne 1 320 euros de plus que les ingénieurs non prépa. Cette estimation brute incorpore les biais de sélection vus plus haut. La seconde ligne (ATT pour *Average Treatment Effect on the Treated*) correspond à la comparaison du salaire des ingénieurs prépa avec celui des ingénieurs non prépa ayant les mêmes caractéristiques qu'eux. D'après la théorie, le salaire des ingénieurs non prépa doit ici se comprendre comme ce qu'aurait gagné les ingénieurs prépa s'ils n'avaient pas fait prépa. Il s'agit là de l'impact en terme de salaire du passage par une prépa pour ceux qui sont passés par une prépa. Il apparaît une différence de 1 437 euros à la faveur des ingénieurs prépa. Ainsi, après trois ans sur le marché du travail, un ingénieur prépa pourrait d'après cette méthode gagner en moyenne 1 437 euros de plus sur une année. La troisième ligne (ATU pour *Average Treatment Effect on the Untreated*) compare le salaire des ingénieurs non prépa à celui des ingénieurs prépa ayant les mêmes caractéristiques qu'eux. Le salaire des ingénieurs prépa doit se lire comme le salaire qu'auraient eu les ingénieurs non prépa s'ils étaient passés par une prépa. Il apparaît ici qu'ils auraient gagné en moyenne 1 200 euros de plus trois ans après leur entrée sur le marché du travail. Enfin la dernière ligne (ATE pour *Average Treatment Effect*) correspond à la synthèse des deux lignes précédentes. Il s'agit de la différence de salaire liée au fait d'être ou non passé par la prépa estimée sur l'ensemble de la population. Autrement dit, cette moyenne combine l'avantage comparatif des ingénieurs prépa et le manque à gagner des ingénieurs non prépa du fait de ne pas être allé en prépa. Cette différence moyenne, c'est-à-dire l'effet moyen global s'élèverait à 1 356 euros en moyenne sur une année.

<sup>15</sup> Module psmatch2 : Edwin Leuven, Department of Economics, University of Amsterdam. Barbara Sianesi, Institute for Fiscal Studies, London, UK.

Tableau 15

**Différences de salaire entre les ingénieurs selon leur formation bac+2.**

En euros	Prépa	Autre	Différence
Non appariés	32 703	31 383	1 320
ATT	32 704	31 267	1 437
ATU	32 583	31 383	1 200
ATE	-	-	1 356

Une fois ces résultats obtenus, il importe cependant d'estimer dans quelle mesure ces différences de salaires peuvent être considérées comme significatives. Se révèlent ici l'avantage et l'inconvénient de la méthode d'appariement et notamment celle basée sur l'estimateur kernel. L'avantage se trouve au niveau du caractère non paramétrique de la méthode qui libère ainsi des vicissitudes liées à l'identification de la fonction caractérisant le mieux les distributions que l'on cherche à étudier. Mais c'est aussi là l'inconvénient puisque n'ayant pas cherché à caractériser ces distributions, il devient impossible de déterminer l'intervalle de confiance des estimations qui en découlent. Pour pallier cet inconvénient, il est possible de recourir à la technique de rééchantillonnage (*bootstrap* en anglais).

Cette technique permet de déterminer l'intervalle de confiance d'un estimateur  $\theta$  en l'absence de paramètres caractérisant la distribution dont il est issu. Pour cela, on procède un certain nombre de fois (100 fois par exemple) au tirage avec remise du même échantillon de base. On a ainsi 100 estimations du  $\theta$  qui constituent une distribution du  $\theta$  dont on peut supposer (loi des grands nombres) qu'elle suit une loi normale. Il devient alors aisé d'estimer son intervalle de confiance. Dans le cas présent, cette procédure a été menée pour la situation ATT (l'avantage comparatif pour les ingénieurs prépa d'avoir fait prépa). La différence moyenne de 1 437 euros à la faveur des ingénieurs prépa s'avère être significativement différente de zéro au seuil de 5 %.

