

*La Note d'Analyse n°120,*  
« **Inégalité des chances : ce qui compte le plus<sup>1</sup>** »

Avril 2023

## **ANNEXE**

### **1. Construction des données de l'enquête *Formation et qualification professionnelle (FQP) 2014-2015***

Nous restreignons l'échantillon d'analyse aux personnes nées en France métropolitaine ayant entre 31 et 46 ans au 31 décembre de l'année de l'enquête, soit environ 8 000 individus enquêtés nés entre 1967 et 1983.

#### ***La construction de l'origine sociale***

L'origine sociale est construite en combinant la catégorie socioprofessionnelle du père (CS\_PERE) et celle de la mère (CS\_MERE), sur la base de la « PCS Ménage » proposée en 2019 par le groupe de travail du Centre national de l'information statistique (Cnis) consacré à la rénovation de la PCS de l'Insee.

On retient le premier chiffre de la variable CSP qui permet d'identifier les groupes socioprofessionnels usuels. Les chefs d'entreprise de dix salariés ou plus (CS : « 23 » ) sont regroupés avec les cadres et professions intellectuelles supérieures.

Au total, on obtient cinq groupes socioprofessionnels :

1. agriculteurs exploitants, artisans et commerçants, chefs d'entreprise ;
2. cadres et professions intellectuelles supérieures (y compris chefs d'entreprise de dix salariés ou plus) ;
3. professions intermédiaires ;
4. employés ;
5. ouvrier

---

<sup>1</sup> Dherbécourt C. et Flamand J. (2023), « **Inégalité des chances : ce qui compte le plus** », *La Note d'analyse*, n° 120, France Stratégie, avril.

En plus de ces cinq groupes, on distingue les inactifs et les parents « séparés », c'est-à-dire ne résidant pas avec leur enfant :

Un parent est considéré comme inactif s'il n'est ni en emploi, ni au chômage, ni à la retraite ou s'il n'a jamais travaillé au moment où l'enquêté a « cessé de fréquenter régulièrement l'école ou l'université », ou à ses 16 ans s'il n'est jamais allé à l'école<sup>2</sup>. Ces concepts sont déclaratifs et se distinguent de ceux du Bureau international du travail (BIT) tels que mesurés dans l'enquête *Emploi*.

L'enquête FQP ne permet pas d'identifier systématiquement les individus qui ne résidaient pas avec leurs deux parents. Par conséquent, la proportion de pères identifiés comme « séparés » est beaucoup plus faible dans FQP (3 % essentiellement des pères inconnus) que dans l'enquête EDP (12 %). Afin d'obtenir une part de famille monoparentale comparable dans l'enquête EDP et dans l'enquête FQP, nous considérons comme absents une partie des pères et des mères divorcés. Conformément à ce que l'on observe dans notre échantillon issu de l'enquête EDP, on considère qu'en cas de divorce dans l'enquête FQP, 92 % des pères sont absents. Afin d'obtenir la même proportion de familles monoparentales dans notre échantillon issu de l'enquête FQP que dans notre échantillon issu de l'enquête EDP, on considère également que, dans un tiers des cas de divorces, il y a remise en couple. Dans ce cas, on attribue au nouveau conjoint la même PCS que celle de l'ancien conjoint.

### **La construction de l'ascendance migratoire**

On considère les personnes nées en France métropolitaine ayant au moins un parent immigré. Lorsque les deux parents sont immigrés, on retient pour déterminer l'origine migratoire de l'enquêté le pays de naissance du père. Lorsqu'un seul des parents est immigré, c'est le pays de naissance de ce parent qui sert à déterminer l'origine migratoire de l'enquêté (PAYSNAISR\_PER ou MER). On distingue des personnes sans ascendance migratoire les descendants d'immigrés d'Afrique (y compris Maghreb) et les autres (descendants d'immigrés européens, asiatiques et américains).

## **2. Écarts bruts versus écarts nets**

Les écarts bruts sont obtenus en régressant le revenu d'activité sur l'une des cinq dimensions étudiées (origine sociale, type de territoire d'origine, région d'origine, ascendance migratoire et sexe), à âge donné (modèles 1 à 5). Les écarts nets sont obtenus en régressant le revenu d'activité sur les cinq dimensions en même temps (modèle 6).

---

<sup>2</sup> On demande à l'enquêté quelle est la situation professionnelle de ses parents (ou tuteurs ou parents adoptifs) s'il est encore en formation initiale, sinon à l'époque où il a cessé de fréquenter régulièrement l'école ou l'université (s'il n'est plus en formation initiale), sinon quand il avait 16 ans (s'il n'a pas fréquenté l'école).

