



HAL
open science

Essays on Contextual Determinants of Educational, Work and Family Trajectories in France

Fanny Landaud

► **To cite this version:**

Fanny Landaud. Essays on Contextual Determinants of Educational, Work and Family Trajectories in France. Economics and Finance. Université Paris sciences et lettres, 2018. English. NNT: 2018PSLEH097 . tel-03168272

HAL Id: tel-03168272

<https://theses.hal.science/tel-03168272>

Submitted on 12 Mar 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à l'Ecole des hautes études en sciences sociales

Essais sur les déterminants contextuels des trajectoires scolaires,
professionnelles et familiales en France

Ecole doctorale n°465

ECOLE DOCTORALE ECONOMIE PANTHEON SORBONNE

Spécialité Sciences Économiques

**Soutenue par Fanny LANDAUD
le 28 Septembre 2018**

Dirigée par **Éric MAURIN**

COMPOSITION DU JURY :

M. PISTOLESI Nicolas
Professeur à TSE et à l'Université
Toulouse 1 Capitole, Rapporteur

Mme. SOLAZ Anne
Directrice de recherche à l'INED,
Rapporteur

Mme. BACACHE Maya
Professeur à Télécom ParisTech,
Membre du jury

Mme. FACK Gabrielle
Professeur à l'Université Paris Dauphine,
Membre du jury

M. MAURIN Éric
Directeur des études à l'EHESS, PSE,
Membre du jury

ÉCOLE DES HAUTES ÉTUDES EN
SCIENCES SOCIALES



PARIS SCIENCES ET LETTRES



ÉCOLE D'ÉCONOMIE DE PARIS



ÉCOLE DOCTORALE : ED 465 – Économie Panthéon Sorbonne

THÈSE

Pour l'obtention du grade de docteur en Sciences Économiques
de l'École des Hautes Études en Sciences Sociales

Présentée et soutenue publiquement le 28 Septembre 2018 par

Fanny LANDAUD

ESSAIS SUR LES DÉTERMINANTS CONTEXTUELS DES TRAJECTOIRES SCOLAIRES, PROFESSIONNELLES ET FAMILIALES EN FRANCE

Sous la direction d'Éric MAURIN

Composition du jury :

Rapporteurs

Nicolas PISTOLESI Professeur à TSE et à l'Université Toulouse I Capitole

Anne SOLAZ Directrice de recherche à l'INED

Examineurs

Maya BACACHE Professeur à Télécom ParisTech

Gabrielle FACK Professeur à l'Université Paris Dauphine

Directeur

Éric MAURIN Directeur d'études à l'EHESS, PSE

ÉCOLE DES HAUTES ÉTUDES EN
SCIENCES SOCIALES



PARIS SCIENCES ET LETTRES



PARIS SCHOOL OF ECONOMICS



ÉCOLE DOCTORALE: ED 465 – Économie Panthéon Sorbonne

PHD THESIS

Submitted to École des Hautes Études en Sciences Sociales
For the Degree of Doctor of Philosophy in Economics

Prepared and defended on September 28, 2018 by

Fanny LANDAUD

ESSAYS ON THE CONTEXTUAL DETERMINANTS OF
EDUCATIONAL, WORK AND FAMILY TRAJECTORIES IN
FRANCE

Thesis Advisor: Éric MAURIN

Jury :

Reviewers

Nicolas PISTOLESI Professor at TSE and University Toulouse I Capitole
Anne SOLAZ Professor at INED

Examinators

Maya BACACHE Professor at Télécom ParisTech
Gabrielle FACK Professor at University Paris Dauphine

Advisor

Éric MAURIN Professor at EHESS and PSE

Remerciements

Je tiens tout d'abord à remercier Éric Maurin qui a été un directeur de thèse exceptionnel. Éric a commencé à m'encadrer il y a maintenant cinq ans, pour mon mémoire de master. Pendant toutes ces années, Éric a été d'une disponibilité incroyable. Que ce soit pour nos collaborations, pour d'autres projets de recherche ou pour la recherche d'un post-doc, il s'est toujours montré très à l'écoute et ses conseils m'ont été d'une grande aide. J'ai énormément appris à ses côtés, et je lui en suis infiniment reconnaissante.

Je tiens également à remercier Gabrielle Fack qui a accepté de faire partie de mon comité de thèse et de mon jury, et m'a donné de nombreux conseils tout au long de ces dernières années. Je remercie aussi chaleureusement les rapporteurs de ma thèse, Anne Solaz et Nicolas Pistolesi, ainsi que Maya Bacache pour avoir accepté de consacrer du temps à mes travaux.

Cette thèse a été écrite à partir de nombreuses sources de données différentes, et je tiens à remercier les différentes institutions et personnes qui m'ont permis ou facilité l'accès à ces données. En particulier, je remercie Caroline Simonis-Sueur de la Direction des Études, de la Performance et de la Prospective du ministère de l'Éducation nationale et Daniel Lejeune du service statistique du concours Mines-Ponts. Je remercie également Isabelle Catto, Pascal Combemale et Sophie Giovachini pour m'avoir permis de collecter de nombreuses informations dans le cadre de ma mission complémentaire de recherche au sein de PSL, et Astrid Aulanier et Yagan Hazard qui m'ont assistée dans cette collecte. J'en suis aussi reconnaissante à Pascal Combemale pour m'avoir initiée et donné goût aux questionnements économiques.

J'ai bénéficié du soutien de très nombreuses personnes pour écrire cette thèse, et je tiens à leur exprimer ma gratitude. Je remercie l'ensemble des membres de l'École d'Économie de Paris pour m'avoir permis de travailler dans les meilleures conditions possibles. En particulier, Béatrice Havet et Véronique Guillotin ont toujours su répondre à mes nombreuses questions administratives, et Radja Aroquiardja a toujours été disponible pour mes questions informatiques. Mes travaux de recherche ont également grandement profité de multiples interactions avec les professeurs de PSE, que ce soit en séminaire, autour d'une machine à café, ou d'un piano à Aussois.

Enfin, l'ensemble des doctorants m'ont beaucoup apporté, tant sur le plan professionnel qu'amical. Je tiens en particulier à remercier Malka Guillot et les membres du bureau A1

devenu R3-13 : Anthony Lepinteur, Clémentine Van Effenterre, Mikkel Høst Gandil, Marion Leroutier, Margaux Vinez, Paul Brandily, Paul Dutronc, Pauline Grégoire-Marchand, Simon Briole et Son-Thierry Ly.

Cette thèse doit également beaucoup aux amis et famille qui m'ont entourée à chaque instant. En particulier, Andréa, Charlotte, Chatoun, Corentin et Édith. Mes parents, Philippe et Sophie ; mes sœurs, Adèle et Léna ; et mes grands-parents, Coco et Papet.

Enfin, Arnaud, pour sa patience à mon égard et le bonheur d'être à ses côtés.

Résumé en français

Cette thèse se compose de trois essais indépendants dont le dénominateur commun est d'étudier le rôle du contexte institutionnel et de l'environnement social dans lesquels se décident les trajectoires scolaires, professionnelles et familiales.

Une moitié environ des lycées parisiens (les plus prestigieux) reçoivent chaque année davantage de demandes d'admission de la part des collégiens qu'ils n'ont de places disponibles. Le logiciel d'affectation utilisé (*Affelnet*) est contraint de définir, pour chacun de ces lycées, un niveau minimum (en termes de note moyenne obtenue en troisième) nécessaire pour l'admission. En comparant les trajectoires des derniers collégiens admis dans des lycées parisiens sélectifs avec celles des premiers recalés, nous montrons dans le premier chapitre que l'accès à un établissement scolaire sélectif n'a aucun impact sur les performances ultérieures au baccalauréat, mais un impact très négatif sur la propension des jeunes lycéennes à s'orienter dans la filière générale scientifique en fin de seconde. Dans le contexte parisien, l'accès à un lycée sélectif s'accompagne surtout d'une amélioration du niveau académique des camarades de classe, notamment en science. Nos résultats concorderaient donc avec les travaux d'économie et psychologie expérimentale suggérant que les filles seraient plus réticentes que les garçons à choisir des voies compétitives.

Le deuxième chapitre porte sur les effets des politiques de redoublement en classes préparatoires scientifiques, en distinguant les effets directs observés pour les redoublants eux-mêmes et les effets indirects induits sur les autres élèves au moment de la préparation des concours. Nous montrons que les redoublements ont des effets très largement positifs pour les redoublants eux-mêmes, mais que la présence de nombreux redoublants dans une classe a des effets plutôt négatifs pour les autres élèves, notamment quand ces redoublants sont forts académiquement. Plus une classe compte de redoublants de haut niveau académique (niveau mesuré lors de leur première participation aux concours) plus les nouveaux arrivants sont en difficulté au moment des concours, alors que les performances des nouveaux arrivants sont largement insensibles au nombre de redoublants de niveau moyen ou faible. Les redoublements ne semblent pas influencer sur les autres élèves parce qu'ils contribuent à surcharger les classes, mais parce qu'ils en modifient le profil et peuvent induire les enseignants à produire des cours d'un niveau trop ambitieux.

Le troisième chapitre s'intéresse aux conséquences pour les trajectoires familiales de l'institutionnalisation des emplois précaires et des difficultés accrues rencontrées par les

jeunes adultes pour s'insérer durablement dans l'emploi. Nous y développons une analyse de durée à partir du modèle d'Abbring et van den Berg (2003) et de deux enquêtes rétrospectives sur de larges échantillons de jeunes adultes, et montrons que l'accès à l'emploi stable constitue une étape bien plus décisive que l'accès à un emploi précaire pour la mise en couple et l'arrivée des premiers enfants. Ainsi, entre les générations nées au milieu des années cinquante et celles nées au début des années soixante-dix, la montée de la part des emplois précaires et du chômage des jeunes expliquerait environ 25% des retards observés dans l'âge de mise en couple, et 40% des retards dans l'âge au premier enfant.

Discipline : Sciences économiques

Mots clefs : Éducation, Effets de pairs, Emploi, Formation de la famille

Summary in English

This thesis is composed of three independent essays studying the role of the schooling and social environment in which individuals make their educational, work or family decisions.

The first chapter studies the impact of enrollment at a more selective Parisian high school on students' performance and choice of field of study. We compare students' educational outcomes depending on whether their 9th grade standardized score fell just above or below an admission threshold, and we find that enrollment at a more selective high school has no impact on students' performance but induces female students to turn away from scientific fields and settle for less competitive ones. Our results are consistent with lab-experiment findings on gender differences in attitude towards competition and bad grades.

The second chapter analyzes grade repetition in higher education and focuses on the spillover effects induced by grade repeaters on undergraduate freshmen. We distinguish between spillovers effects induced by higher- or lower- achieving repeaters to disentangle class size from composition effects, and we find that grade repetition generates little congestion effects but has important negative composition effects. We show that the performances of freshmen are very sensitive to the number of higher-achieving repeaters while they are not impacted by the number of lower-achieving repeaters. One potential mechanism would be distortion in teaching practices.

The last chapter studies the impact of temporary contracts and youth unemployment to explain observed delays in age at first cohabiting relationship and in age at first child. Using French data on the work and family history of large samples of young adults, this chapter provides evidence that access to permanent jobs has a much stronger impact than access to temporary jobs for family formation. According to our estimates, about 25% of the increase in age at first cohabitation and about 40% of the increase in age at first child observed during the second half of the century can be explained by the rise in unemployment and in the share of temporary jobs among young workers.

Field: Economics

Key words: Education, Peer effects, Employment, Family Formation

Table des matières

Remerciements	vii
Résumé en français	ix
Summary in English	xi
Introduction générale	1
1 Competitive Schools and the Gender Gap in the Choice of Field of Study	7
1 Introduction	8
2 Institutional Context	11
2.1 The Assignment of Middle School Students to High Schools	11
2.2 Major Field of Study and High School Exit Exams	13
2.3 Selective Undergraduate Programs	14
3 Data and Methods	15
3.1 Data	15
3.2 Cut-off Scores	16
4 Graphical Evidence	17
4.1 First-Stage : Effect of Eligibility on Enrollment	18
4.2 Field of Study at the End of 10 th Grade	19
5 Regression Results	20
5.1 First Stage Effects on School Environment	21
5.2 Major Field of Study and Performance on Exams	23
5.3 Robustness and Falsification Tests	24
6 Mechanisms	25
6.1 Attitude towards Competition	26
6.2 Rank Consideration	26
6.3 Comparative Advantages and Capacity Constraints	27
6.4 Teachers' Characteristics	29
7 Conclusion	30
Appendices	38
1.A Additional Tables and Figures	38
1.B Cut-off Scores	57
2 Les redoublants nuisent-ils aux autres étudiants ? Le cas des classes préparatoires scientifiques	59
1 Introduction	60

2	Contexte	63
2.1	Les concours	64
2.2	Les classes préparatoires	65
2.3	Classes étoilées	66
3	Données utilisées	67
3.1	Les données du concours Mines-Ponts	67
3.2	Les données du ministère	68
3.3	Identification des étudiants qui redoublent après avoir été classés . . .	68
3.4	Échantillon de travail	69
4	Les redoublants : profils et performances	70
4.1	Les effets directs du redoublement sur les redoublants	70
4.2	Le niveau académique des redoublants classés et non classés	71
5	L'influence des redoublants sur les non-redoublants	72
5.1	Effets des redoublements sur la taille des classes l'année suivante . . .	74
5.2	Effets des 5/2 sur les performances des 3/2	74
5.3	Redoublants classés vs. non classés	75
6	Conclusion	77
	Annexe	81

3	From Employment to Engagement? Stable Jobs, Temporary Jobs, and Cohabiting Relationships	85
1	Introduction	86
2	Stable Jobs, Temporary Jobs, and Cohabiting Relationships	89
2.1	Data	89
2.2	Event Study	91
2.3	Timing of Events Analysis	93
2.4	Stable Jobs, Temporary Jobs, Cohabiting Relationships and Fertility . .	98
3	Gender, Employment and Cohabiting Relationships before the 1950s	99
3.1	Data	99
3.2	Event Study	101
3.3	Timing of Events Analysis	102
4	Discussion	103
5	Conclusion	105
	Appendix	112

	Conclusion générale	135
--	----------------------------	------------

	References	138
--	-------------------	------------

Introduction générale

Avec la parution des ouvrages de Gary Becker "*Human Capital, A Theoretical and Empirical Analysis, with Special Reference to Education*" en 1964 et "*A Treatise on the Family*" en 1981, le champ de la science économique s'est progressivement élargi aux domaines de l'éducation et de la famille, jusqu'alors chasse gardée des sociologues.