En résumé, l'utilisation d'une méthode d'appariement pour évaluer l'impact du passage par une prépa sur le salaire des ingénieurs tend à faire apparaître une différence positive et significative de cette formation par rapport aux autres que sont le DEUG, le DUT et le BTS. Si les méthodes d'appariement s'avèrent moins robustes que les méthodes expérimentales, elles sont à leur tour plus robustes que les méthodes corrélatives. Cependant on voit bien dans quelle mesure la qualité de leurs résultats est très fortement liée à celle du modèle de sélection et à travers lui à la qualité et au nombre de variables dont on dispose pour modéliser le processus de sélection. C'est là une des limites manifestes de ce type de méthode qui renvoie à la question de la prise en compte des biais de sélection liés à des variables dites inobservables parce qu'absentes de la base de données. Connaissant ces limites, il importe de voir également combien ces méthodes d'appariement peuvent s'avérer pertinentes et efficaces en évaluation. Leur caractère non paramétrique constitue un avantage certain par rapport aux méthodes corrélatives souvent utilisées en évaluation. S'adaptant au cadre social, elles peuvent éviter le recours à des expérimentations toujours très lourdes à mettre en place tant en termes de protocole que de coûts. Par ailleurs, elles s'avèrent extrêmement pratiques pour l'évaluateur travaillant sur des bases de données déjà constituées.

## **5. LA RECHERCHE DE DIFFÉRENCES DE SALAIRES DUES À DES CARACTÉRISTIQUES INOBSERVABLES**

L'appariement permet de prendre en considération les biais liés aux différences observables, autrement dit, liés à des variables présentes dans la base de données utilisées. Cependant



comme nous l'avons vu plus haut, il existe une seconde source de biais, celle résultant de différences inobservables. La notion de biais de sélection désigne normalement ces dernières même si un certain usage tend à l'appliquer également aux biais résultant des différences observables. L'étude de cette question doit beaucoup, notamment, aux travaux de l'économiste américain James Heckman. Comme cela a déjà été évoqué, l'idée de base, concernant notre cas, est qu'il existe peut-être des caractéristiques inobservables qui agissent à la fois sur l'orientation en CPGE et sur le salaire (ambition, aptitudes, niveau d'information). Nous avons alors eu recours à la correction de biais de sélection telle qu'elle est proposée par Heckman.

### La correction des biais de sélection selon la procédure en deux étapes d'Heckman\*

Soit l'équation de salaire :

$$y_i = \beta' X_i + \delta PREPA_i + \varepsilon_i$$

avec  $y_i = \log(S_i)$  le logarithme du salaire de l'individu  $i$ . Soit  $PREPA_i$  variable binaire indiquant si oui ou non l'individu a suivi une classe préparatoire. La question est de savoir dans quelle mesure  $\delta$  représente la valeur du passage par une telle filière (en supposant que le reste de l'équation soit correctement spécifié). La réponse est négative si l'individu qui souhaite entrer dans une classe préparatoire et qui y est admis aurait eu des revenus plus élevés, qu'il soit passé ou non par une classe préparatoire. Le problème est celui de la sélection : auto-sélection d'une part et sélection par l'institution de l'autre. On peut faire l'hypothèse que les bacheliers qui souhaitent entrer dans une classe préparatoire ont des caractéristiques particulières, en termes de motivations par exemple. Par conséquent, si ces caractéristiques inobservées des bacheliers sont corrélées avec le fait d'intégrer une classe préparatoire, les estimations par MCO du coefficient associé à la variable de passage par une classe préparatoire sera biaisé car corrélé au terme d'erreur de l'équation de salaire. Si l'observation est correcte, l'estimation par les moindres carrés ordinaires de  $\delta$  surestimera certainement l'effet du traitement (ici le passage par la classe prépa).

En fait, le passage par une classe préparatoire est le résultat d'une interaction entre une demande individuelle et une offre institutionnelle (les individus sont acceptés sur dossier). On peut estimer que le choix se fonde à partir de la valeur de  $PREPA_i^*$  différence des valeurs présentes des gains espérés du passage en prépa et d'une orientation alternative (gains privés espérés par l'individu ou gains sociaux espérés par l'institution qui sélectionne).

$PREPA_i^*$  est fonction d'un ensemble de caractéristiques notées  $w_i$

$$PREPA_i^* = \gamma' w_i + u_i$$

$PREPA_i = 1$  si  $PREPA_i^* > 0$  et  $PREPA_i = 0$  si  $PREPA_i^* \leq 0$ ,

où  $\gamma$  est un vecteur de coefficients associés à chaque variable  $w$  affectant la probabilité d'entrer en classe préparatoire.  $u_i$  est une erreur symétrique aux propriétés standards ( $u \sim N(0, \sigma_u^2)$ ).