**Tableau 1 – Résultats des estimations des écarts de revenus d'activité bruts et nets entre catégories**

		Variable dépendante: Revenu d'activité individuel						
		Écarts bruts				Écarts nets		
		(1)	(2)	(3)	(4)	(5)	(6)	
Origine sociale (ref:favorisée)	intermédiaire	-834*** (18)					-770*** (18)	
	défavorisée	-1200*** (21)					-1080*** (21)	
Type de territoire d'origine (ref:rural)	périurbain		194*** (27)				104*** (27)	
		urbain non ZUS		282*** (22)				142*** (22)
	urbain ZUS			-233*** (32)				-104*** (33)
					397*** (19)			289*** (20)
Région d'origine (ref: autre)	Ile-de-France							
	Nord Pas-de-Calais Languedoc-Roussillon			-198*** (23)			-151*** (22)	
Ascendance migratoire (ref: autre)	sans ascendance				35 (25)		6*** (25)	
	ascendance Afrique Maghreb				-394*** (36)		-175*** (36)	
Sexe (ref: femmes)	Hommes					622*** (14)	616*** (14)	
Année de naissance		-41*** (2)	-40*** (2)	-39*** (2)	-39*** (2)	-38*** (2)	-42*** (1)	
Constante		84846***	81942***	79082***	78365***	77446***	84632***	
		-2975	-3022	-3014	-3019	-2996	-295	
Observations		109 245	109 245	109 245	109 245	109 245	109 245	
R <sup>2</sup>		0,05	0,01	0,01	0,01	0,03	0,08	

Champ : individus nés entre 1972 et 1987, ayant été recensés entre 10 et 18 ans en 1990 ou 1999 en France métropolitaine, résidant en France métropolitaine en 2018 et ayant rempli une déclaration d'impôt sur leurs revenus de 2018.

Source : France Stratégie, à partir de l'Échantillon démographique permanent (EDP) – millésime 2019

### 3. Revenu d'activité par origine sociale détaillée

Le Tableau 2 présente le revenu d'activité moyen par origine sociale détaillée, à type de territoire d'origine, région d'origine, ascendance migratoire et âge donnés.

**Tableau 2 – Revenu d'activité moyen à l'âge adulte par origine sociale détaillée**

Origine sociale	PCS ménage des parents	Part dans total	Coefficient de régression	Revenu moyen « net »
favorisée	1a- Cadre avec cadre	3,0 %	ref.	3 481 €
	1b- Cadre avec profession intermédiaire	4,4 %	- 554	2 927 €
	2a- Cadre avec employé ou ouvrier	3,5 %	- 792	2 689 €
	2b- Cadre avec inactif ou sans conjoint	5,3 %	- 444	3 037 €
	2c- Profession intermédiaire ou cadre avec petit indépendant	2,1 %	- 955	2 526 €
	2d- Profession intermédiaire avec profession intermédiaire	4,2 %	- 949	2 532 €
intermédiaire	3a- Profession intermédiaire avec employé ou ouvrier	10,9 %	- 1 214	2 267 €
	3b- Profession intermédiaire avec inactif ou sans conjoint	6,4 %	- 1 254	2 227 €
	3c- employé avec employé	4,2 %	- 1 434	2 047 €
	4a- Petit indépendant avec petit indépendant, avec inactif	7,2 %	- 1 345	2 136 €
	4b- Petit indépendant avec employé ou ouvrier	4,1 %	- 1 288	2 193 €
	5a- Ouvrier avec employé	13,6 %	- 1 507	1 974 €
	5b- Ouvrier avec ouvrier	6,2 %	- 1 628	1 853 €
modeste	6a- Employé avec inactif ou sans conjoint	8,2 %	- 1 584	1 897 €
	6b- Ouvrier avec inactif ou sans conjoint	16,8 %	- 1 759	1 722 €

Note : le revenu moyen « net » est calculé à partir du revenu moyen de la catégorie 1a (deux parents cadres), soit 3 470 euros, auquel on ajoute l'écart de revenus net estimé pour chaque catégorie (à âge, type de territoire d'origine, région d'origine et ascendance migratoire donnés).

Champ : individus nés entre 1972 et 1987, ayant été recensés entre 10 et 18 ans en 1990 ou 1999 en France métropolitaine, résidant en France métropolitaine en 2018 et ayant rempli une déclaration d'impôt sur leurs revenus de 2018.

Source : France Stratégie, à partir de l'Échantillon démographique permanent (EDP) – millésime 2019

## 4. Méthode de décomposition des écarts

L'analyse de médiation consiste à mesurer la part des écarts entre catégories qui transitent par des variables « intermédiaires » ou variables de médiation.

Dans la note d'analyse, nous décomposons les écarts de revenus entre un écart lié au parcours éducatif, un écart résiduel de position sur le marché du travail (à parcours éducatif donné) et un écart inexpliqué.

Cette analyse de médiation se déroule en deux étapes.

### Première étape :

La première étape consiste à estimer des équations de revenu.

On estime d'abord les écarts de revenus entre catégories d'individus, au moyen d'une régression linéaire :

$$(1) \quad Y = a + \sum b_i.CATEGORIE_i + c1.CONTROLES$$

Où  $Y$  correspond au revenu d'activité déclaré par l'individu (salaires et revenus d'indépendant),  $CATEGORIE_i$  aux catégories d'individu et  $CONTRÔLES$  aux variables de contrôle.

Pour l'estimation des équations, la population est divisée en deux « catégories » lorsqu'on s'intéresse aux écarts entre sexe (hommes/femmes), et en trois catégories lorsqu'on étudie les écarts entre origines sociales (origine favorisée/intermédiaire/modeste). Pour les analyses des écarts entre catégories croisant plusieurs critères, le nombre de catégories s'élève. La population est divisée en six catégories lorsqu'on croise sexe et origine sociale, en neuf lorsqu'on croise origine sociale et région de naissance (Île-de-France/Nord-Pas-de-Calais/Languedoc-Roussillon/autres régions) ou origine sociale et ascendance migratoire (sans ascendance migratoire/ascendance migratoire africaine/autre ascendance migratoire).