Dans ces deux ouvrages, Gary Becker propose de modéliser les choix scolaires et les décisions relatives à la famille (mariage, naissance, divorce,...) comme des arbitrages purement individuels, très largement indépendants du contexte dans lequel ils sont rendus, essentiellement déterminés par les préférences et les ressources des individus décisionnaires eux-mêmes.

Depuis lors, de nombreuses recherches se sont attelées à compléter cette approche en sorte de mieux rendre compte de l'importance du contexte social dans lequel les décisions individuelles sont prises, notamment celles qui conditionnent les parcours scolaires ou les carrières familiales et professionnelles. Cette thèse s'inscrit dans cette lignée de travaux et présente trois essais dont le dénominateur commun est d'étudier le rôle du contexte institutionnel et de l'environnement social dans lesquels se décident les trajectoires individuelles.

Le premier chapitre de cette thèse étudie ainsi le rôle des camarades de classe dans les choix d'orientation des élèves de seconde générale de l'académie de Paris. Le second chapitre porte sur l'influence qu'ont les redoublants sur les autres élèves dans les classes préparatoires scientifiques, notamment au moment de la préparation des concours. Enfin, le troisième chapitre étudie la façon dont l'institutionnalisation des emplois précaires et les difficultés rencontrées par les jeunes pour s'insérer durablement dans l'emploi se répercutent dans les autres sphères de leur vie, en contribuant notamment à retarder l'âge auquel ils peuvent vivre en couple et avoir des enfants.

Pour de nombreux observateurs, il ne fait guère de doute que les élèves d'une même classe, voire d'une même école, s'influencent mutuellement, tant dans leurs comportements en classe que dans leurs choix d'orientation. En fait, il ne fait même guère de doute pour beaucoup que le mieux pour un enfant est d'avoir les meilleurs camarades de classe possibles. De nombreux parents s'attachent ainsi à protéger leurs enfants d'influences qu'ils jugent problématiques en essayant de les scolariser avec des élèves du meilleur niveau scolaire possible, que ce soit en choisissant des options réputées difficiles (comme le latin) ou en venant résider à proximité d'établissements réputés sélectifs.

Force est de constater qu'un élève est en moyenne d'autant plus performant à l'école que ces camarades le sont, ce qui semble apporter une justification aux stratégies des parents les plus inquiets. Pourtant, comme le souligne – par exemple – Joshua Angrist dans "*The*

Perils of Peer Effects", il n'est en réalité guère évident d'interpréter ce type de corrélation comme reflétant un effet proprement causal des camarades de classe, c'est-à-dire comme reflétant une véritable influence vertueuse qu'exerceraient les élèves les plus performants sur leur entourage. Les deux premiers chapitres de cette thèse utilisent les spécificités du système scolaire français – et notamment les spécificités du système d'affectation des collégiens dans les lycées Parisiens – pour revisiter cette question et essayer d'isoler l'influence réelle exercée par des camarades de classe d'un meilleur niveau académique. Paradoxalement, ces deux chapitres mettent en lumière que ces effets sont souvent bien loin de ce à quoi on les associe généralement et peuvent même être tout à fait contre-productifs. Dans le premier chapitre, nous montrons par exemple qu'être scolarisé dans un lycée plus sélectif avec de meilleurs élèves n'apporte aucun bénéfice scolaire au moment du baccalauréat, mais détourne les filles de la filière scientifique, la plus prestigieuse et recherchée. Dans le second chapitre, nous nous intéressons aux classes préparatoires scientifiques et démontrons que les redoublants les plus forts académiquement tendent à y déprimer les performances aux concours de leurs camarades de classe non-redoublants.

Le premier chapitre – coécrit avec Son-Thierry Ly et Éric Maurin et intitulé "*Competitive Schools and the Gender Gap in the Choice of Field of Study*" – s'inscrit dans deux courants de la littérature : celui analysant la sous-représentation des femmes en science et celui étudiant l'effet d'être admis dans un établissement scolaire sélectif.

S'agissant du premier courant de recherches, il s'est intéressé au rôle joué par les différences de capacités cognitives, d'aspirations ou de préférences existant entre hommes et femmes pour expliquer la sous-représentation des femmes dans les domaines scientifiques, lesquels sont également les domaines les plus rémunérateurs (voir par exemple Zafar, 2013; Stinebrickner et Stinebrickner, 2014 ou Wiswall et Zafar, 2015). Dans un autre registre, Joensen et Nielsen (2009); Carrell, Page et West (2010); Lavy et Sand (2015) ou Jackson (2012) et Eisenkopf et al. (2015) ont souligné l'importance de facteurs plus institutionnels pour expliquer le déficit de femmes en sciences. Ces recherches ont notamment étudié le rôle des enseignants et des pratiques enseignantes comme facteurs explicatifs potentiels des trajectoires des filles et des garçons. Face à ce corpus de recherches, l'apport de notre premier chapitre est d'identifier un autre déterminant de la sous-représentation des femmes en sciences, à savoir la sélectivité de l'environnement scolaire dans lequel elles sont amenées à faire leurs choix. En concentrant l'analyse sur les lycées parisiens les plus sélectifs et en comparant les trajectoires des derniers collégiens admis dans ces lycées avec celles des premiers recalés (au sens du score utilisé par le logiciel d'affectation des élèves, *Affelnet*), nous démontrons que l'accès en fin de troisième

à un lycée plus sélectif n'a aucun impact sur les performances ultérieures au baccalauréat, mais un impact très négatif sur la propension des jeunes filles à choisir la filière scientifique à l'issue de leur année de seconde (les choix d'orientation des garçons restant quant à eux totalement insensibles au lycée fréquenté). Ces résultats font écho aux travaux d'économie et de psychologie expérimentale qui suggèrent que les filles tendent à être plus rétives que les garçons à entrer en compétition avec leurs camarades (voir par exemple Niederle et Vesterlund, 2007, 2011; Croson et Gneezy, 2009; ou Buser, Niederle et Oosterbeek, 2014). Dans la mesure où l'accès à un lycée sélectif accroît surtout la compétition dans les matières scientifiques (ce que semble confirmer l'étude de l'évolution des classements des élèves dans les différentes matières après la troisième), l'aversion spécifique pour la compétition des jeunes filles pourrait expliquer leur rejet des mathématiques en fin de seconde dans les lycées les plus cotés. Au-delà de la question des filles et des sciences, ce premier chapitre contribue également à la littérature sur l'impact des établissements sélectifs. Cette littérature a principalement étudié le rôle des établissements sélectifs sur les performances scolaires (voir par exemple Pop-Eleches et Urquiola, 2013 ou Abdulkadiroğlu, Angrist et Pathak, 2014). Les spécificités du système français nous permettent d'élargir cette perspective et d'étudier les effets sur les choix d'orientation des élèves, jamais envisagés jusqu'à présent.

Le second chapitre de cette thèse – coécrit avec Éric Maurin et intitulé *"Les redoublants nuisent-ils aux autres étudiants? Le cas des classes préparatoires scientifiques"* – propose une analyse des effets des politiques de redoublement, en distinguant les effets directs observés sur les redoublants eux-mêmes et les effets indirects induits sur les autres élèves. L'une des originalités de notre analyse est de porter non pas sur l'enseignement primaire ou secondaire (comme l'immense majorité de la littérature sur le redoublement), mais sur l'une des filières de l'enseignement supérieur, à savoir celle des classes préparatoires scientifiques.

Le redoublement est une pratique largement répandue dans les premiers cycles de l'enseignement supérieur, notamment les plus sélectifs (médecine, droit, classes préparatoires, ...). Pourtant, il n'existe que peu d'éléments attestant de son efficacité dans ce contexte. À notre connaissance, il n'existe même qu'un seul article ayant abordé la question du redoublement dans l'enseignement supérieur, à partir de données obtenues d'une université Suisse (voir Tafreschi et Thiemann, 2016). Dans ce chapitre, nous rassemblons un ensemble de données administratives retraçant (pour la période 2011-2016) les résultats aux concours d'écoles d'ingénieurs obtenus par les élèves de classes préparatoires scientifiques avant et après un éventuel redoublement. Ces données permettent tout d'abord de démontrer très directement un premier résultat qui n'avait jamais été établi rigoureusement, à savoir que le redoublement de la dernière année de classe préparatoire permet un accroissement très net

des performances aux concours pour une très grande majorité des redoublants concernés (accroissement que nous quantifions précisément).

Au-delà des redoublants eux-mêmes, les redoublements sont susceptibles d'affecter les autres élèves, ne serait-ce que parce qu'ils contribuent à augmenter la taille des classes et à rendre a priori plus difficiles les conditions d'enseignement. Il n'existe à notre connaissance que deux articles ayant abordé la question des effets indirects des politiques de redoublement et tous deux portent sur l'enseignement secondaire (Lavy, Paserman et Schlosser, 2012 et Hill, 2014). Notre travail est ainsi le premier à analyser les effets induits par les redoublements sur les non-redoublants dans l'enseignement supérieur. L'un des intérêts des données disponibles est de permettre de distinguer deux types de redoublants selon le niveau académique qu'ils avaient atteint en fin de première année. Notre analyse révèle que les redoublants de niveau moyen ou faible (à l'aune de leurs résultats aux concours l'année précédant le redoublement) n'ont aucun effet sur les non-redoublants, même quand ils contribuent à augmenter très significativement la taille des classes. Les effets de congestion liés à la taille des classes semblent ainsi relativement négligeables dans le contexte des classes préparatoires. Les seuls redoublants apparaissant exercer une influence sur les autres élèves de leur classe de *Mathématiques Spéciales* sont finalement les redoublants les plus forts académiquement, ceux qui étaient admis dans une très bonne école et n'ont redoublé que dans le but d'accéder à une école plus prestigieuse encore. En outre, nous révélons que cette influence tend à être négative : plus la classe compte de redoublants de haut niveau académique, plus les non-redoublants sont en difficulté au moment des concours, notamment dans les classes dites "étoilées", où précisément se concentrent les meilleurs redoublants. Les redoublants les plus forts ne semblent ainsi générer aucun des effets d'entraînement vertueux attribués traditionnellement aux bons élèves, mais semblent en revanche induire les enseignants à produire des cours d'un niveau trop ambitieux pour la moyenne des nouveaux arrivants.

Le dernier chapitre de cette thèse s'intéresse à des effets de contexte d'une autre nature. Dans ce chapitre – intitulé "*From Employment to Engagement? Stable Jobs, Temporary Jobs, and Cohabiting Relationships*" – nous abordons en effet la question des déterminants des trajectoires professionnelles et familiales. Depuis les années soixante-dix, l'âge de la première cohabitation conjugale et l'âge au premier enfant ont nettement reculé en France (voir par exemple Prioux, 2003 et Pison, 2018). Ces évolutions coïncident avec d'importants changements de normes sociales, symbolisés par les revendications de Mai 68 et matérialisés entre autres par les lois Neuwirth ou Veil. Ces changements de normes se sont accompagnés d'un accroissement de la demande d'éducation des filles et d'un recul notable de leur âge de fin d'études. La question se pose toutefois de savoir si ces changements de normes sociales

peuvent être tenus pour seuls responsables du recul de l'âge auquel se forment les familles. Dans ce chapitre, nous étudions un autre élément de contexte susceptible d'avoir influé sur les trajectoires familiales : l'augmentation du chômage des jeunes et l'institutionnalisation des contrats précaires à l'entrée de la vie active.