$\text{Prob}(PREPA_i = 1) = \Phi(\gamma' w_i)$  et  $\text{Prob}(PREPA_i = 0) = 1 - \Phi(\gamma' w_i)$ ,

où  $\Phi$  représente la fonction de répartition de la loi normale.

\* Voir par exemple Heckman J., "Sample selection bias as a specification error", *Econometrica*, janvier, 1979, Vol. 47, N°1, p. 153.

Nous supposons donc que  $u_i$  et  $\varepsilon_i$  suivent une distribution normale bivariée de moyenne nulle et de coefficient de corrélation  $\rho$ .

Nous pouvons augmenter l'équation de salaire d'une correction, dite d'Heckman, résultant de l'estimation d'une équation de sélection représentant l'entrée dans une classe préparatoire (approche à deux étapes développée par Heckman, 1978, 1979). Cette équation de sélection reprendra la spécification précédente, qui offre l'avantage de recourir à des variables indépendantes, qui ne sont pas reprises dans l'équation de gains. Nous pouvons donc adopter l'approche en deux étapes : *two-stage treatment effect*. Heckman (1979) propose d'estimer l'équation Probit pour obtenir une valeur de l'inverse du ratio de Mill, puis de l'inclure comme régresseur dans l'équation de salaire en deuxième étape. De cette façon, le contrôle des effets de sélection produira des estimations non biaisées des rendements de du passage en classe préparatoire dans l'équation de salaire.

$$\begin{aligned} \text{Dans ce cas, } E[y_i / \text{PREPA}_i = 1] &= \beta' x_i + \delta + E[\varepsilon_i / \text{PREPA}_i = 1] \\ &= \beta' x_i + \delta + \rho \sigma_\varepsilon \lambda \end{aligned}$$

avec  $\lambda$ , l'inverse du ratio de Mill, tel que  $\lambda = \frac{\phi(\gamma' w_i / \sigma_u)}{\Phi(\gamma' w_i / \sigma_u)}$ , où  $\phi$  représente la fonction de densité de la loi normale centrée réduite et  $\Phi$  la fonction de répartition de cette même loi. Pour les individus qui ne sont pas passés par une classe préparatoire,

$$E[y_i / \text{PREPA}_i = 0] = \beta' x_i + \rho \sigma_\varepsilon \frac{-\phi(\gamma' w_i / \sigma_u)}{1 - \Phi(\gamma' w_i / \sigma_u)}$$

Concrètement, la procédure revient à estimer l'équation de sélection pour calculer  $\lambda$  et l'introduire dans l'équation de gains.

Les variables du modèle de sélection sont ici les mêmes que celles utilisées dans le modèle de sélection lors de l'appariement. La seule différence tient à la spécification du modèle, qui est ici un Probit, les résultats obtenus étant pratiquement les mêmes que ceux obtenus précédemment.

Tableau 16  
Régression Probit sur la probabilité d'être en classe préparatoire.

Régression Probit sur la probabilité d'être en classe préparatoire	Coef.	Sig.
Constante	5,9	***
<b>Cadres, chefs d'entreprise, professions intermédiaires et agriculteurs (Référence)</b>		
Artisans	-0,443	***
Employés	-0,271	***
Ouvriers	-0,697	***
<b>Age au bac</b>	-0,298	***
<b>n</b>	<b>911</b>	
<b>D de Sommers</b>	<b>28 %</b>	

Note : "Coef." : valeur du coefficient de régression ; "Sig." la significativité de celui-ci.

Lecture : \* : significatif à moins 10 % ; \*\* : significatif à moins 5 % ; \*\*\* : significatif à au moins 1 %.

Dans ce modèle la personne de référence est un ingénieur dont l'origine socio professionnelle du père est soit cadre, soit profession intermédiaire ou agriculteur.

Le modèle de gain avec ratio de Mill révèle deux résultats importants. D'une part, il ne semble pas que l'on puisse conclure à l'existence d'un biais de sélection puisque le ratio n'est pas significatif. D'autre part, le passage par une classe préparatoire n'apparaît pas conduire à une différence significative de gains. Ce résultat est apparemment en contradiction avec celui obtenu précédemment. Cependant, dans la mesure où le ratio de Mill n'est pas significatif, il semble cohérent de réestimer le modèle sans l'inclusion du lambda. Dans ce cas, le passage par une classe préparatoire montre un effet positif et significatif à 5 % en termes salariaux.