Les variables de contrôle sont incluses dans les régressions afin d'obtenir des écarts « nets » entre catégories, selon la définition retenue dans la note d'analyse<sup>3</sup>. Le Tableau 3 précise les variables de contrôle incluses dans les régressions selon les catégories comparées.

<sup>3</sup> Les écarts de revenus sont « nets » au sens où ils sont mesurés à « caractéristiques comparables ». Par exemple, on mesure les écarts entre catégories d'origine sociale entre individus de sexe, d'origine migratoire et d'origine régionale comparables.

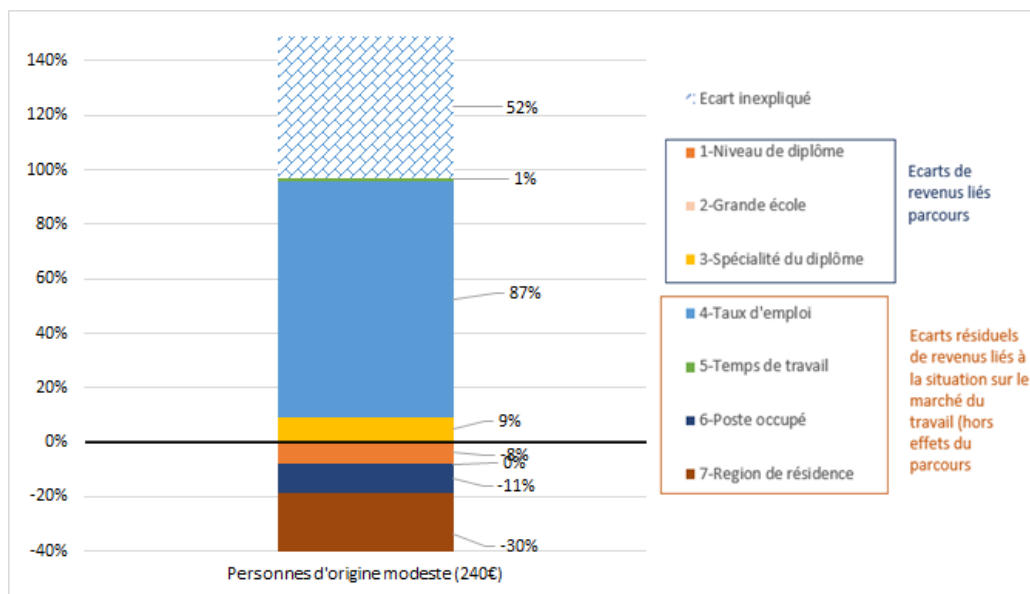
**Tableau 3 – Variables de contrôle utilisées, selon les catégories comparées**

Individus regroupés par...	Catégories comparées	Variables de contrôle incluses dans les régressions	N° de graphique dans la NA
<b>Sexe</b>	Hommes <b>vs</b> Femmes	Origine sociale, ascendance migratoire, région de naissance, âge	5
	Hommes origine favorisée <b>vs</b> Femmes origine favorisée	Ascendance migratoire, région de naissance, âge	
	H origine intermédiaire <b>vs</b> F origine intermédiaire		
	H origine modeste <b>vs</b> F origine modeste		
<b>Origine sociale</b>	Personnes d'origine favorisée <b>vs</b> Personnes d'origine modeste	Sexe, ascendance migratoire, région de naissance, âge	6
	Femmes d'origine favorisée <b>vs</b> Femmes d'origine modeste	Ascendance migratoire, région de naissance, âge	
	Hommes d'origine favorisée <b>vs</b> Hommes d'origine modeste		
<b>Région de naissance</b>	Origine sociale favorisée natifs IDF <b>vs</b> Origine sociale favorisée natifs NPDC/LR	Sexe, ascendance migratoire, âge	2 (annexe)
	Origine sociale modeste natifs IDF <b>vs</b> Origine sociale modeste natifs NPDC/LR		
<b>Ascendance migratoire</b>	Origine sociale modeste avec ascendance africaine <b>vs</b> Origine sociale modeste sans ascendance migratoire	Sexe, région de naissance, âge	1 (annexe)

Lecture : lorsqu'on décompose les écarts entre hommes et femmes, on inclut les variables suivantes comme contrôles dans les équations de régression : l'origine sociale, l'ascendance migratoire, la région de naissance et l'âge des individus.

Les Graphiques 1 et 2 présentent les décompositions des écarts de revenu liés à l'ascendance migratoire et à la région de naissance, évoqués dans l'Encadré 3 de la note d'analyse.