Ce dernier chapitre s'inscrit ainsi dans la vaste littérature sur les liens entre l'accès à l'emploi et la formation de la famille (voir par exemple Ekert-Jaffé et Solaz, 2001 ; Prioux, 2003 ; Adsera, 2005 ou Goldstein et al., 2013). Notre principale contribution est de mettre en lumière l'importance de la nature de l'emploi obtenu – stable ou précaire – dans la décision d'emménager avec son conjoint, puis d'avoir des enfants. D'un point de vue méthodologique, il n'est guère aisé d'identifier le sens et l'ampleur des liens de causalité susceptibles d'exister entre l'accès à l'emploi et la formation de la famille. La corrélation existant entre la date du premier emploi et la date de la première cohabitation peut tout autant indiquer un lien causal de l'emploi vers le couple qu'un lien causal du couple vers l'emploi, voire de simples effets de sélection. Afin d'éclairer ces questions, ce chapitre développe une analyse de durée utilisant le modèle d'Abbring et van den Berg, 2003. Sous l'hypothèse que les individus ne peuvent anticiper longtemps à l'avance les dates exactes d'obtention de leur premier emploi ou de formation de leur premier couple (ou, en tous cas, ne prennent pas de décisions en lien avec ces événements plusieurs mois en amont), le modèle d'Abbring et van den Berg permet d'identifier les liens de causalité entre accès à l'emploi et formation de la famille. L'utilisation de ce modèle – menée sur deux enquêtes retraçant les trajectoires professionnelles et familiales d'un large échantillon de jeunes adultes – révèle qu'un premier emploi stable a un impact nettement plus important sur la mise en couple qu'un emploi précaire (et quantifie précisément la différence). Nos résultats révèlent par ailleurs que la stabilité de l'emploi importe également pour la décision d'avoir un enfant. Nous démontrons également que les liens entre emploi, cohabitation et genre ont beaucoup évolué au fil des générations. Pour celles nées avant le milieu des années cinquante, l'emploi constituait plus fréquemment une condition nécessaire pour fonder une famille chez les hommes que chez les femmes et la mise en couple jouait un rôle très négatif sur la participation des femmes au marché du travail. Pour les générations suivantes, les liens entre mise en couple et emplois apparaissent bien plus similaires entre hommes et femmes. D'une génération à l'autre, la montée du chômage des jeunes et de la part des contrats précaires expliquerait environ 25% des retards observés dans l'âge à la première mise en couple et environ 40% des retards dans l'âge au premier enfant.

Chapter 1

Competitive Schools and the Gender Gap in the Choice of Field of Study

JOINT WITH SON-THIERRY LY AND ÉRIC MAURIN

Abstract

In most developed countries, students have to choose a major field of study during high school. This is an important decision as it largely determines subsequent educational and occupational choices. Using French data, this paper reveals that enrollment at a more selective high school, with higher-achieving peers, has no impact on boys, but a strong impact on girls' choices: they turn away from scientific fields and settle for less competitive ones. Our results are not consistent with two commonly-advanced explanations for gender differences in field of study, namely disparities in prior academic preparation and in sensitivity to rank in class.

JEL: I21, I24.

1 Introduction

In most developed countries, male and female students still choose very different major fields of study during high school or during college. In French high schools for instance, male students are about 40% more likely than female students to specialize in science¹. These gender differences have attracted considerable attention as they likely explain a significant part of labor market differentials across gender groups. The choice of science as a major field of study is typically associated with the best prospective outcomes, but female students are still dramatically underrepresented in this field.

A long-standing literature has explored the causes of the gender gap in the choice of field of study, with a specific emphasis on gender differences in ability, expectations or preferences. Several influential studies have also emphasized the role of teaching practices and teachers' stereotypes². In this paper we analyze the role of another potential determinant of students' choices, namely the school environment in which they make their decisions. Specifically we look at whether the choice of field of study of girls and boys depends on the academic level of the schoolmates with whom they have to compete. In more selective schools, with higher-achieving peers, students may be induced to form new expectations about their chances of success in the different fields, which may eventually affect their choices.

It has long been recognized that enrollment at a more selective school, with higher-achieving peers, may affect students' subsequent performance and graduation probability, even though empirical evidence are mixed (see Pop-Eleches and Urquiola, 2013 or Abdulkadiroğlu, Angrist and Pathak, 2014 for example). Much less is known, however, on whether enrollment in selective schools affects students' choice of field of study. One basic reason for this lack of evidence is that the choice of major field of study often occurs at the same time as the high school choice, so that it is logically very difficult to define the impact of the latter on the former. Another basic difficulty is that the choice of enrolling in a more

¹See Direction de l'évaluation, de la prospective et de la performance (2014). The requirement to choose major fields of study during high school is not specific to the French system. Similar requirements exist in the UK, Spain, Italy, Switzerland, Denmark, Norway or Korea for instance. For evidence on the gender gap in field of study in secondary education, see for example Buser, Niederle and Oosterbeek (2014) for the Netherlands; Joensen and Nielsen (2016) for Denmark; Office for Standards in Education (OFSTED) for the UK; Buser, Peter and Wolter (2017) for Switzerland. For evidence on the gender gap in the choice of field of study in US post-secondary education system see for example Chen and Weko (2009).

²For an analysis of teachers and teaching practices, see Carrell, Page and West (2010); Joensen and Nielsen (2009); Lavy and Sand (2015). For an analysis of the role of attitude towards competition, see Croson and Gneezy (2009); Niederle and Vesterlund (2011, 2010); Buser, Niederle and Oosterbeek (2014). For an analysis of gender differences in preferences and expectations see Zafar (2013).

selective school and the choice of field of study likely depend on the same explanatory factors, such as students' willingness to compete. In such a context, where two decisions potentially share the same causes, it is typically very hard to evaluate the influence that they exert on each other, even when they do not take place at the same time. Students enrolled in selective schools tend to choose more demanding fields of study, but it does not follow that their choice is influenced by their school environment.

In this paper, we build on the features of the Parisian high school system to overcome these issues. One first feature of this system is that middle school students are assigned to high schools through a centralized process that gives priority to students with the best average grades in middle school. About half of high schools receive more applications than they have places to offer and enrollment at each of these schools is restricted to students whose middle school average grade is above a specific cut-off level computed by the system. The second basic feature of the system is that students do not have to choose their major field of study at the start of high school, but only after one year of familiarization (that is, grade 10). In this context, it is possible to isolate the impact of enrollment at an oversubscribed high school on subsequent choices of field of study by comparing students whose middle school achievements were either just above or just below the specific cut-off level of this more selective high school.

This Regression Discontinuity (RD) analysis first confirms that eligibility for enrollment in a more selective school is associated with a very significant increase in the ability level of high school peers. It also reveals that this increase in peer ability is even more significant in science than in humanities, so that enrollment in a more selective school is first and foremost associated with an increase in competition in science. Importantly these first stages effects on peer ability are very similar for boys and girls, consistent with the assumption that, in our set-up, there is no gender gap in the willingness to attend higher ranked schools.

By contrast, eligibility for admission into a more selective school, with higher-achieving peers, has very different effects on the choice of major field of study made by boys and girls one year later, at the end of grade 10. Specifically, it has no effect on boys, but induces a significant decrease in the probability that girls choose science and a symmetrical increase in the probability that they choose humanities. Eventually, enrollment at a more selective school is not followed by any significant change in students' overall graduation probability, but by a significant decrease in the share of girls who graduate in science.

Generally speaking, our results are consistent with experimental findings showing that female students are more likely to turn away from competitive settings than their male counterparts (Croson and Gneezy, 2009; Niederle and Vesterlund, 2011). They are also consistent with Rask and Tiefenthaler (2008) or Goldin (2015), who suggest that female students tend to be more responsive to a decline in performances than males. Our paper contributes to showing the decisive impact of these gender differences on the choices made by girls and boys during high school. Because science is the field of study where competition increases the most in sought-after schools, gender differences in attitudes towards grades and competition appear to induce many female students to turn away from science. Our findings are also reminiscent of the literature on college major choice in the US and on the role played by students' expectations in this choice (Stinebrickner and Stinebrickner, 2014; Wiswall and Zafar, 2015). In a related contribution, Arcidiacono, Aucejo and Hotz (2016) show that the probability that a minority student graduates in science may be much lower in more selective Californian universities than in less selective ones.

Finally, our paper contributes to the literature on the effect of going to a more selective school, with higher-achieving peers (Abdulkadiroğlu et al., 2017; Clark and Del Bono, 2016; Cullen, Jacob and Levitt, 2006; Jackson, 2013). Several recent papers build on a similar regression discontinuity design to provide evidence on the effect of selective schools on students' performance in various institutional contexts (Abdulkadiroğlu, Angrist and Pathak, 2014; Dobbie and Fryer Jr, 2014; Pop-Eleches and Urquiola, 2013). This literature finds mixed evidence on the impact of elite schools on student academic performance³. Because French students have to choose a major field of study at the end of their first year of high school, we are able to look not only at the impact on academic performance, but also on the choice of field of study. The effect of gaining admission to a more selective school appears to be much stronger on field of study than on academic performance. There is a small recent literature which documents similar findings about tracking within schools (He, 2016; Dougherty et al., 2017).

The paper is organized as follows. Section 2 describes the institutional context. Section 3 describes our administrative data sources while Section 4 provides basic graphical evidence on the impact of being eligible for admission into a selective school on students subsequent choices of field of study or graduation probabilities. Section 5 develops our Regression Dis-

³Most studies on elite schools in the US find little or no effects on academic achievement (Cullen, Jacob and Levitt, 2006; Abdulkadiroğlu, Angrist and Pathak, 2014; Dobbie and Fryer Jr, 2014), but positive effects are found in other contexts (Jackson, 2010; Estrada and Gignoux, 2017; Pop-Eleches and Urquiola, 2013). There is a small related literature on the impact of students' rank within their school on their subsequent academic performance (see Murphy and Weinhardt, 2014).

continuity analysis. Section 6 explores the candidate mechanisms that may explain that more competitive school contexts induce female students to turn away from science.

2 Institutional Context

In this section, we provide information on how middle school students are assigned to high schools in Paris as well as on the exams that they have to take and the choices that they have to make during their high school years. In the following sections, the main research question will be whether the high school to which a middle school student is assigned affects her subsequent choices and performance on exams.

2.1 The Assignment of Middle School Students to High Schools

In France, middle school runs from grade 6 to grade 9. Students complete grade 9 the year they turn 15. The curriculum is the same in all middle schools and there is no streaming by ability. At the end of grade 9, students enter into high school, which runs from grade 10 to grade 12. This paper focuses on students who completed 9th grade in public middle schools in the education region of Paris, in either 2009 or 2010.

France is divided into thirty education regions and the education region of Paris represents about 3% of French middle school students. In this region, there are about 100 public middle schools and about 50 public high schools where middle school students can pursue general education courses. Middle school students are assigned to public high schools through a centralized process called *Affelnet*, which is completely gender-blind and which is described in detail in Hiller and Tercieux (2014) or in Fack, Grenet and He (2015). Students are first asked to list up to six choices of public high schools in descending order of preference⁴. Paris is divided into four geographical districts (West, East, North, and South), and there is a very strong incentive to apply to high schools in one's district of residence since the system gives priority to home-district over out-of-district applications. Also, within each district, a priority is given to low-income students, namely the 20% students eligible to means-tested financial assistance. For the other students, the system ranks their applications according to the average of their 9th grade marks across all subjects and assigns them to as many seats as possible using a deferred acceptance algorithm (Roth,

⁴Students from public middle schools can also apply to private high schools, but this is not processed by *Affelnet*. As discussed below, we checked that eligibility for admission into a more selective public high school has no effect on the probability that students from public middle schools go to a private high school. Students from private middle schools can also apply to public high schools, but their applications are processed separately, after those of students from public middle schools.

1982) and a multi round process⁵. There are twelve subjects (Mathematics, Physics, Biology, Technology, French, History/Geography, Sport, two foreign languages, Art, Music and Discipline) and the marks used to compute the average score used by the system are first standardized at the region level. Standardization amounts to weighting each mark by the inverse of its standard deviation. These weights being revealed only ex post (that is, after all students have submitted their choices), the weighted average marks used by the system are ex ante impossible to precisely predict or manipulate.

In substance, the algorithm first assigns the students with the best 9th grade average marks to their preferred schools until one school starts being oversubscribed. This top school is then dropped from the application lists of the remaining (not yet assigned) students. These students are then re-ranked and the process is reiterated until another school starts being oversubscribed, and so on. At the end of this first round, there are no seats left in a fraction of schools (that is, the oversubscribed ones), whereas the other fraction is still undersubscribed. Similarly, a fraction of students are assigned to one of the schools of their list whereas the other fraction are still unassigned (that is, they applied for oversubscribed schools only). To further improve the assignment rate, each unassigned student is then asked to form new choices, namely to apply to at least one of the undersubscribed schools, and the process is reiterated. At the end of this second round, some students are still unassigned, and the education administration helps them find a seat in one of the remaining undersubscribed schools in an informal way. Undersubscribed schools are typically those that end up admitting a significant proportion of out-of-district students.

The key feature of this assignment process is that it is possible to define a minimum admission score for a large fraction of high schools, namely the oversubscribed ones. As discussed below, in years 2009 and 2010, about half of the public high schools of the region of Paris appear to be oversubscribed, with very significant discontinuities in the rate of enrollment of students at specific cut-off points of the distribution of 9th grade scores. This feature will make it possible to build on a regression discontinuity analysis to evaluate the effect of being admitted to these schools on subsequent educational outcomes. Specifically, we will focus on students whose 9th grade scores fall either just above or just below the cut-off point of an oversubscribed school and we will compare the outcomes of those just above with those just below. The vast majority of these students continue general education in high school, and the question will be whether enrollment at a higher ranked school affects either their major field of study in high school or the level of their academic

⁵A bonus is granted for students who apply to high schools where they have a brother or a sister.

performance⁶.

Generally speaking, our research strategy relies on the assumption that individual 9th grade scores and schools' minimum admission scores cannot be manipulated and predicted. As discussed above, there is little scope for manipulation of individual scores. With respect to schools' minimum admission scores, they depend on many factors that are no easier to predict than individual scores, such as the exact number of low-income students and the exact distribution of their choices across schools in each district. In this set up, there is again little scope for manipulation or prediction⁷. As discussed below, we find no evidence of discontinuities in the density of individual scores – or in students' pre-assignment characteristics – at the cut-off.