Tableau 17

**Régression du log du salaire avec introduction de l'inverse du ratio de Mill.**

Régression du log du salaire	Coef.	Sigma	Sig.
Constante	7,77	0,025	***
Femme	-0,055	0,017	***
<b>Cadres, artisans, chefs d'entreprise, professions intermédiaires et ouvriers (Référence)</b>			
Agriculteurs	-0,045	0,03	ns
Employés	-0,030	0,02	ns
<b>Informatique, électronique (Référence)</b>			
Agriculture, agroalimentaire	-0,180	0,02	***
Transformation des matériaux et énergie	-0,080	0,02	***
<b>Travail en Île-de-France (Référence)</b>			
Travail en province	-0,112	0,014	***
<b>Classe préparatoire (Référence.)</b>			
Formation bac+2 autre que prépa	-0,033	0,065	ns
<b>Lambda</b>	0,001	0,04	ns
<b>n</b>		<b>849</b>	
<b>R<sup>2</sup> (%)</b>		<b>19,9</b>	

Note : "Coef." : valeur du coefficient de régression. "Sigma" : valeur de l'écart-type ; "Sig." : significativité du coefficient. "ns" : non significatif.

Lecture : \* : significatif à moins 10 % ; \*\* : significatif à moins 5 % ; \*\*\* : significatif à au moins 1 %.

Dans ce modèle la personne de référence est un ingénieur homme dont l'origine socioprofessionnelle du père est soit cadre, soit artisan, chef d'entreprise, soit profession intermédiaire ou ouvrier. Il est diplômé soit en informatique, électronique, soit en technologies industrielles, soit en mécanique. Il travaille en Ile de France et est passé par une classe préparatoire pour intégrer l'école d'ingénieurs.

Tableau 18

**Régression linéaire sans l'inverse du ratio de Mill.**

Régression du log du salaire	Coef.	Sigma	Sig.
Constante	7,77	0,013	***
Femme	-0,055	0,017	***
<b>Cadres, artisans, chefs d'entreprise, professions intermédiaires et ouvriers (Référence)</b>			
Agriculteurs	-0,045	0,03	ns
Employés	-0,030	0,02	ns
<b>Informatique, électronique (Référence)</b>			
Agriculture, agroalimentaire	-0,180	0,02	***
Transformation des matériaux et énergie	-0,080	0,02	***
<b>Travail en Île-de-France (Référence)</b>			
Travail en province	-0,112	0,015	***
<b>Classe préparatoire (Référence.)</b>			
Formation bac+2 autre que prépa	-0,035	0,015	**
<b>n</b>		<b>849</b>	
<b>R<sup>2</sup> (%)</b>		<b>19,9</b>	

Lecture : les ingénieurs passés par une classe préparatoire ont, en moyenne, un salaire supérieur à celui des autres de 3,5 % trois ans après avoir obtenu leur diplôme. Note : l'application de la correction de Kennedy au coefficient de la variable formation bac+2 (Formation bac+2 autre que prépa) conduit à un ajustement de ce dernier qui le situe à -0,0355. La correction de Kennedy s'exprime de la façon

suivante:  $\hat{G} = \left[ c + \frac{1}{2} \times V(c) \right]$ ,  $c$  étant le coefficient.

Au regard de ces résultats, notre conclusion doit être relativement prudente. En effet, en nous étant entouré du maximum de précautions économétriques, il ressort un avantage salarial en faveur des ingénieurs qui ont débuté leurs études supérieures par une classe préparatoire. Cet avantage correspond trois ans après l'achèvement des études à une différence de l'ordre de 3,5 %. On pourra remarquer que cette différence est en cohérence avec celle mise en lumière par la démarche précédente d'appariement.

## CONCLUSION

Dans le cadre des discussions engagées sur l'avenir du système éducatif français, il est important de s'interroger sur l'ensemble des structures de notre système éducatif, y compris les plus prestigieuses d'entre elles. Les classes préparatoires aux grandes écoles méritent un examen attentif dans la mesure, d'une part, où elles constituent la clé de voûte de la préparation de l'élite française mais, d'autre part, où elles représentent l'une des filières les plus coûteuses en termes de dépenses par étudiant. Il s'agit donc d'analyser si cette dépense plus élevée se traduit par une productivité supérieure de leurs anciens élèves dans leur activité professionnelle.