**Graphique 1 – Décomposition des écarts de revenus selon l'ascendance migratoire**



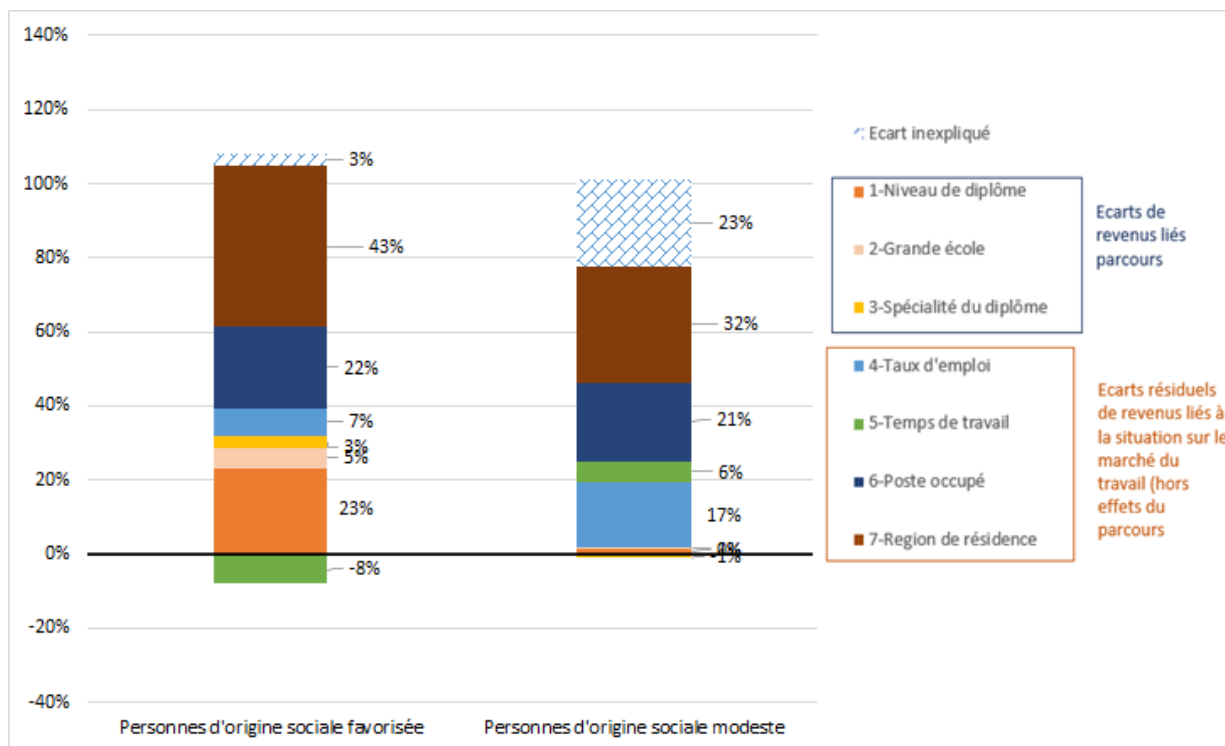
Note : sans ascendance migratoire *versus* enfant d'au moins un parent immigré d'Afrique. Faute d'un nombre suffisant d'observations, l'effet n'est pas distingué selon le sexe.

Champ : individus nés et résidant en France métropolitaine, âgés de 32 à 46 ans en 2014.

Lecture : parmi les personnes d'origine modeste, 87 % de l'écart de revenu d'activité entre personnes sans ascendance migratoire et descendants d'immigrés d'Afrique, est lié au fait que les premières ont un taux d'emploi plus élevé que les seconds.

Source : France Stratégie, à partir de l'enquête Formation et qualification professionnelle (FQP) 2014-2015

**Graphique 2 – Décomposition des écarts de revenus selon la région d'origine, à origine sociale et migratoire données**



Note : faute d'un nombre suffisant d'observations, l'effet n'est pas distingué selon le sexe.

Champ : individus nés et résidant en France métropolitaine, âgés de 32 à 46 ans en 2014.

Lecture : parmi les personnes d'origine favorisée, 221 % de l'écart de revenu d'activité entre les personnes nées en Île-de-France et dans les anciennes régions Nord-Pas-de-Calais/Languedoc-Roussillon est lié au fait que les premières occupent des postes mieux rémunérés que les secondes.

Source : France Stratégie, à partir de l'enquête Formation et qualification professionnelle (FQP) 2014-2015

On estime ensuite sept équations en ajoutant successivement les variables de médiation.

On commence par ajouter les variables concernant le parcours éducatif :

$$(2) \quad Y = a + \sum bi.CATEGORIE_i + c1.CONTROLES + d1.DIPLOME + d2.GRANDE_ECOLE$$

$$(3) \quad Y = a + \sum bi.CATEGORIE_i + c1.CONTROLES + d1.DIPLOME + d2.GRANDE_ECOLE$$

$$(4) \quad Y = a + \sum bi.CATEGORIE_i + c1.CONTROLES + d1.DIPLOME + d2.GRANDE_ECOLE + d3.SPECIALITE$$

Avec GRANDE\_ECOLE, une variable binaire valant 1 lorsque l'individu a fréquenté une grande école (construite à partir des variables POSTBAC et PREPA dans l'enquête) ; diplôme, le plus haut diplôme obtenu (variable DIPL) ; et SPECIALITE, pour spécialité du diplôme (construite en croisant les variables DIPL et SPE).