2.2 Major Field of Study and High School Exit Exams

At the end of their first year of high school (grade 10), French students can either pursue general education or enter a technical or a vocational education program. Furthermore, those who pursue general education have to choose a major field of study. There are three possible fields: science (field "S"), economics and social sciences (field "E/S") or languages and literature (field "L"). The number of students per school and field of study is not set once and for all. It can vary significantly from one year to another so as to meet the choices of students⁸. At the end of 10th grade, students are asked the field of study that they prefer and, eventually, the vast majority is allowed to pursue the track that they prefer. This is a key choice: each field of study corresponds to a very specific curriculum, specific high school examinations, and specific opportunities after high school.

For example, for those who choose to specialize in science, the scientific subjects represent 50% to 60% of compulsory courses in grade 11 and 12. By contrast, for those who choose to specialize in languages and literature, scientific subjects represent less than 5% of compulsory courses. With respect to post-secondary education, it is virtually impossible to enter an engineering school or a medical school after non-scientific studies in high school⁹.

⁶It should be emphasized that no school is oversubscribed with low-income students (the maximum proportion of low-income students is 60%). Hence, there is no school that can define a minimum admission score for low-income students. Minimum admission scores are defined for non low-income students only.

⁷We checked that most schools do not have the same minimum admission score for the first and for the second cohort. About half of schools do not even have the same exact rank within their district (in terms of minimum admission score) from one cohort to another.

⁸We checked that the number of students who choose science as major field of study actually varies significantly from one year to the other within most schools. The within-school variation rate is on average 10%.

⁹According to a national longitudinal survey conducted by the Ministry of Education (so called *Panel d'élèves du second degré, recrutement 1995*), only about 1% of students who graduate in humanities pursue a science track after high school (see also Direction de l'évaluation, de la prospective et de la performance, 2012).

Generally speaking, science is the most prestigious field of study. Students in the scientific track performed on average 60% of a standard deviation higher in middle school than students in the social sciences track and 75% of a standard deviation higher than those in the languages and literature track. In our working sample, about 31% of students choose the scientific track, 13% specialize in literature and languages, 22% specialize in social sciences whereas about 34% opt for a more technical or vocational program, or drop out from education.

The first year of high school (grade 10) is dedicated to exploring the different subjects and to choosing a major field of study. After this exploration year, students have very little leeway to change their major field of study. In Paris, only about 2.5% of students change their major field of study after 10th grade. The two last years of high school (11th and 12th grade) are dedicated to the preparation of the national high school exit exam, the *baccalauréat*, which is a prerequisite for entry into post-secondary education. Students have to take one exam per subject, and they obtain their diploma if their weighted average mark across subjects is 10/20 or more, where subjects taken and weights depend strongly on the major field of study. For instance, the weights of exams in scientific subjects represent about 50% of the total for students who choose these subjects as major field of study, whereas the weights of these subjects is only about 20% for those who choose social sciences and 5% for those who choose languages and literature. Most exams are taken at the end of 12th grade, except for exams in French (oral and written) which are taken at the end of 11th grade. Students whose weighted average mark across subjects is 12/20 or more graduate with honors. Graduation with honors is granted to about half of the students each year in each field.

2.3 Selective Undergraduate Programs

High school graduation is a prerequisite for admission into post-secondary education programs. About half of these programs are selective, and selection depends on the grades obtained in the two last years of high school as well as on students' ranks within their class. The *Classes Préparatoires aux Grandes Écoles* (CPGE) are among the most prestigious such selective programs. These two-year programs prepare students to take the entry exams of the most prestigious graduate programs (so called *Grandes Écoles*). The last important question addressed in this paper is whether enrollment at a more selective high school at the end of grade 9 affects the subsequent probability of gaining access to CPGE programs at the end of grade 12.

There are different types of CPGE programs: some are specialized in science (they prepare

for entry exams at engineering graduate schools), some are specialized in economics and business (they prepare for entry exams at business schools) and some are specialized in literature and languages (they prepare for entry exams at top graduate programs in this field). Each year, in Paris, about 20% of students from high school general education programs gain access to a CPGE. The vast majority graduate from high school with honors. When a student applies to a CPGE program, her high school has to provide the average marks (as assessed by teachers) that the student obtained in each subject for each quarter during 11th and 12th grade, as well as the corresponding rank within their class. Hence students from more selective high schools may benefit from the prestige of their schools, but may suffer from being less well ranked within their class.

3 Data and Methods

3.1 Data

In our empirical analysis, we use administrative data providing detailed information on students who finished middle school (9th grade) in either 2009 or 2010 in the education region of Paris. For each student, we know the high school to which she was assigned after 9th grade, the field of study chosen at the end of 10th grade as well as the field of study in which she graduated at the end of 12th grade. For each student, we also know whether she repeated a grade during high school, whether she dropped out from education before graduation and whether she got admitted into a CPGE program after high school.

With respect to students' academic performance, we know the average marks given by teachers in each subject during 9th grade as well as results at the national exams taken in Mathematics and French at the end of 9th grade (externally set and marked). We also know students' results at the national examination (*baccalauréat*, externally set and marked) taken at the end of high school (12th grade). As discussed previously, the score used to assign middle school students to high schools corresponds to an average of the average marks given by 9th grade teachers.

To construct this dataset, we used schools' registration records as well as administrative records with exhaustive information on results at the end of middle school and at the end of high school national exams, for each year between 2009 and 2014. We were able to match these different data sources using students' ID number.

Finally, we augmented our individual-level database with information coming from

school-level administrative datasets, namely information on high school size as well as on the proportion of female teachers and on the distribution of teachers' age.

3.2 Cut-off Scores

In this section, we consider the 52 public high schools which enter the centralized assignment system in 2009 or 2010. For each cohort and each district, we focus on the 9th grade students whose applications went through the standard assignment process, that is, they are not low-income and come from a public middle school of the district¹⁰. Our purpose is to identify the public high schools which received more applications than they could accommodate, and to estimate the lowest 9th grade score that students had to earn to gain admission into these schools.

Our data do not provide information on students' rank order lists, so that it is not possible to identify which schools are oversubscribed (and their minimum admission scores) directly from the data. To overcome this issue, we built on a method developed by Hansen (2000). Among all possible minimum admission scores, the method first identifies for each school those which coincide with a significant discontinuity in enrollment rates. If any such thresholds exist, the method amounts to choosing for each school the specific threshold which corresponds to the most significant discontinuity. More details on how we implemented this procedure are provided in Appendix 1.B.

This method was used by Card, Mas and Rothstein (2008) to estimate for each large American city the minimum proportion of minorities above which white residents start to flee a neighborhood. This method was also used by Hoekstra (2009) in his analysis of the benefits of attending a flagship university. In a recent contribution, Porter and Yu (2015) show that RD estimates and standard errors are unaffected when we use this two-stage method, where cut-offs are estimated first, followed by a standard RD model.

In our case, this method makes it possible to identify 23 schools that were likely significantly oversubscribed in 2009, and 26 in 2010 (out of 52). For each oversubscribed school j , it is possible to define the sub-sample S_j of students who are not low-income, who come from a public middle school located in the same district as j and whose 9th grade scores are closer to the minimum admission score of j (denoted q_j^*) than to any other minimum admission score in the district. This sub-sample S_j covers the set of students who are either just above or just below the minimum required score for admission into j ,

¹⁰As mentioned above, low-income students and students from private middle schools are processed separately by the centralized system.

but for no other school. For each student i in S_j , we can also define her percentile rank distance D_i to the cut-off admission score of j , where $D_i > 0$ corresponds to students above the cut-off and $D_i < 0$ to students below the cut-off¹¹.

In the remainder of the article, we develop a regression discontinuity analysis using D_i as running variable, and focusing on the students in the pooled sample of all S_j whose distance D_i to the admission cut-off is smaller than 25 centiles¹². We also drop from our working sample students whose 9th grade score falls just between the two quantiles below and above q_j^* , namely students whose 9th grade scores are too close to the cut-off to know for sure whether they are above or below.

Overall, our working sample consists of about 7,500 students whose 9th grade scores fall close to an admission cut-off, and our basic research question will be whether we observe discontinuous behavior and performance above and below the $D_i = 0$ cut-off. In Appendix 1.A Tables 1.A1-1.A4, we report basic descriptive statistics for students in this pooled sample. In this sample, about 24% of girls choose to specialize in science after 10th grade, against about 38% for boys.

4 Graphical Evidence

Before moving on to a more comprehensive regression analysis, we start by providing graphical evidence on the effect of eligibility for admission into a higher ranked school on three basic outcomes, namely the type of school in which students end up going to after grade 9, their choice of field of study at the end of grade 10 and their performance at the high school exit exam at the end of grade 12. As discussed above, we focus on the pooled sample of students whose 9th grade score is either just above or just below the minimum admission score of an oversubscribed high school. The basic question is whether we observe discontinuities in enrollment, choice and performance between students above and below the cut-off score. Another important issue is whether we observe similar discontinuities for boys and girls.

¹¹Within each sample S_j , we first define the percentile rank of each student using their 9th grade score. We then re-scale this percentile rank variable so that the re-scaled rank is zero for students whose 9th grade score is equal to the minimum admission score. This running variable is the same as the one used by Abdulkadiroğlu, Angrist and Pathak (2014).

¹²We obtain this bandwidth using Calonico, Cattaneo and Titiunik (2014). As discussed in the next section, we checked that our results are robust to alternative bandwidths.

4.1 First-Stage: Effect of Eligibility on Enrollment

To start with, Figures 1.1a and 1.1b show the probability of enrollment at a higher ranked high school for students whose 9th grade score is either just above or just below the minimum admission score¹³. Figure 1.1a focuses on girls whereas Figure 1.1b focuses on boys. Comfortingly, they show a very significant shift at the cut-off, the probability of enrollment at a higher ranked high school being about 17 percentage points higher for students just above the cut-off than for students just below the cut-off. This finding is suggestive that, among students whose 9th grade score falls close to the minimum admission score of an oversubscribed school, about 17% prefers this school to any lower ranked school. Results are very similar for girls and boys. It is consistent with the assumption that girls are no less willing than boys to enroll into selective schools¹⁴.

The first potential reason for why enrollment at a higher ranked high school may matter is that it is likely associated with a strong increase in the ability level of high school peers and, consequently, with a strong decline in students' own ability rank. To provide evidence on this issue, we define for each student her percentile rank in the distribution of 9th grade scores across her high school peers. Figures 1.2a and 1.2b show the variation in this rank variable for students just above and just below the minimum admission score. On the left hand side of the Figures 1.2a and 1.2b, we observe a smooth increase in students' rank as their (absolute) score increases and gets closer to the cut-off. On the right hand side, we observe a similar mechanical improvement as we consider students with better scores. But when we compare the first non-eligible students (just below the cut-off) with the last eligible ones (just above), we observe a significant downward shift of about -6 percentile ranks. Consistent with Figures 1.1a and 1.1b, the shift is similar for boys and girls. Figures 1.3a and 1.3b further explore the nature of the change in peer competition at the cut-off. We considered marks in Mathematics and French (using results at end-of-middle school national exams, externally set and marked) and we divided students' marks in Mathematics by their marks in French, so as to construct (after standardization) a measure of students' relative ability in Mathematics. The Figures plot the average of this Mathematics/French ratio across high school peers, above and below the cut-off. They reveal a significant upward shift in peers' rel-

¹³For each Figure, the solid line shows smoothed values of a kernel-weighted local polynomial regression of the dependent variable on the standardized distance to the threshold. We use a triangular kernel and a degree 1 polynomial smoothing. Smoothed values are estimated separately on each side of the cut-offs. Plotted points represent the mean of the dependent variable for all applicants in a one-unit binwidth. The bandwidth is computed following Calonico, Cattaneo and Titiunik (2014).

¹⁴It should be noted that enrollment rates are not exactly equal to zero for students to the left of the cut-off in Figure 1.1. This fact reflects that some schools are allowed to develop programs that are focused on a specific theme (sports, music, arts...) and that applications to these specific programs are processed directly by school, not by the centralized system.

ative ability in Mathematics at the cut-off, for both male and female students. We checked that these results are very similar when we use marks given by 9th grade teachers to construct our Mathematics/French ratio. These findings likely reflect that the willingness to attend a higher ranked school is stronger among students who are relatively stronger in Mathematics. More importantly, these results show that enrollment at a higher ranked school is associated with an increase in peer competition which is even stronger in science than in humanities.

4.2 Field of Study at the End of 10th Grade

Enrollment in a higher ranked school is associated with an increase in the level of peer competition and the next basic question is whether it affects students' subsequent choices and performance. Figures 1.4a and 1.4b show the probability of choosing science as a major field of study at the end of 10th grade for students just above and just below the minimum admission score. Similarly, Figures 1.4c and 1.4d show the probability of choosing humanities as a major field of study, namely the probability to choose either social sciences or languages and literature. Finally, Figures 1.4e and 1.4f show the probability of opting for technical or vocational education at the end of 10th grade. Taken together, these Figures reveal that eligibility for admission into a higher ranked school induces a significant fraction of girls to choose humanities rather than science as a major field of study. We observe a significant rise in the proportion of girls who choose humanities¹⁵, a significant decline in the proportion who choose science as well as a marginal decline in the proportion who opt for technical or vocational education. By contrast, we do not observe any significant shifts for boys.