Le premier résultat de notre investigation est qu'aujourd'hui, compte tenu de l'expansion massive des écoles d'ingénieurs universitaires à partir de 1990, les classes préparatoires ne conduisent qu'un élève sur dix vers une école qui leur est spécifiquement destinée. Les autres poursuivent dans des écoles qui accueillent également des titulaires de diplômes universitaires ou de brevets de techniciens supérieurs dans des proportions variables. Les écoles à recrutement mixte vont donc faire se côtoyer des élèves qui ne sont sans doute pas les plus brillants des classes préparatoires et des élèves issus des meilleurs, des plus motivés, et des mieux informés des autres filières.

On peut dès lors s'interroger sur le fait de savoir si pour ceux qui transitent par une école à recrutement mixte le passage par une classe prépa marque la trajectoire de l'individu, y compris dans son parcours professionnel. Les données recueillies par l'enquête Génération 98 du Céreq nous aident à répondre à cette question.

Tout d'abord, il apparaît que la très grande majorité des anciens élèves atteignent rapidement le statut de cadre, qu'ils soient donc anciens d'une prépa ou d'une autre filière préalable à leur entrée dans une école. Cependant une légère différence s'observe en faveur des anciens de prépa (cadres 93 % contre 90 % pour les autres). En deuxième lieu, le passage par une classe prépa semble conférer un léger avantage en terme de rémunération, qui se révèle significatif tant dans la procédure d'appariement que dans le modèle de gains. Cet avantage correspond, d'après nos données et trois ans après la sortie du système éducatif, à 3,6 % de rémunération.

Les constats présentés dans cet article reposent sur des observations conduites en début de carrière, il pourrait être intéressant de les reprendre avec une période d'observation plus longue lorsque les données le permettront. Par ailleurs, il serait intéressant de s'interroger sur les compétences développées par les différentes filières conduisant aux écoles d'ingénieurs et sur les effets à long terme que pourraient avoir ces compétences sur les trajectoires professionnelles des diplômés.

## RÉFÉRENCES BIBLIOGRAPHIQUES

- Bourdieu P., *La Noblesse d'État*, Editions de Minuit, 1988
- Brand J., Halaby C., Regression and matching estimates of the effects of elite college attendance on career outcomes, communication présentée au congrès annuel de l'Association américaine de sociologie, Atlanta, 2003, [www.ssc.wisc.edu/~jlowry/elite\\_jbrand\\_rev1203.pdf](http://www.ssc.wisc.edu/~jlowry/elite_jbrand_rev1203.pdf).
- Gaulupeau Y., *La France à l'école*, Découverte Gallimard, n°147, 1997.
- Cier B., "Les admissions parallèles dans les écoles de commerce", Espace Études Éditions, 1999.
- Danin-Boileau H., "Ruptures psychiques en classes préparatoires" in *Classes préparatoires, des étudiants pas comme les autres*, sous la direction de Claire-Marine François-Poncet et d'Alain Braconnier, Bayard Éditions-Fondation de France, 1998.
- Direction de l'Évaluation et de la Prospective, "le coût de l'éducation en 2002", *Note d'information* 03.57, novembre 2003.
- DPD, Direction de la Programmation et du Développement, *Tableaux statistiques* 6787, juin, 2001.
- DPD, Direction de la Programmation et du Développement, *Tableaux statistiques* 6787, juin, 2001.
- Duru-Bellat Marie, Kieffer Annick, Adangnikou Noël, "Efficacité et équité de l'enseignement supérieur français dans la production des élites", Premier colloque scientifique du Réseau d'étude sur l'enseignement supérieur (RESUP), mai 2002.
- Green W. H., "Econometric analysis", 4<sup>e</sup> édition, Prentice Hall, 2000.
- Heckman J., "Sample selection bias as a specification error", *Econometrica*, january, 1979, Vol. 47, N°1, p. 153
- Heckman J., "Selectivity bias : new developments, varieties of selection bias", *The American Economic Review*, May, 1990, p. 313
- Heckman J., Ichimura H., Todd P. Smith J., "Characterizing selection bias using experimental data", *Econometrica*, septembre, 1998, p. 1017, disponible sur Proquest.
- Heckman J., Ichimura H., Todd P., "Matching as an econometric evaluation estimator", *The Review of Economic Studies*, 1998, vol. 65, p. 261, disponible sur Proquest.
- Lahire B., Les manières d'étudier, *Cahier de l'Observatoire de la Vie Étudiante* n°2, La Documentation française. 1997.
- Le Pellec L., Roux S., "Les salaires des ingénieurs diplômés en 2000", *Insee première*, Insee, n°812, novembre 2001.
- Magnac T., "L'apport de la microéconométrie à l'évaluation des politiques publiques", *Cahiers d'économie et sociologie rurales*, n°54, 2000.
- ONISEP, "Ecoles d'ingénieurs, niveaux d'admission", in *Après le bac... réussir ses études*, 2000.
- Sianesi B., "Implementing propensity score matching estimators with Stata", document de travail, Réunion des utilisateurs de Stata Royaume-Uni, Londres, Mai 2001.