On ajoute ensuite les variables de situation sur le marché du travail :

$$(5) \quad Y = a + \sum bi.CATEGORIE_i + c1.CONTROLES +$$



$$\begin{aligned}
 & d1.DIPLOME + d2.GRANDE\_ECOLE + d3.SPECIALITE + \\
 & e1.EMPLOI \\
 (6) \quad Y = a + \sum bi.CATEGORIE\_i + c1.CONTROLES + \\
 & d1.DIPLOME + d2.GRANDE\_ECOLE + d3.SPECIALITE + \\
 & e1.EMPLOI + e2.TEMPS\_TRAVAIL \\
 (7) \quad Y = a + \sum bi.CATEGORIE\_i + c1.CONTROLES + \\
 & d1.DIPLOME + d2.GRANDE\_ECOLE + d3.SPECIALITE + \\
 & e1.EMPLOI + e2.TEMPS\_TRAVAIL + e3.POSTE
 \end{aligned}$$

Avec EMPLOI, l'exercice ou non d'un emploi (construit à partir de la variable ACTEU) ; TEMPS\_TRAVAIL, l'exercice d'un emploi à temps plein ou partiel selon la quotité travaillée (variable TXTPRED) ; et POSTE, le type de poste occupé. Le POSTE est lui-même une combinaison du domaine professionnel (variable P agrégée au domaine professionnel)<sup>4</sup>, de la catégorie socioprofessionnelle (CSTOT), de l'appartenance au secteur privé ou public (CHPUB et STCT), de l'exercice d'une tâche d'encadrement (ENCADR), de la taille de l'entreprise (EFFTR), du secteur d'activité (NAFG010UN).

On ajoute enfin la variable de région de résidence (variable REGION)

$$\begin{aligned}
 (8) \quad Y = a + \sum bi.CATEGORIE\_i + c1.CONTROLES + \\
 d1.DIPLOME + d2.GRANDE\_ECOLE + d3.SPECIALITE + \\
 e1.EMPLOI + e2.TEMPS\_TRAVAIL + e3.POSTE + \\
 f1.REGION\_RESIDENCE
 \end{aligned}$$

Toutes les équations, quelles que soient les catégories comparées, sont estimées sur l'ensemble de l'échantillon de 7 377 individus observés dans FQP, âgés de 31 à 46 ans au 31 décembre 2014, pour lesquels toutes les variables de l'équation 8 sont observées.

Le Tableau 4 présente les coefficients de régression obtenus lorsque les individus sont distingués selon le sexe et l'origine sociale par exemple.

---

<sup>4</sup> Les professions sont regroupées par domaines professionnels, selon la nomenclature de la Dares.

**Tableau 4 – Coefficients de régression associés aux catégories d'individus croisant sexe et origine sociale**

		<i>Variable dépendante: Revenu d'activité mensuel</i>							
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
<b>Femmes d'origine favorisée</b>		<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>	<i>ref.</i>
<b>Femmes d'origine intermédiaire</b>		-431,77*** (52,72)	-92,29* (50,38)	-70,61 (50,08)	-58,43 (49,34)	-44,88 (47,87)	-45,34 (47,05)	40,67 (43,43)	51,5 (43,37)
<b>Femmes d'origine modeste</b>		-658,71*** (61,16)	-163,17*** (58,99)	-137,36** (58,62)	-140,33** (57,87)	-108,05* (56,14)	-90,99* (55,19)	48,4 (51,20)	53,28 (51,06)
<b>Hommes d'origine favorisée</b>		833,61*** (62,88)	919,68*** (58,73)	896,63*** (58,35)	809,57*** (58,32)	799,66*** (56,59)	688,24*** (56,07)	612,01*** (52,57)	616,01*** (52,41)
<b>Hommes d'origine intermédiaire</b>		68,49 (52,98)	550,11*** (51,58)	565,71*** (51,25)	521,09*** (51,44)	498,02*** (49,94)	359,85*** (49,86)	385,94*** (47,44)	400,84*** (47,38)
<b>Hommes d'origine modeste</b>		-110,86* (60,91)	502,48*** (59,76)	520,89*** (59,35)	477,41*** (59,17)	461,19*** (57,46)	317,96*** (57,18)	324,67*** (54,21)	335,04*** (54,12)
<b>Variables de contrôle</b>	<b>Âge</b>	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
	<b>Région de naissance</b>	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
	<b>Origine migratoire</b>	Oui	Oui	Oui	Oui	Oui	Oui	Oui	Oui
	<b>Diplôme</b>	Non	Oui	Oui	Oui	Oui	Oui	Oui	Oui
	<b>Grande école</b>	Non	Non	Oui	Oui	Oui	Oui	Oui	Oui
	<b>Spécialité</b>	Non	Non	Non	Oui	Oui	Oui	Oui	Oui
	<b>Emploi</b>	Non	Non	Non	Non	Oui	Oui	Oui	Oui
	<b>Quotité de travail</b>	Non	Non	Non	Non	Non	Oui	Oui	Oui
	<b>Poste</b>	Non	Non	Non	Non	Non	Non	Oui	Oui
	<b>Région de résidence</b>	Non	Non	Non	Non	Non	Non	Non	Oui
<b>Observations</b>		7377	7377	7377	7377	7377	7377	7377	7377
<b>R<sup>2</sup></b>		0,12	0,15	0,25	0,27	0,32	0,34	0,45	0,46

Note : chaque colonne présente les coefficients des régressions (1) à (8) associés à chaque catégorie, en écart à la catégorie de référence (les femmes d'origine favorisée). Seuil de significativité des coefficients : \*p<0.1 ; \*\*p<0.05 ; \*\*\*p<0.01.