Figure 1.5a further shows that the negative shift in the probability that female students choose science at the end of 10th grade is followed by a shift of similar magnitude in the probability that they graduate in science at the end of 12th grade. Similarly, Figure 1.5c shows that the positive shift in the probability that girls choose humanities at the end of 10th grade is followed by a parallel shift in the probability that they graduate in humanities at the end of 12th grade. Taken together, Figures 1.5a and 1.5c are suggestive that girls who are induced by competition to turn away from science actually succeed in graduating in humanities at the end of 12th grade, but would also have succeeded in graduating in science, had they been assigned to a less competitive school. Finally, consistent with Figures 1.4b and 1.4d, Figures 1.5b and 1.5d show no shift in the probability that boys graduate in humanities and no shift in the probability that they graduate in science. Hence, enrollment at a more selective school has no effect neither on boys' choice of field of study

¹⁵There is an increase in both the proportion of girls who specialize in social sciences and the proportion who specialize in language/literature, though the increase is stronger for specialization in social sciences.

at the end of grade 10 nor on their performance at end-of-high school exams.

Overall, enrollment at a higher ranked school does not seem to affect students' graduation probability, only the field in which girls choose to graduate. The fact that we find similar reduced form effects on the field of study chosen by boys and girls at the end of 10th grade as on the field in which they graduate at the end of 12th grade is also consistent with the fact that very few students change fields after grade 10.

Before moving on to our regression analysis, it should be emphasized that our figures represent the effects of eligibility for enrollment (reduced form effects), not the effects of enrollment (local average treatment effects). Given that we find that about 17% of eligible students actually enroll in the higher ranked school at the cut-off (the compliers), enrollment effects on compliers are likely about 6 times larger than eligibility effects (where $6 \approx \frac{1}{0.17}$). For example, under the maintained assumption that eligibility matters only insofar as it affects enrollment, a -6 percentage point effect of eligibility on the probability to graduate in science can be interpreted as a -36 percentage point effect of enrollment on the same probability for the 17% of students whose assignment is actually affected by eligibility at the cut-off.

5 Regression Results

The previous graphical analysis focused on the most basic high school outcomes: students' ability rank, main field of study, and performance on exit exams. To complement and test the robustness of these graphical findings, this section provides a regression analysis of the effect of eligibility for admission into a higher ranked school on a more comprehensive set of outcomes, using a large set of control variables and a standard regression discontinuity model (Lee and Lemieux, 2010). We keep on using the same pooled sample of students as in the previous section, namely the pooled sample of students whose 9th grade score is either just above or just below the minimum score required for admission into a higher ranked school. For each student in this sample, we still denote D_i the percentile rank difference between the 9th grade score of student i and the minimum admission score to which she is close. Variable D_i is positive if student i is just eligible for admission into the higher ranked school to which her score is close to, and D_i is negative if student i is not eligible for admission into this higher ranked school. For each outcome Y_i available in our longitudinal database, we estimate the following model:

$$Y_i = \alpha \mathbb{1}\{D_i \geq 0\} + f(D_i) + X_i\gamma + u_i \quad (1.1)$$

where variable X_i is a set of control variables which includes controls for students' age, gender, family background, average marks in grade 9 as well as a full set of dummies indicating the high school that corresponds to the nearest cut-off. Function $f(D_i)$ is a first order spline function of D_i , and we use a tent-shaped edge kernel centered around the admission cut-offs¹⁶. Variable u_i represents the unobserved determinants of students' choices and performance in high school. The parameter of interest is α . Under the maintained assumption that there is no discontinuity in the distribution of unobserved u_i at the cut-off, this parameter can be interpreted as the causal effect of eligibility for admission in a higher ranked school on the outcome under consideration. Tables 1.A2 to 1.A4 in Appendix 1.A provide the means and standard deviations of the different set of outcomes used in this regression analysis. In the same appendix, Table 1.A5 reports the results of regressing the different observed baseline characteristics (gender, age, family background and average marks in grade 9) on a dummy indicating eligibility for admission into a higher ranked school (that is, $\mathbb{1}\{D_i \geq 0\}$) using Model (1.1). Consistent with our identifying assumption, we do not find evidence of any discontinuous variation in baseline characteristics at the eligibility cut-off, that is, the "effect" of $\mathbb{1}\{D_i \geq 0\}$ on baseline characteristics is never statistically significant at standard level. Also, when we regress the eligibility dummy $\mathbb{1}\{D_i \geq 0\}$ on the full set of baseline variables, an F-test does not reject the null assumption that the coefficients are jointly equal to zero. Figures 1.A1a-1.A1k in Appendix 1.A provide additional graphical evidence on the smoothness of baseline characteristics in the neighborhood of admission cut-offs. Finally, as a last specification test, we build on McCrary (2008) to test for manipulation of the running variable around the cut-off. The test does not show any significant difference in the (log) height at the cut-off and does not reject the null assumption of continuous distribution of the running variable at standard level (p-value = 0.75). The result holds true for both male and female samples.

5.1 First Stage Effects on School Environment

The first panel of Table 1.1 shows the estimated effects of eligibility for enrollment into a higher ranked Parisian public school at the end of middle school (grade 9) on the type of school attended during the following academic year. It shows that eligibility has no significant effect on the probability that students repeat 9th grade, nor on the probability that they drop out from education. Put differently, there is no evidence that middle school students who fail to be eligible in their preferred high school had rather repeating 9th grade (or dropping out from education) than going into another high school. The Table also

¹⁶This specification is similar to the non parametric specification used by Abdulkadiroğlu, Angrist and Pathak (2014). In Subsection 5.3, we check that our results are robust to using higher order spline functions with a uniform kernel.

shows that eligibility for admission into a higher ranked public school in Paris has no effect on the probability that students go into a private high school or leave the education region of Paris. Students who do not obtain their preferred choices do not fly away from public education nor from Paris. Overall, consistent with Parisian institutions, the Table shows that the main effect of eligibility for enrollment into a selective high school is to induce a very significant rise in the probability of enrollment at this very high school. Findings are very similar for boys and girls.

As we find no impact on the proportion of students who repeat 9th grade nor on the proportion of students who drop out from education, it is possible to focus on the subsample of students who go to high school and to look at how eligibility affects their peers' characteristics and school environment at entry into high school. The second panel of Table 1.1 shows that eligibility has no significant effect on the proportion of female students or on the proportion of low-income students among high school peers¹⁷. It is associated, however, with a significant rise in their level of ability. This rise in peer ability translates into a significant decline in students' ability rank within their school as well as with a very significant rise in students' probability to fall in the bottom quartile of the distribution of peer ability within their school (+14 percentage points). These shifts are about as significant for female and male students.

The Table also confirms that the rise in peer ability in higher ranked schools is even stronger in Mathematics than in French, regardless of whether we focus on girls or on boys and regardless of whether ability is measured through teachers' assessment or through results at the end-of-middle school exams. When we divide baseline marks in Mathematics by baseline marks in French, we find that the average of this Maths/French ratio across high school peers is significantly stronger for students eligible for enrollment at a higher ranked school. Again, we checked that this result holds true regardless of whether ability is measured through teachers' assessment or through results at the end-of-middle school exams. Overall, we get an array of results suggesting that enrollment in a higher ranked school is associated with an increase in peer competition which is even stronger in Mathematics than in French.

To further explore why enrollment at a more selective school may make a difference, Table

¹⁷This result is a consequence of the fact that low-income students have a priority access to the different high schools of their district and, consequently, are not constrained by the admission thresholds under consideration. In that sense, the Parisian set-up offers the possibility to analyze the effects of a rise in peer ability holding peer family background and peer gender composition constant. Recent studies by Jackson (2012) or Eisenkopf et al. (2015) are suggestive that the gender composition of schools may affect students' achievement and choices.

1.A6 in Appendix 1.A shows the results of re-estimating Model (1.1) using the characteristics of teachers and schools as dependent variables. It reveals that there is little variation in teachers' gender and age across higher and lower ranked schools. Teachers tend to be older in higher ranked schools, but the age shift is small (about +0.5 year) and not significantly different for girls (+0.6 year) and boys (+0.3 year). The size of schools and the number of teachers per student are also very similar above and below the cut-off. Similarly, there is no significant shift in the distance (in meters) between students' past middle school and students' current high school. These results are suggestive that enrollment at a more selective school does not come at the cost of more crowded classrooms or longer travel distance from home to high school¹⁸. Not surprisingly, the main shift in school characteristics is observed for past results at the high school exit exam (the *baccalauréat*). Students just above the cut-off are assigned to high schools that obtained significantly better results at the *baccalauréat* the year before the assignment. This finding is consistent with schools' results at the exit exam being one the main driver of middle school students' choices¹⁹. Further explorations reveal that the increase in schools' graduation rates at the cut-off is significant in science, but not in humanities. This latter result is consistent with our previous finding that enrollment at a higher ranked school is first and foremost associated with a rise in peer competition in science.

5.2 Major Field of Study and Performance on Exams

Table 1.2 shows the effect of eligibility for admission into a more selective school on students' choices and performance. Consistent with graphical evidence, it shows that eligibility has no effect on boys, but induces a very significant decline in the probability that girls choose science as major field of study at the end of 10th grade (-7.2 percentage points) and a rise in the probability that they choose humanities (11.0 percentage points). These effects on girls' choices translate into a significant negative effect on the probability that they graduate in science at the end of 12th grade, and into a significant positive effect on graduation in humanities.

Given that boys' field of study are the same below and above the threshold, the fact that their performance at exit exams are also the same above and below the threshold can be interpreted as meaning that enrollment at a more selective school has no effect on their

¹⁸The resources allocated to each school are defined each year by the central administration so as to guarantee maximum equality across schools. In this set up, it should come as no surprise that we find little variation in the quality and quantity of teachers across schools.

¹⁹Schools' results at high school exit exams is the only information on schools' outcomes that is publicly available. Each year, early April, the ministry of education publishes the results of all high schools for the previous academic year. High school league tables are published extensively in the main newspapers.

performance at exit exams, be they taken in science or in humanities. For girls, the diagnosis is less straightforward as they do not choose the same fields of study above and below the threshold. Our regression results suggest, however, that girls who are induced to fly away from science would have succeeded in graduating in science, had they not been admitted into a higher ranked school. In fact, most of them would have succeeded in graduating in science with honors, as suggested by the significant decline in the proportion who graduate in science with honors observed at the cut-off (-5.6 percentage points). Changing field of study does not appear to be a way for girls to improve their overall probability to graduate with honors.

Another reason for why many students enrolled at higher ranked schools choose humanities rather science may be that humanities classes are of a specific quality in these schools, so that students have better chances to be admitted into the most selective undergraduate programs (CPGE) specialized in humanities. But this is not what we observe, regardless of whether we focus on female or male students. In fact, there is no significant increase in the probability of being admitted into a CPGE program specialized in humanities and no significant decline in the probability of being admitted into a CPGE program specialized in science at the cut-off²⁰.

It should be again emphasized that these estimates capture the reduced form effects of eligibility for enrollment, not the effect of enrollment per se. Tables 1.A7-1.A9 in Appendix 1.A provide estimates of the corresponding LATE, under the standard assumptions that eligibility matters only insofar as it affects actual enrollment and that eligibility never induces students not to enroll (monotonicity assumption). They are suggestive that enrollment at a higher ranked school induces a rise of about 67 percentage points in the probability to fall in the bottom quartile of the distribution of peer ability, followed by a decline of about 36 percentage points in the probability that female compliers graduate in science. These LATE suggest that falling in the bottom of one's class has very direct effects on girls' choices, but little effects on boys' ones.

5.3 Robustness and Falsification Tests

As a falsification test, we first replicated our regression analysis on the sample of low income students as well as on the sample of students coming from private middle schools, namely the two groups of students whose applications are processed separately by the

²⁰The latter finding also likely reflects that our working sample consists of students who are good students, but not top students. As shown in Table 1.A4 in Appendix 1.A, on average only about 2% of our sample obtain admission into a CPGE specialized in science each year, so that there is little room for this proportion to be negatively impacted at the cut off.

centralized system. Comfortingly, we find no effects on enrollment and no effects on field choice for both groups of students (see Table 1.A10 and Table 1.A11 in Appendix 1.A). Effects on enrollment and field choice are significant only for students whose applications are actually constrained by the thresholds under consideration, namely only for non low-income students from public middle schools.

As a robustness check, we also replicated our basic regression analysis separately on cohorts 2009 and 2010 (see Table 1.A12 in Appendix 1.A). Comfortingly, both first-stage effects on enrollment and reduced form effects on field of study appear to be similar for both cohorts, even though they are a bit less well estimated with the first cohort. Our findings do not seem to be driven by the behaviors and choices of a specific cohort.

We also replicated our analysis separately on subgroups defined by family background (see panel A of Table 1.A13 in Appendix 1.A). Both first stage and reduced form effects tend to be stronger for girls with a better family background. We also divided our sample according to whether the closest admission threshold is located in the top or the bottom tercile of the distribution of admission thresholds. Again, both first stage and reduced form effects tend to be stronger for students whose scores are close to top admission thresholds (see panel B of Table 1.A13 in Appendix 1.A). However, it should be emphasized that effects are less well estimated on the various subsamples than on the full sample and differences across subgroups (be they defined by family background or by the level of the threshold) are not significant at standard level.