## ANNEXES

### • Annexe 1

Tableau 19

Les nouveaux inscrits en cycle ingénieur dans les écoles d'ingénieurs 2000/2001 en distinguant les élèves issus de classes préparatoires ATS de ceux issus des CPGE.

		Nombre de nouveaux inscrits en cycle ingénieur 2000/2001						
		CPI	CPGE	CPATS	BTS	DUT	DEUG	Ensemble
Écoles d'ingénieurs dépendant du MEN	Écoles universitaires (95)	1 245	5 231	43	231	1 631	1 294	9 675
	ENI (4)	229	39		41	161	10	480
	INSA (4)	1 101	110		12	172	105	1 500
	ENSAM (12)	47	1 135	25	68	125	1	1 401
	Autres écoles (13)	43	1 582	10	47	15	17	1 714
	<b>Sous Total (128)</b>	<b>2 665</b>	<b>8 097</b>	<b>78</b>	<b>399</b>	<b>2 104</b>	<b>1 427</b>	<b>14 770</b>
	<b>%</b>	<b>50,4 %</b>	<b>57,4 %</b>	<b>32,4 %</b>	<b>53,6 %</b>	<b>76,7 %</b>	<b>80,8 %</b>	<b>59,4 %</b>
Écoles d'ingénieurs dépendant d'autres ministères	Agriculture (15)	50	738	115	13	17	96	1 029
	Ville de Paris (2)		86				4	90
	Industrie (8)	348	244	11		10	15	628
	Équipement transport (5)		358			1	13	372
	Santé (1)							0
	Télécommunications (3)		424		4			428
	Défense (11)		1 060			1	16	1 077
	<b>Sous Total (45)</b>	<b>398</b>	<b>2 910</b>	<b>126</b>	<b>17</b>	<b>29</b>	<b>144</b>	<b>3 624</b>
<b>%</b>	<b>7,5 %</b>	<b>20,6 %</b>	<b>52,3 %</b>	<b>2,3 %</b>	<b>1,1 %</b>	<b>8,2 %</b>	<b>14,6 %</b>	
<b>Écoles privées (64)</b>		<b>2 226</b>	<b>3 090</b>	<b>37</b>	<b>329</b>	<b>609</b>	<b>195</b>	<b>6 486</b>
<b>%</b>		<b>42,1 %</b>	<b>21,9 %</b>	<b>15,4 %</b>	<b>44,2 %</b>	<b>22,2 %</b>	<b>11,0 %</b>	<b>26,1 %</b>
<b>Total écoles d'ingénieurs (237)</b>		<b>5 289</b>	<b>14 097</b>	<b>241</b>	<b>745</b>	<b>2 742</b>	<b>1 766</b>	<b>24 880</b>
<b>%</b>		<b>100,0 %</b>	<b>100,0 %</b>	<b>100,0 %</b>	<b>100,0 %</b>	<b>100,0 %</b>	<b>100,0 %</b>	<b>100,0 %</b>
<b>Part de chaque formation dans l'ensemble</b>		<b>21,3 %</b>	<b>56,7 %</b>	<b>1,0 %</b>	<b>3,0 %</b>	<b>11,0 %</b>	<b>7,1 %</b>	<b>100,0 %</b>

Sources : Direction de la Programmation et du développement, *Tableaux Statistiques*, n°6787, "les écoles d'ingénieurs, effectifs des élèves en 2000-2001", juin 2001.