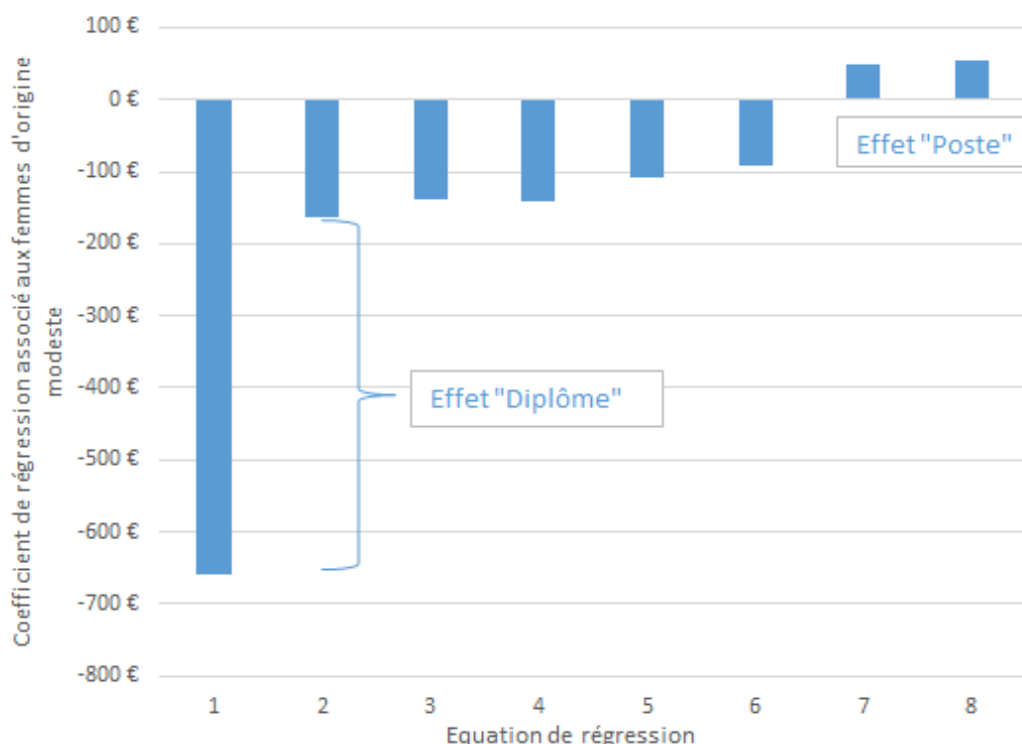
Champ : individus nés et résidant en France métropolitaine, âgés de 32 à 46 ans en 2014.

Source : France Stratégie, à partir de l'enquête Formation et qualification professionnelle (FQP) 2014-2015

## Seconde étape :

Dans la seconde étape, on calcule la contribution des variables de médiations aux écarts entre catégories. On utilise pour ce faire les coefficients associés à chaque catégorie  $i$  (les  $b_i$ ) obtenus lors des estimations successives. Les différents effets sont obtenus par soustraction, comme l'illustre le Graphique 3.

**Graphique 3 – Illustration de la mesure de la contribution des effets « Diplôme » et « Poste » dans les écarts entre femmes d'origine modeste et favorisée**



Champ : individus nés entre 1972 et 1987, ayant été recensés entre 10 et 18 ans en 1990 ou 1999 en France métropolitaine, résidant en France métropolitaine en 2018 et ayant rempli une déclaration d'impôt sur leurs revenus de 2018.

Lecture : l'écart entre femmes d'origine favorisée et modeste passe de 660 à 160 euros lorsqu'on ajoute la variable « Diplôme » à la régression (passage de l'équation de régression 1 à 2). On en déduit que l'effet « Diplôme » correspond à 600 euros environ.

Source : France Stratégie, à partir de l'Échantillon démographique permanent (EDP) – millésime 2019

Dans la note d'analyse, les écarts sont exprimés en pourcentage de l'écart net entre catégories, obtenu dans la première régression. Le Tableau 3 fournit les formules de calcul pour chaque effet ; on note  $b_i(j)$  le coefficient associé à la catégorie d'individu  $i$  dans l'équation de revenu  $(j)$ . L'effet résiduel correspond au coefficient  $b_i(8)$ , soit le coefficient obtenu pour la catégorie  $i$  dans la régression 8, lorsque toutes les variables ont été incluses.

**Tableau 5 – Formules de calcul de la décomposition de l'écart de revenus entre la catégorie « i » et la catégorie de référence**

	Effet	Formule de calcul
Ecart lié au parcours éducatif	Accès aux grandes écoles	$\frac{bi(1) - bi(2)}{bi(1)}$
	Plus haut diplôme obtenu	$\frac{bi(2) - bi(3)}{bi(1)}$
	Spécialité du diplôme	$\frac{bi(3) - bi(4)}{bi(1)}$
Ecart résiduel lié au marché du travail ou à la région de résidence	Taux d'emploi	$\frac{bi(4) - bi(5)}{bi(1)}$
	Temps de travail	$\frac{bi(5) - bi(6)}{bi(1)}$
	Poste occupé	$\frac{bi(6) - bi(7)}{bi(1)}$
	Région de résidence	$\frac{bi(7) - bi(8)}{bi(1)}$
Ecart inexplicé		$\frac{bi(8)}{bi(1)}$

Note : les  $bi(j)$  correspondent aux coefficients de régression associés à la catégorie « i » d'individus dans les équations (1) à (8).

### 3 – Méthode de l'*Event study* sur la naissance des enfants

Nous appliquons la méthode créée par Kleven *et al.* aux données de l'enquête EDP.