To further assess the robustness of our results, we re-estimate Model (1.1) with a bandwidth computed following Imbens and Kalyanaraman (2012) rather than with a bandwidth computed with the method by Calonico et al. (2014). We obtain qualitatively similar results (see Appendix 1.A Table 1.A14). We also find similar results when we re-estimate Model (1.1) without any controls for pre-assignment characteristics, consistent with the assumption that there is no discontinuity in pre-assignment characteristics at the cut-off (see Table 1.A15 in Appendix 1.A). Finally, Table 1.A16 in Appendix 1.A shows that our basic results are unchanged when we use a uniform kernel and alternative spline functions.

6 Mechanisms

The results obtained so far are suggestive that enrollment in a higher ranked school is associated with an increase in peer ability which is even more significant in science than in humanities. They also suggest that this change in peer group composition induces a signifi-

cant decline in the proportion of female students who choose science as major field of study whereas it has no effect on male students. In this section, we discuss in turn the main mechanisms that may help to explain this finding. In particular, it is still to be explained why an increase in peer competition in science should affect female students and not male students.

6.1 Attitude towards Competition

One first possible explanation for our findings is simply that female students are more responsive than male students to an increase in peer competition, as suggested by several recent studies (Niederle and Vesterlund, 2007; Croson and Gneezy, 2009; Niederle, 2016; Niederle and Vesterlund, 2011). In particular, choosing less competitive and prestigious fields of study may be a way to avoid signaling traits that could be sanctioned by potential male partners, such as ambition and assertiveness (Bursztyn, Fujiwara and Pallais, 2017). In this scenario, it should come as no surprise that female students end up turning away from science, if science is the field of study that is the most prestigious and where competition increases the most.

A reinforcing mechanism may be that female 10th graders are more concerned about grades, and maybe also about their future educational prospects, than their male counterparts (Rask and Tiefenthaler, 2008; Goldin, 2015). There is a long standing psychology literature suggesting that female students are more prone to anxiety than male students, especially when they have bad grades, maybe because they are more concerned with pleasing adults and professors (see Pomerantz, Altermatt and Saxon, 2002 for example). Under this assumption, female students may turn away from science in order to reduce the psychological costs associated with poor academic performance.

Overall, gender differences in attitude towards competition and bad grades represent plausible mechanisms behind our main findings, even though we have no direct evidence for or against this type of explanations in our setting.

6.2 Rank Consideration

Another possible explanation for our findings is that female students choose their field of study so as to be assigned to classes in which they can reach better ranks and where they have better chances of gaining admission into more selective undergraduate programs. As mentioned above, about half of the programs into which students can enter after high school are selective and most of these programs take account of students' class rank to decide upon admission. Given this fact, it may appear to be a good strategy to choose one's

field of study so as to maximize class rank. In this scenario, estimated field effects on female students would not reflect their specific exposure to anxiety or their specific attitude towards competition, but their being more aware of the importance of being well ranked.

To explore this assumption, we constructed a variable describing the percentile rank that students can expect after grade 10 if they choose science as a major field of study and the percentile rank that they can expect if they choose humanities²¹. We then analyzed these two expected rank variables (as well as the difference between them) using the same RD design as in the previous sections (see Table 1.A17 in Appendix 1.A). This analysis confirms that enrollment at a higher ranked high school is associated for both male and female students with a decline in expected ranks. But we find that this negative effect is similar in science and in humanities. Hence, for both male and female students, we do not find any significant variation in the difference between expected ranks in science and expected rank in humanities at the cut-off. In that sense, enrollment at a higher ranked school does not appear to induce specific incentive to choose humanities rather than science as a major field of study. This result is consistent with our previous findings showing that admission into a higher ranked school has no effect on the probability to graduate with honors, meaning no effect on the probability to be in the top half of the distribution of high school graduation scores across students with the same field of study. If girls' choices were driven by the possibility to improve their relative ranks, we would likely observe something different, namely a positive effect on their probability to graduate with honors.

6.3 Comparative Advantages and Capacity Constraints

Another possible explanation for our findings may be that female and male students do not have *ex ante* the same strong points, namely female students tend to be relatively stronger in humanities whereas boys are relatively stronger in science. Specifically, at each point of the distribution of 9th grade score, female students have relatively better grades in humanities whereas male students have relatively better grades in science. In our working sample of students whose 9th grade score is close to an admission cut-off, we checked that

²¹We have information on the score obtained by each student in each elementary subject during 9th grade (Maths, French, Physics, History, etc.). Hence, for each student, it is possible to compute a weighted average score using these elementary scores and the weights that corresponds to high school graduation in science (hereafter, science score). Similarly, we can compute for each student a weighted average score using the weights that correspond to high school graduation in humanities. Eventually, for each cohort t and each school s , it is possible to compute the percentiles of the distribution of science scores for science students (hereafter, the science track percentiles) as well as the percentiles of the distribution of humanities scores for humanities students (the humanities track percentiles). Eventually, for each student in school s and cohort t , our measure of her expected rank in science corresponds to her science track percentile rank within school s and cohort $t-1$ whereas our measure of her expected rank in humanities corresponds to her humanities track percentile rank.

girls' pre-admission grades in Mathematics are on average about 17% of a SD below that of boys with similar average scores whereas their pre-admission grades in French are about 16% of a SD above.

In this context, one reason for why girls just above an admission cut-off tend to be more responsive than boys to an increase in competition in science may be that they are relatively weaker in this subject and more exposed to fall behind their peers if they choose this subject as a major field of study. It may even be that girls are overrepresented among students who are encouraged by teachers and schools to choose humanities rather than science as a major field of study.

In theory, there is no constraint on the number of students who can pursue in the science track at the end of 10th grade. In practice, given that science is the most prestigious and sought-after track, it may be that schools induce students who are weaker in science than in humanities to choose humanities, if only to have a minimum number of students in their humanities courses at the start of 11th grade. In this scenario, the underrepresentation of female students in science would not reflect students' preferences, but constraints on field choices. Specifically, the stronger effects of selective schools on female students would reflect that female students are overrepresented among students who are initially relatively weak in science and that science is the most popular and constrained choice.

To further explore this assumption, we replicated our regression discontinuity analysis separately on the four subsamples defined by the quartiles of the distribution of students' relative strength in Mathematics (as measured by the standardized ratio between their score in Mathematics and their score in French at the end-of-middle school national exam). For each subsample, we estimated jointly the main effect of eligibility for enrollment at a higher ranked school and the interaction between eligibility for enrollment at a higher ranked school and a dummy indicating that the student is a girl (Table 1.A18). For each subsample, the main effect captures the impact of eligibility on boys whereas the interacted effect captures the differential impact on girls and boys.

Assuming that the gender-biased effects on field choices reflect girls' relative weakness in science, these effects should become much less significant when we compare boys and girls with similar relative strength in science, meaning no differential effects between boys and girls in the different subsamples. This is not what we find. Table 1.A18 shows that we keep on finding significant field effects on girls only, even when we compare boys and girls with similar relative strength in Mathematics. By contrast, Table 1.A18 does not show any

significant field effect on boys, regardless of their initial strength in science. Enrollment at a higher ranked school does not induce any decline in boys' probability to graduate in science, even when we focus on those who are relatively weak in Mathematics.

The most significant field effects are actually observed for girls in the third, and, to a lesser extent, in the second quartile of the distribution of relative strength in Mathematics. The weaker field effects on girls in the bottom and top quartiles are suggestive that students who are top achievers in a given subject (be it science or humanities) choose this subject regardless of the context, that is regardless of whether they are admitted into a higher ranked school, with higher-achieving peers, or not. The school context makes a difference mostly for students with an initially balanced academic profile.

Overall, the fact that admission at a higher ranked school has stronger effects on girls' choices than on boys' choices does not seem to be a mere consequence of the fact that girls are overrepresented among students who perform relatively poorly in Mathematics.

6.4 Teachers' Characteristics

Until now, our interpretations have focused on the fact that enrollment at a higher ranked school is associated with an increase in peer competition in science. But enrollment at a higher ranked school may be associated with other changes in the school environment. In particular, it may be associated with a change in who teaches science. There is evidence that a same-gender teacher may improve students' performance, especially female students' ones (see Rothstein, 1995; Bettinger and Long, 2005; Dee, 2007; Lim and Meer, 2017 for example). Our estimated field effects could be explained in part by the fact that enrollment at a higher ranked school is associated with a decrease in the relative proportion of women among science teachers, which could induce female students to choose alternative fields of study. To investigate this assumption we augmented our dataset with administrative information on teachers' gender (and age) in each school and field of study. We then looked at whether enrollment at a higher ranked high school was associated with a change in the proportion of women among science teachers, but we found no evidence of such change. As discussed in previous sections, teachers tend to be slightly older in higher ranked schools which is consistent with the fact that teachers' access to more sought-after schools depends mostly on their level of seniority. But the age shifts are similar for science and humanities teachers. When we focus on the sample of female students, the age shift is only about +0.7 year for science teachers and +0.3 year for humanities teachers, the difference between the two effects being statistically non-significant at standard levels (these results are reported in Appendix 1.A Table 1.A6). Overall, it seems difficult to relate our main findings to changes in

the age or gender composition of science teachers at the cut-off.

7 Conclusion

French students must choose a major field of study at the end of their first year of high school. This is a very important decision as it is a key determinant of the higher education programs to which they can get access after high school. The results reported in this paper suggest that peer competition has little effect on the field of study chosen by male students, but a very significant impact on the field of study chosen by female students. Specifically, we find that enrollment at the most sought-after Parisian high schools is associated with both a significant increase in peer competition (especially in Mathematics) and a significant decline in the proportion of female students who choose science as a major field of study. As it happens, many female students fly away from science, even though they would be able to graduate in science, were they assigned to different schools, with less competitive schoolmates. The effect is particularly strong for female students with an initially balanced academic profile.

In terms of policy implication, our results suggest that the process which assigns students to high schools is a potentially important determinant of the gender gap in science and its variation across ability groups. When an assignment process gives priority to students with higher academic ability, it likely increases ability segregation across schools and likely increases the proportion of female students who opt for humanities, even though they could succeed in science.

The region of Paris is the French region where districts include the largest number of high schools and where ability segregation across schools is maximal. One way to go could be to modify the design of the assignment process, so as to reduce the importance of middle school grades and give more weight to place of residence (with smaller districts and fewer schools per district, for instance). It would likely reduce ability segregation across Parisian schools and increase the correlation between students' rank within their school and their rank within the overall population. Eventually, it would likely reduce the influence of schoolmates on girls' choices. It should be emphasized, however, that such a reform would not necessarily be neutral in terms of students' overall performance. It may be that ability segregation across schools has a positive influence on students' average performance, if only because it makes possible to have more homogeneous schools and classes (see Booij, Leuven and Oosterbeek, 2017). More research is needed to assess how (and why) changes in the design of students' assignment process affect the relationships between gender, ability and

field of study.

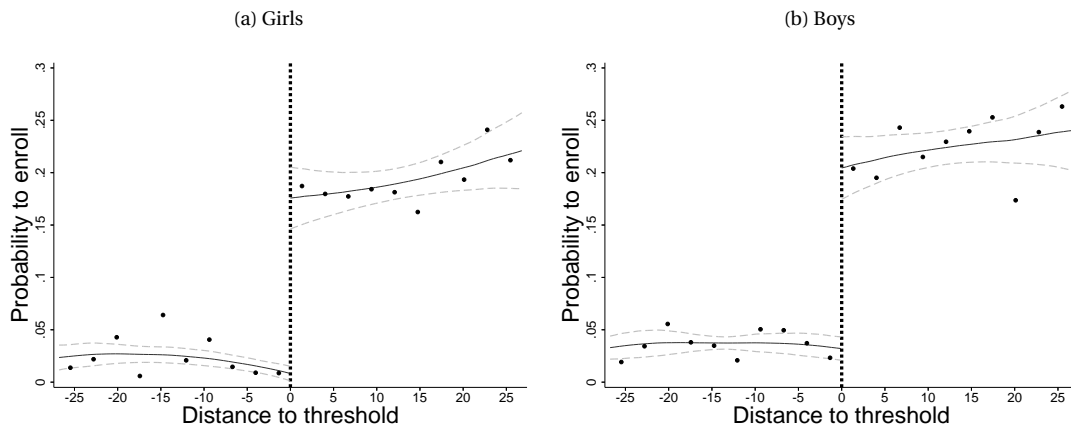


Figure 1.1 – Enrollment probability

Note: The figure refers to the sample of middle school students who were in their last year of middle school (9th grade) in 2009 or 2010, and whose 9th grade score fell either just below or just above the admission threshold of an oversubscribed high school. The bandwidth is computed following Calonico, Cattaneo and Titiunik (2014). The figure shows the probability to enroll into this school, plotted against the standardized distance to the threshold.