### • Annexe 2

Tableau 20

Évolution de la proportion de CPGE parmi les nouveaux inscrits en écoles d'ingénieurs depuis 1990

Pourcentage de CPGE parmi les nouveaux inscrits	1990-1991		1995-1996		2000-2001		2002-2003	
	(I)	(II)	(I)	(II)	(I)	(II)	(I)	(II)
<b>Écoles MEN</b>	47,4	60,3	46,6	60,2	51,9	67,3	50,4	65,8
dont écoles universitaires	19	20,3	45,7	52,9	52,3	62,4	45,9	52,9
dont ENSAM	84,1	87,9	84,4	88,4	81,4	85,4	84,4	88,4
<b>Écoles autres ministères</b>	86,8	87,9	90,1	91,3	91,8	93,9	87,0	92,1
<b>Écoles privées</b>	48,1	78,9	51,8	77,7	44,4	73,2	41,8	64,4
<b>Total</b>	<b>54,3</b>	<b>69,8</b>	<b>54,0</b>	<b>69,5</b>	<b>54,8</b>	<b>72,9</b>	<b>53,3</b>	<b>69,5</b>

Note : nouveaux inscrits : il s'agit des élèves admis au niveau du bac et des élèves admis en première année du cycle ingénieur à l'issue d'une formation bac+2 (CPGE, DEUG, DUT, BTS).

La colonne (I) indique le pourcentage de CPGE par rapport à l'ensemble des inscrits au niveau bac et ceux en première année du cycle ingénieur issus d'une formation bac+2.

La colonne (II) indique le pourcentage de CPGE par rapport à l'ensemble des inscrits en première année du cycle ingénieur issus d'une formation bac+2.

Sources : *Tableaux Statistiques*, n° 5134, 6036, 6436, 6787, "les écoles d'ingénieurs, effectifs des élèves".

*Tableau 21*  
**Effectifs des formations bac+2 scientifiques en France.**

Effectifs des formations scientifiques bac+2 en France							
		BTS	DUT	DEUG	CPI	CPGE	TOTAL
2001	Effectifs	80 730	63 228	115 980	5 855	43 437	309 230
	%	26,1 %	20,4 %	37,5 %	1,9 %	14,0 %	100,0 %

Sources : Direction de la Programmation et du Développement, *Repères et références statistiques*, édition 2001 ;  
*Tableaux Statistiques*, n°6787, "Les écoles d'ingénieurs, effectifs des élèves en 2000-2001", juin 2001 ;  
*Tableaux Statistiques*, n°6746, "Classes préparatoires aux grandes écoles 2000-2001", décembre 2000 ;  
*Tableaux Statistiques*, n°6784, "Effectifs dans les instituts universitaires de technologie 2000-2001", mai 2001 ;  
*Tableaux Statistiques*, n°6749, "Sections de techniciens supérieurs, préparations diverses post-baccalauréat 2000-2001", janvier 2001.

Pour des raisons de comparabilité, les chiffres ci-dessus concernent la France métropolitaine.

Pour le DEUG, addition des effectifs de premier cycle en Sciences et structure de la matière, Sciences et technologie (sciences pour l'ingénieur), Sciences de la nature et de la vie.

Pour les CPI, chiffres tels qu'ils sont fournis par la DEP (cycle préparatoire dans le Tableau Statistique 6787).

Pour les CPGE, chiffres tels qu'ils sont fournis par la DEP (TS 6746)

Pour les DUT, addition des départements du secteur secondaire auxquels ont été ajoutés ceux d'informatique, de services et réseaux de communication ainsi que statistique et traitement informatique des données (TS 6784).

Pour les BTS, ensemble des BTS en deux ans moins certains domaines tels que services aux personnes sauf analyses biologiques, secrétariat bureautique, journalisme communication, échange et gestion, habillement, cuirs et peaux (TS 6749).