On rassemble d'abord les individus ayant déclaré des revenus en continu entre 2010 et 2018 et ayant eu leur premier enfant en 2013, 2014 et 2016. L'échantillon compte 46 000 individus.

Du fait des contraintes liées aux sources fiscales, nous ne pouvons observer que les individus âgés de 25 à 35 ans en 2010 et de 34 à 44 ans en 2018<sup>5</sup>. Les individus des générations 1975-1985 ayant eu leur premier enfant avant 25 ans ou après 35 ans sont donc exclus de l'analyse.

<sup>5</sup> Nous n'avons utilisé les données fiscales que si les individus étaient âgés de 25 ans ou plus, âge auquel le rattachement au foyer fiscal des parents n'est plus autorisé.

Par comparaison, les données de revenu utilisées par Kleven *et al.* couvrent deux décennies et permettent d'identifier l'ensemble des individus ayant eu au moins un enfant durant leur vie.

La base de données contient les observations de revenu d'activité, avec une observation par individu et par année.

On crée une variable qui mesure, pour chaque année et pour chaque individu, le nombre d'années d'écart avec la naissance du premier enfant de l'individu. Dans notre échantillon, cette variable prend des valeurs comprises entre -6 (l'année 2010, pour les individus devenus parents en 2016) et +5 (en 2018 lorsque la naissance a eu lieu en 2013).

On crée ensuite onze variables binaires  $T_n$  valant 1 si le nombre d'années d'écart est égal à  $n$  ( $n$  étant différent de -1 et compris entre -6 et 5).

On estime l'équation suivante par régression linéaire :

$$(1) \quad Y = a + \sum b_n \cdot T_n + c \cdot ANNEE + d \cdot AGE$$

L'équation est estimée séparément pour les hommes et les femmes.

Les coefficients  $b_n$  s'interprètent alors comme l'effet en euros de la naissance des enfants sur le revenu d'activité de l'année  $n$  (l'effet étant nul pour  $n=-1$ ). Afin d'obtenir des évolutions de revenus en pourcent, on divise ces coefficients  $b_n$  par le revenu moyen en  $n$  à défaut d'enfant<sup>6</sup>.

Pour l'estimation des effets selon le sexe et l'origine sociale, on se focalise sur les individus dont l'origine sociale est observable dans les recensements de 1990 et 1999, soit environ 10 000 individus sur 46 000<sup>7</sup>. On crée six sous-échantillons pour chaque catégorie (hommes et femmes d'origine favorisée, intermédiaire ou modeste). On applique la méthode présentée ci-dessous aux six sous-échantillons.

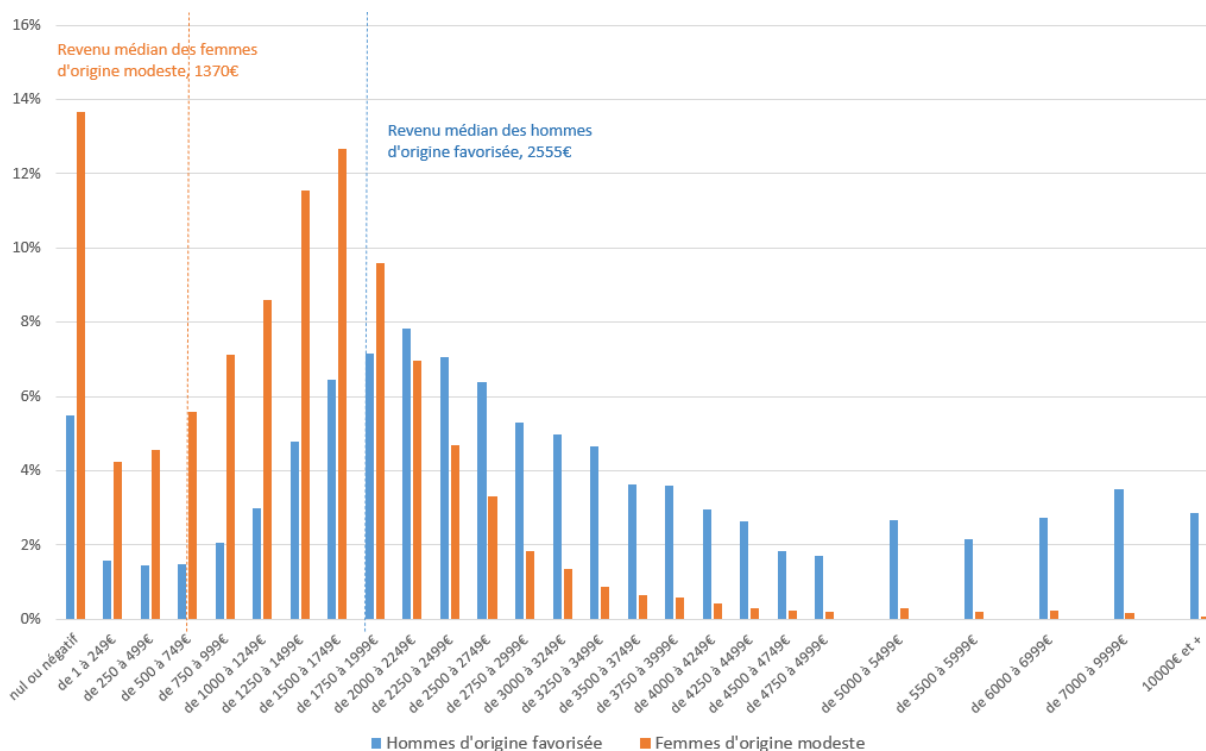
<sup>6</sup> Pour le calcul du revenu moyen en  $n$  à défaut d'enfant, on prédit d'abord le revenu d'activité à partir des coefficients obtenus dans la régression (1), en considérant  $T_n=0$  pour tout  $n$ , comme dans l'article de Kleven *et al.* On calcule ensuite la moyenne de ce revenu prédit, pour tout  $n$ .