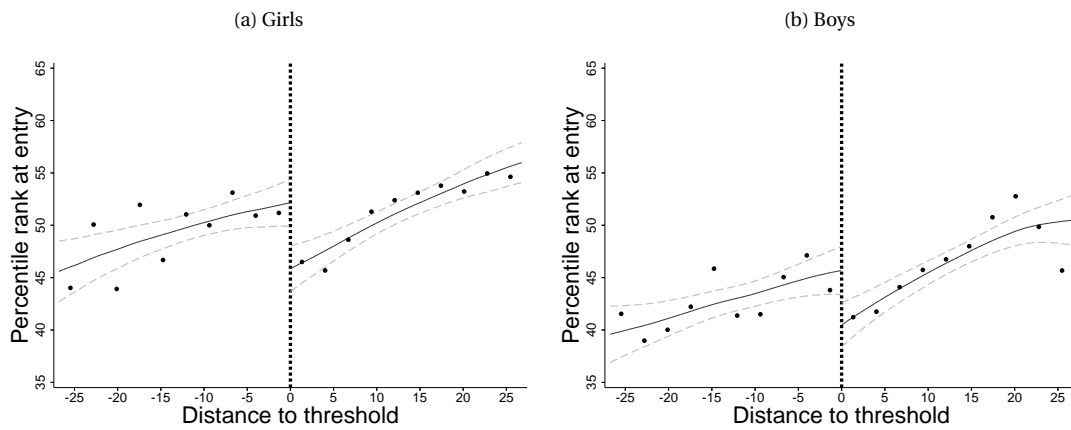


Figure 1.2 – Rank at entry

Note: Same sample as for Figure 1.1, restricted to students who enrolled in general education in a public high school in Paris. For each student in this sample, we computed her percentile rank in the distribution of 9th grade scores among students admitted into her high school. The figure shows average percentile ranks, plotted against the standardized distance to the threshold.

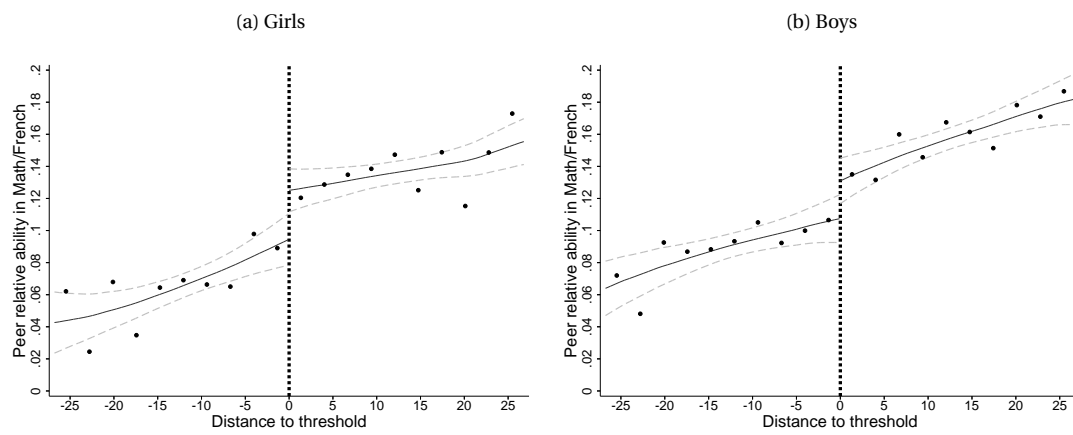


Figure 1.3 – Peer average relative mark Mathematics/French

Note: Same sample as for Figure 1.1, restricted to students who enrolled in general education in a public high school in Paris. For each student in this sample, we computed her 10th grade peers' relative mark in Mathematics/French (using results at end-of-middle school national examinations). The figure plots a standardized version of this measure of peers' relative ability in Mathematics, against the standardized distance to the threshold.

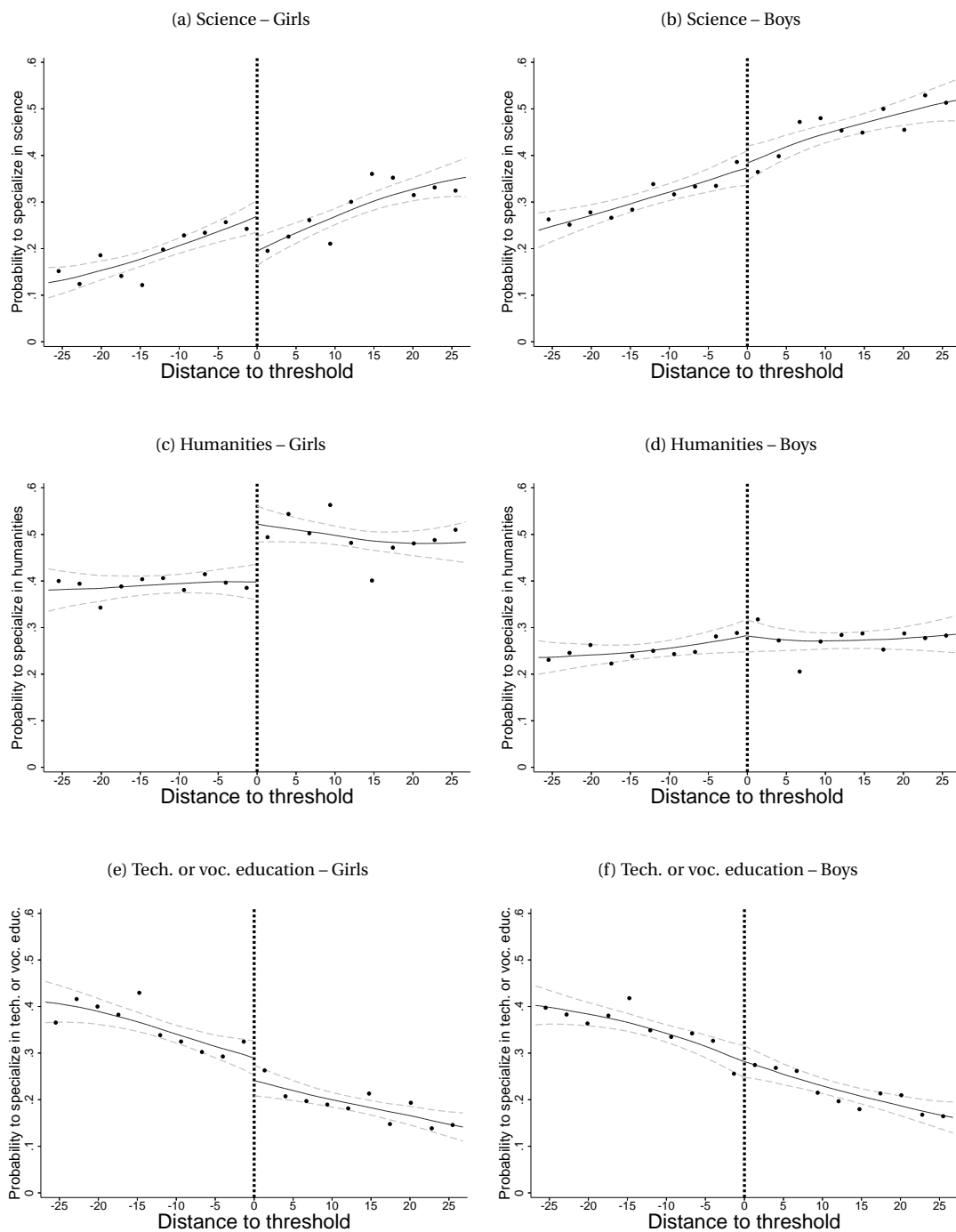


Figure 1.4 – Main field of study

Note: Same sample as for Figure 1.1. The figure shows the proportion of students who choose science as their major field of study (Figures 1.4a and 1.4b), humanities (Figures 1.4c and 1.4d), or technical/vocational education (Figures 1.4e and 1.4f), plotted against the standardized distance to the threshold.

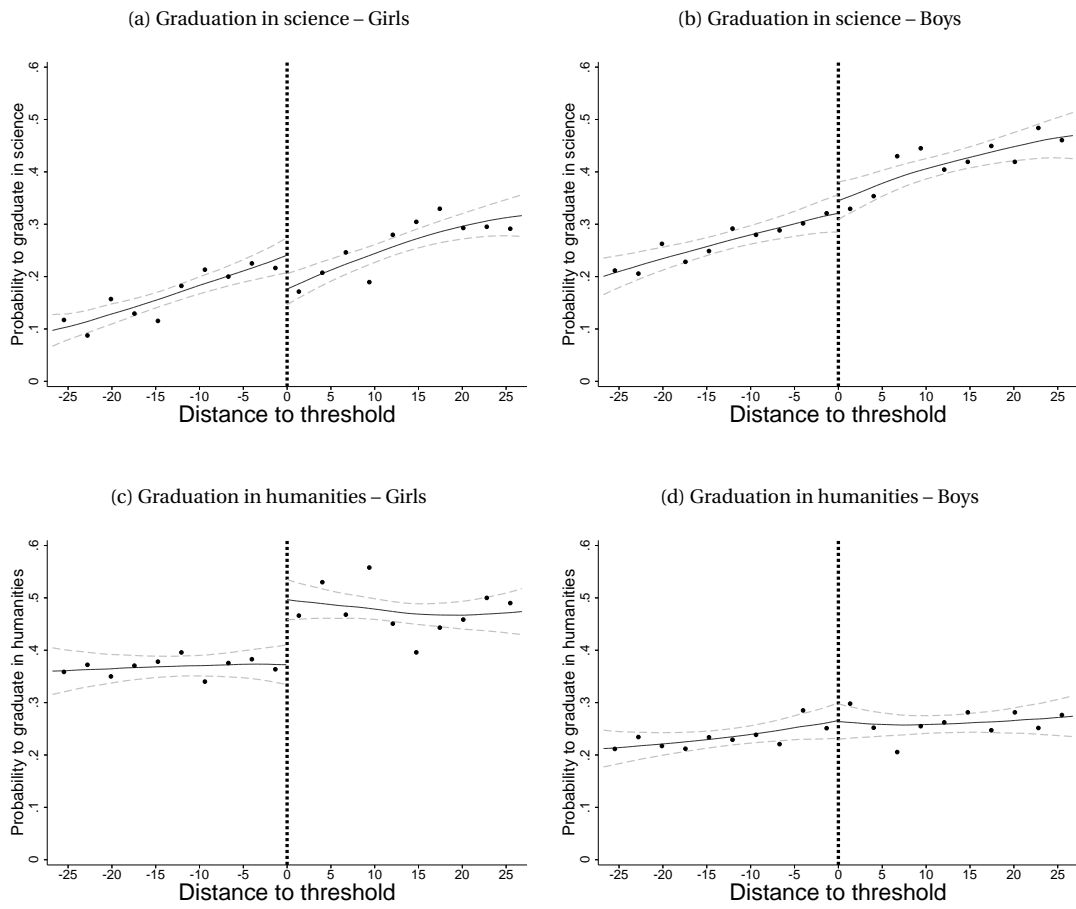


Figure 1.5 – Graduation

Note: Same sample as for Figure 1.1. The figure shows the proportion of students who graduate in science (Figures 1.5a and 1.5b), or in humanities (Figures 1.5c and 1.5d), plotted against the standardized distance to the threshold.

Table 1.1 – Type of school attended and characteristics of high school peers

	All (1)	Girls (2)	Boys (3)
<i>Panel A: Type of schools attended</i>			
Higher ranked high school	0.173** (0.016)	0.172** (0.021)	0.172** (0.021)
Private high school	-0.013 (0.010)	-0.027 (0.015)	0.001 (0.011)
Non-Parisian high school	0.002 (0.006)	0.004 (0.008)	0.000 (0.008)
Middle school (grade repetition)	0.006 (0.008)	-0.000 (0.013)	0.014 (0.011)
Dropout (after 9 th grade)	0.002 (0.008)	-0.004 (0.009)	0.008 (0.013)
N	7573	3691	3882
<i>Panel B: Characteristics of high school peers</i>			
Peers' average (9 th grade) score	0.066** (0.014)	0.080** (0.019)	0.051* (0.020)
Students' rank in the distribution of peers' scores	-5.9** (1.1)	-6.4** (1.5)	-5.3** (1.4)
Stud.' prob. to fall in the bottom quartile of peers' scores	0.140** (0.021)	0.144** (0.027)	0.136** (0.031)
Proportion of girls among peers	-0.003 (0.005)	0.005 (0.006)	-0.011 (0.007)
Proportion of low-income students among peers	-0.004 (0.003)	0.003 (0.005)	-0.010* (0.005)
Average mark in Mathematics	0.065** (0.014)	0.070** (0.019)	0.058** (0.020)
Average mark in French	0.047** (0.012)	0.056** (0.016)	0.039* (0.017)
Average relative mark Mathematics/French	0.026** (0.007)	0.027** (0.008)	0.024* (0.010)
N	5903	2937	2966

Note: Panel A refers to the sample of middle school students who were in their last year of middle school (9th grade) in 2009 or 2010, and whose 9th grade score fell either just below or just above the admission threshold of an oversubscribed public high school in Paris. Panel B refers to the same sample restricted to students who enrolled in general education in a public high school in Paris. For both panels, the bandwidth is computed following Calonico, Cattaneo and Titiunik (2014). Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable, namely enrollment at the higher ranked high school (row 1), enrollment at a private high school (row 2), enrollment at a non-Parisian high school (row 3), grade 9 repetition (row 4), and dropout after 9th grade (row 5). Row 6 corresponds to grade 10 peers' average ability. Row 7 corresponds to students' percentile rank within the distribution of their 10th grade high school peers, and row 8 shows students' probability to fall in the bottom quartile of their high school. For those three rows, ability is measured by the 9th grade score. Row 9 corresponds to the proportion of girls and row 10 to the proportion of low-income students among 10th grade peers. Rows 11 and 12 correspond to grade 10 peers' average ability in Mathematics (rows 11) and French (rows 12), and row 13 shows peers' relative ability in Mathematics and French, where relative ability is measured by the standardized ratio between peers' grades in Mathematics and French. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell corresponds to a specific regression. All regressions include controls for students' age, gender, family background, average marks in grade 9 as well as a full set of dummies indicating the nearest cut-off.