- **Annexe 3**

**Classement des écoles les plus rémunératrices (étude Insee) :**

Polytechnique  
Mines de Paris  
Centrale Paris  
Ponts et chaussées  
ENSTA Paris  
Mines de Saint-Étienne  
Sup Aéro  
ENAC Toulouse  
ENSCMu, Mulhouse  
INA-PG, Paris  
ECPM Strasbourg (ex EHICS)  
Supelec  
ECAM Lyon  
EC Lille  
ENSIC Nancy  
ECL Lyon  
ENSCP, Paris  
ENSCL ; Lille  
ESIEA, Paris, Laval

- **Annexe 4**

*Tableau 22*  
**Recrutement hors classes préparatoires des écoles les plus rémunératrices.**

Recrutement hors classes préparatoires des écoles les plus rémunératrices							EFFECTIFS	
	STATUT	EU	Écoles	BTS	DUT	DEUG	BAC+3	TOTAL
1	DEFENSE		Polytechnique	-	-	-	399	1 345
2	INDUSTRIE		Mines de Paris	-	-	-	91	321
3	EDUCATION		Centrale Paris	-	-	-	443	1 301
4	EQUIPEMENT		Ponts et chaussées	-	-	-	105	629
5	DEFENSE		ENSTA Paris	-	-	-	96	453
6	INDUSTRIE		Mines de Saint-Étienne	-	-	-	102	339
7	DEFENSE		Sup Aéro	-	-	-	119	555
8	EQUIPEMENT		ENAC Toulouse	-	-	A	92	304
9	EDUCATION	EU	ENSCMu, Mulhouse	A	A	A	66	229
10	AGRICULTURE		INA-PG, Paris	A	A	A	219	748
11	EDUCATION	EU	ECPM Strasbourg (ex EHICS)	A	A	A	78	270
12	PRIVEE		Supelec	-	A	A	328	1 081
13	PRIVEE		ECAM Lyon	-	-	-	110	306
14	EDUCATION		EC Lille	-	-	-	274	854
15	EDUCATION	EU	ENSIC Nancy	-	A	A	99	339
16	EDUCATION		ECL Lyon	-	-	-	315	1 010
17	EDUCATION	EU	ENSCP, Paris	A	A	A	66	221
18	EDUCATION	EU	ENSCL, Lille	A	A	A	78	233
19	PRIVEE		ESIEA, Paris, Laval	A	A	A	304	1 345
<b>TOTAL Hors Groupe</b>							<b>3 384</b>	<b>11 883</b>

Note : L'indication "STATUT" renvoie à la classification des écoles en trois groupes telle qu'elle est opérée par la DPD. Sont ainsi distinguées les écoles dépendant du ministère de l'Éducation nationale, celles dépendant d'autres ministères et les écoles privées. L'indication "EU" désigne les écoles universitaires. "A" indique les formations qui sont admises et "M" les admissions au niveau des titulaires d'une maîtrise. "BAC+3" indique le nombre d'élèves au niveau bac+3 lequel correspond à la première année du cycle ingénieur au sens "classique". "TOTAL" indique l'effectif total de l'école.

Source : DPD, *Tableaux Statistiques*, n°6787, "les effectifs des écoles d'ingénieurs en 2000/2001", juin 2001.  
ONISEP, "Écoles d'ingénieurs, niveaux d'admission", in *Après le bac... réussir ses études*, 2000.

Il importe ici de nuancer l'image d'un recrutement "exclusivement prépa" donnée par le guide de l'ONISEP par celle d'un recrutement "quasi exclusivement prépa". Ainsi l'école polytechnique référencée parmi les écoles ne recrutant que des prépa admet aussi quelques élèves de DEUG.

- **Annexe 5**

*Tableau 23*  
**Effectifs des écoles universitaires dans l'échantillon et au niveau national.**

	Échantillon		NATIONAL	
	N	%	N	%
<b>Formations universitaires</b>	519	-	11 445	-
<b>Écoles dépendant du MEN</b>	689	75,3 %	18 002	63,6 %
<b>Ensemble des écoles</b>	972	53,4 %	31 048	36,9 %

Lecture : les effectifs des formations universitaires s'élèvent à 519 ingénieurs sur un total de 972 pour les écoles de l'échantillon. Ce qui correspond à 53,4 % de ce total. Par ailleurs les effectifs des formations universitaires représentent 75,3 % des écoles publiques dépendant du ministère de l'Éducation nationale.

Source niveau national : DPD, *Tableaux Statistiques*, n°6787, "les effectifs des écoles d'ingénieurs en 2000/2001", juin 2001.