<sup>7</sup> L'EDP contient les individus nés les quatre premiers jours de chaque trimestre. Les données du recensement ne sont disponibles que pour les individus nés les quatre premiers jours d'octobre.

## 4 – Distributions de revenus d'activité selon le sexe et l'origine sociale

Le Graphique 4 et le Tableau 6 présentent la distribution des revenus d'activité par sexe et origine sociale. Il s'agit de distribution « brutes », c'est-à-dire non redressées des différences liées à l'ascendance migratoire, à la région d'origine, au type de territoire et à l'âge.

**Graphique 4 – Répartition des hommes d'origine favorisée et des femmes d'origine modeste, par tranche de revenu d'activité**



Champ : individus nés entre 1972 et 1987, ayant été recensés entre 10 et 18 ans en 1990 ou 1999 en France Métropolitaine, résidant en France Métropolitaine en 2018 et ayant rempli une déclaration d'impôt sur leurs revenus de 2018.

Lecture : 13,7 % des femmes d'origine modeste de l'échantillon ont un revenu négatif ou nul.

Source : France Stratégie, à partir de l'Échantillon démographique permanent (EDP) – millésime 2019

**Tableau 6 – Distribution des revenus d'activité, selon le sexe et l'origine sociale**

Tranche de revenu	Hommes d'origine favorisée	Hommes d'origine intermédiaire	Hommes d'origine modeste	Femmes d'origine favorisée	Femmes d'origine intermédiaire	Femmes d'origine modeste
nul ou négatif	5,5%	4,6%	7,7%	5,7%	6,6%	13,7%
de 1 à 249€	1,6%	1,6%	2,4%	2,0%	2,6%	4,2%
de 250 à 499€	1,4%	1,9%	2,5%	1,7%	2,8%	4,6%
de 500 à 749€	1,5%	2,2%	3,1%	2,9%	3,9%	5,6%
de 750 à 999€	2,1%	3,0%	3,6%	3,6%	5,5%	7,1%
de 1000 à 1249€	3,0%	4,0%	5,4%	5,1%	7,7%	8,6%
de 1250 à 1499€	4,8%	6,8%	8,4%	6,5%	11,1%	11,5%
de 1500 à 1749€	6,5%	11,2%	12,6%	9,0%	14,0%	12,7%
de 1750 à 1999€	7,2%	11,9%	12,3%	10,2%	12,4%	9,6%
de 2000 à 2249€	7,8%	11,2%	11,2%	10,1%	9,5%	7,0%
de 2250 à 2499€	7,0%	9,1%	7,8%	8,1%	6,9%	4,7%
de 2500 à 2749€	6,4%	7,0%	5,6%	6,6%	4,8%	3,3%
de 2750 à 2999€	5,3%	5,1%	4,1%	5,1%	3,0%	1,8%
de 3000 à 3249€	5,0%	4,2%	2,8%	3,9%	2,2%	1,4%
de 3250 à 3499€	4,7%	3,1%	2,2%	3,1%	1,6%	0,9%
de 3500 à 3749€	3,6%	2,4%	1,6%	2,4%	1,1%	0,6%
de 3750 à 3999€	3,6%	2,0%	1,2%	2,0%	0,9%	0,6%
de 4000 à 4249€	3,0%	1,3%	0,9%	1,9%	0,6%	0,4%
de 4250 à 4499€	2,6%	1,2%	0,8%	1,6%	0,5%	0,3%
de 4500 à 4749€	1,8%	0,9%	0,5%	1,3%	0,4%	0,2%
de 4750 à 4999€	1,7%	0,8%	0,4%	0,9%	0,3%	0,2%
de 5000 à 5499€	2,7%	1,2%	0,7%	1,5%	0,4%	0,3%
de 5500 à 5999€	2,2%	0,8%	0,5%	1,1%	0,3%	0,2%
de 6000 à 6999€	2,7%	0,9%	0,6%	1,4%	0,4%	0,2%
de 7000 à 9999€	3,5%	1,1%	0,6%	1,5%	0,4%	0,2%
10000€ et +	2,8%	0,5%	0,4%	0,9%	0,2%	0,1%
Total	100%	100%	100%	100%	100%	100%
médiane	2 555 €	2 060 €	1 815 €	2 079 €	1 673 €	1 370 €

Champ : individus nés entre 1972 et 1987, ayant été recensés entre 10 et 18 ans en 1990 ou 1999 en France métropolitaine, résidant en France métropolitaine en 2018 et ayant rempli une déclaration d'impôt sur leurs revenus de 2018.

Lecture : le revenu médian des hommes d'origine favorisée est de 2 555 euros. 5,5 % d'entre eux ont un revenu d'activité négatif ou nul.

Source : France Stratégie, à partir de l'Échantillon démographique permanent (EDP) - millésime 2019