* significant at 5%. ** significant at 1%.

Table 1.2 – Choice of major field of study and performance on high school exit exams

	All (1)	Girls (2)	Boys (3)
<i>Main field of study</i>			
Science	-0.033 (0.020)	-0.072** (0.025)	-0.001 (0.026)
Humanities	0.050* (0.021)	0.110** (0.029)	-0.005 (0.027)
Technical or vocational education	-0.011 (0.016)	-0.035 (0.022)	0.015 (0.023)
Dropout (after 10 th grade)	-0.016 (0.013)	-0.015 (0.017)	-0.018 (0.020)
<i>Graduation</i>			
Graduation in science	-0.022 (0.019)	-0.061** (0.023)	0.012 (0.025)
Graduation in science with honors	-0.028 (0.016)	-0.056** (0.020)	-0.002 (0.021)
Graduation in humanities	0.049* (0.021)	0.109** (0.030)	-0.006 (0.027)
Graduation in humanities with honors	0.014 (0.015)	0.036 (0.025)	-0.011 (0.019)
Graduation (any field)	0.027 (0.017)	0.048 (0.024)	0.007 (0.025)
Graduation with honors (any field)	-0.014 (0.018)	-0.020 (0.026)	-0.013 (0.024)
Selective undergraduate program in science	-0.001 (0.010)	-0.002 (0.007)	0.001 (0.017)
Selective undergraduate program in humanities	0.006 (0.008)	0.018 (0.014)	-0.002 (0.010)
Selective undergraduate program (any field)	0.005 (0.013)	0.015 (0.016)	-0.001 (0.020)
N	7573	3691	3882

Note: Same sample of students as in Table 1.1 Panel A. Column 1 refers to the full sample, column 2 to the female subsample and column 3 to the male subsample. Each row corresponds to a specific dependent variable. The first three rows correspond to dependent variables describing major fields of study at the end of grade 10, namely science (row 1), humanities (row 2) or technical/vocational education (row 3). The fourth row corresponds to a variable indicating whether students drop out from education after 10th grade. The next six rows correspond to variables indicating graduation in science at the end of grade 12 (row 5), graduation in science with honors (row 6), graduation in humanities (row 7), graduation in humanities with honors (row 8), graduation in general education (row 9), and graduation in general education with honors (row 10). The three last rows corresponds to dependent variables indicating whether students gain access to a selective program in science (row 11), in humanities (row 12), or in any such selective undergraduate program (row 13) after high school. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Appendix

1.A Additional Tables and Figures

Table 1.A1 – Descriptive Statistics – Baseline characteristics

	All (1)	Girls (2)	Boys (3)
<i>Baseline academic level</i>			
Standardized results – teachers grades and national examinations	0.240	0.302	0.181
	[0.589]	[0.587]	[0.584]
Standardized results in French – teachers grades	0.172	0.320	0.031
	[0.718]	[0.701]	[0.706]
Standardized results in Mathematics – teachers grades	0.148	0.130	0.166
	[0.768]	[0.771]	[0.764]
Standardized results in French – national examinations	0.181	0.326	0.044
	[0.801]	[0.793]	[0.784]
Standardized results in Mathematics – national examinations	0.159	0.087	0.227
	[0.829]	[0.813]	[0.837]
<i>Demographic characteristics</i>			
Female students	0.487		
	[0.500]		
Age above 15 at the end of 9 th grade	0.166	0.143	0.187
	[0.372]	[0.350]	[0.390]
High socio-economic status	0.549	0.524	0.572
	[0.498]	[0.499]	[0.495]
N	7573	3691	3882

Note: Same sample of students as in Table 1.1 Panel A. Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific variable. The first five rows correspond to variables describing students' baseline academic level during and at the end of 9th grade, namely their standardized average mark for all subjects as assessed during 9th grade and at the national exams (row 1), their standardized average marks in French (row 2) and in Mathematics (row 3) during 9th grade, and their standardized average marks in French (row 4) and in Mathematics (row 5) at the national exams. Rows 6 to 8 describe students' baseline demographic characteristics, that is whether they are female students (row 6, column 1), whether they were older than 15 at the end of 9th grade (row 7), and whether the household head has a high socio-economic status occupation (row 8). For each variable and each sample, the table reports the average of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 1.A2 – Descriptive Statistics – Type of school attended and characteristics of high school peers

	All (1)	Girls (2)	Boys (3)
<i>Panel A: Type of schools attended</i>			
Higher ranked high school	0.120 [0.325]	0.110 [0.313]	0.129 [0.336]
Private high school	0.044 [0.204]	0.045 [0.207]	0.043 [0.202]
Non-Parisian high school	0.017 [0.127]	0.017 [0.129]	0.016 [0.126]
Middle school (grade repetition)	0.048 [0.213]	0.040 [0.197]	0.055 [0.228]
Dropout (after 9 th grade)	0.022 [0.146]	0.017 [0.128]	0.027 [0.161]
N	7573	3691	3882
<i>Panel B: Characteristics of high school peers</i>			
Peers' average (9 th grade) score	0.360 [0.395]	0.375 [0.396]	0.345 [0.392]
Students' rank in the distribution of peers' scores	47.7 (25.4)	50.6 (25.0)	44.8 (25.5)
Stud.' prob. to fall in the bottom quartile of peers' scores	0.226 [0.418]	0.183 [0.387]	0.269 [0.444]
Proportion of girls among peers	0.521 [0.100]	0.539 [0.088]	0.504 [0.108]
Proportion of low-income students among peers	0.132 [0.091]	0.135 [0.093]	0.128 [0.088]
Average mark in Mathematics	0.307 [0.407]	0.307 [0.411]	0.306 [0.403]
Average mark in French	0.307 [0.337]	0.322 [0.334]	0.293 [0.339]
Average relative mark Mathematics/French	0.116 [0.169]	0.107 [0.170]	0.126 [0.168]
N	5903	2937	2966

Note: Same sample of students as in Table 1.1. Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific variable, namely enrollment at the higher ranked high school (row 1), enrollment at a private high school (row 2), enrollment at a non-Parisian high school (row 3), grade 9 repetition (row 4), and dropout after 9th grade (row 5). Row 6 corresponds to grade 10 peers' average ability. Row 7 corresponds to students' percentile rank within the distribution of their 10th grade high school peers, and row 8 shows students' probability to fall in the bottom quartile of their high school. For those three rows, ability is measured by the 9th grade score. Row 9 corresponds to the proportion of girls and row 10 to the proportion of low-income students among 10th grade peers. Rows 11 and 12 correspond to grade 10 peers' average ability in Mathematics (rows 11) and French (rows 12), and row 13 shows peers' relative ability in Mathematics and French, where relative ability is measured by the standardized ratio between peers' grades in Mathematics and French. For each variable and each sample, the table reports the average of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 1.A3 – Descriptive Statistics – Characteristics of high schools attended

	All (1)	Girls (2)	Boys (3)
Prop. of female teachers	0.600 [0.086]	0.610 [0.078]	0.590 [0.092]
Prop. of female teachers in science	0.513 [0.126]	0.519 [0.126]	0.507 [0.125]
Prop. of female teachers in humanities	0.700 [0.131]	0.701 [0.129]	0.700 [0.132]
Mean age of teachers	47.576 [2.811]	47.546 [2.845]	47.606 [2.778]
Mean age of science teachers	46.172 [3.227]	46.060 [3.267]	46.283 [3.183]
Mean age of humanities teachers	50.762 [3.402]	50.822 [3.380]	50.703 [3.424]
Number of teachers per student	0.100 [0.015]	0.099 [0.014]	0.101 [0.016]
School size (nb. of 10 th graders)	272 [92]	276 [91]	268 [93]
Distance to middle school (meters)	1590 [1531]	1627 [1541]	1555 [1521]
N	5903	2937	2966
Average graduation rate	0.871 [0.117]	0.874 [0.118]	0.867 [0.117]
N	5732	2828	2904
Average graduation rate in science	0.852 [0.137]	0.855 [0.136]	0.849 [0.137]
N	5598	2762	2836
Average graduation rate in humanities	0.891 [0.114]	0.891 [0.115]	0.890 [0.113]
N	5525	2802	2723

Note: Same sample of students as in Table 1.1 Panel B. Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific variable. Row 1 to 6 correspond to high school teachers' characteristics, namely the proportion of female teachers (row 1), the proportion of female teachers in science (row 2), and the proportion of female teachers in humanities (row 3). Row 4 shows the proportion of teachers aged 40 or more, row 5 the mean age of science teachers, and row 6 the mean age of humanities teachers. Row 7 corresponds to the number of teachers per student, and row 8 to the number of students in 10th grade. Row 9 corresponds to the distance between students' middle and high schools. Rows 10 to 12 correspond to the students' high school graduation results. For each variable and each sample, the table reports the average of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 1.A4 – Descriptive Statistics – Choice of major field of study and performance on high school exit exams

	All (1)	Girls (2)	Boys (3)
<i>Main field of study</i>			
Science	0.314 [0.464]	0.242 [0.428]	0.382 [0.486]
Humanities	0.353 [0.478]	0.448 [0.497]	0.264 [0.441]
Technical or vocational education	0.276 [0.447]	0.266 [0.442]	0.286 [0.452]
Dropout (after 10 th grade)	0.089 [0.285]	0.078 [0.268]	0.100 [0.300]
<i>Graduation</i>			
Graduation in science	0.281 [0.449]	0.217 [0.412]	0.342 [0.474]
Graduation in science with honors	0.163 [0.369]	0.121 [0.326]	0.202 [0.402]
Graduation in humanities	0.335 [0.472]	0.427 [0.495]	0.248 [0.432]
Graduation in humanities with honors	0.175 [0.380]	0.229 [0.420]	0.123 [0.329]
Graduation (any field)	0.616 [0.486]	0.644 [0.479]	0.590 [0.492]
Graduation with honors (any field)	0.338 [0.473]	0.350 [0.477]	0.326 [0.469]
Selective undergraduate program in science	0.048 [0.213]	0.020 [0.140]	0.074 [0.262]
Selective undergraduate program in humanities	0.047 [0.211]	0.056 [0.230]	0.038 [0.192]
Selective undergraduate program (any field)	0.095 [0.293]	0.076 [0.265]	0.113 [0.317]
N	7573	3691	3882

Note: Same sample of students as in Table 1.1 Panel A. Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific variable. The first three rows correspond to dependent variables describing the possible choices of major fields of study at the end of grade 10, namely science (row 1), humanities (row 2) or technical/vocational education (row 3). The fourth row corresponds to a dependent variable indicating whether students drop out from education after 10th grade. The next six rows correspond to variables indicating graduation in science at the end of grade 12 (row 5), graduation in science with honors (row 6), graduation in humanities (row 7), graduation in humanities with honors (row 8), graduation in general education (row 9), and graduation in general education with honors (row 10). The three last rows corresponds to dependent variables indicating whether students gain access to a selective program in science (row 11), in humanities (row 12), or in any such selective undergraduate program (row 13) after high school. For each variable and each sample, the table reports the mean of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 1.A5 – Balance checks

	All (1)	Girls (2)	Boys (3)
<i>Baseline academic level</i>			
Standardized results – teachers grades and national examinations	-0.002 (0.004)	-0.004 (0.007)	-0.001 (0.006)
Standardized results in French – teachers grades	0.022 (0.020)	0.027 (0.026)	0.023 (0.031)
Standardized results in Mathematics – teachers grades	0.011 (0.020)	-0.018 (0.027)	0.038 (0.029)
Standardized results in French – national examinations	0.030 (0.020)	0.030 (0.030)	0.026 (0.027)
Standardized results in Mathematics – national examinations	0.010 (0.016)	0.018 (0.024)	0.003 (0.024)
<i>Demographic characteristics</i>			
Female students	-0.026 (0.023)		
Age above 15 at the end of 9 th grade	0.003 (0.015)	0.005 (0.022)	-0.003 (0.022)
High socio-economic status	0.002 (0.020)	-0.025 (0.031)	0.026 (0.028)
N	7573	3691	3882

Note: Same sample of students as in Table 1.1 Panel A. Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable. The first five rows correspond to variables describing students' baseline academic level during and at the end of 9th grade, namely their standardized average mark for all subjects as assessed during 9th grade and at the national exams (row 1), their standardized average marks in French (row 2) and in Mathematics (row 3) during 9th grade, and their standardized average marks in French (row 4) and in Mathematics (row 5) at the national exams. Rows 6 to 8 describe students' baseline demographic characteristics, that is whether they are female students (row 6, column 1), whether they were older than 15 at the end of 9th grade (row 7), and whether the household head has a high socio-economic status occupation (row 8). For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at individual level are in parenthesis. Each cell of the table corresponds to a specific regression.

* significant at 5%. ** significant at 1%.

Table 1.A6 – Characteristics of high schools attended

	All (1)	Girls (2)	Boys (3)
Prop. of female teachers	-0.007 (0.005)	-0.004 (0.006)	-0.010 (0.007)
Prop. of female teachers in science	-0.008 (0.006)	0.002 (0.009)	-0.016 (0.009)
Prop. of female teachers in humanities	0.005 (0.007)	0.004 (0.009)	0.007 (0.009)
Mean age of teachers	0.458** (0.121)	0.624** (0.161)	0.262 (0.168)
Mean age of science teachers	0.518** (0.142)	0.737** (0.191)	0.280 (0.204)
Mean age of humanities teachers	0.330* (0.148)	0.297 (0.196)	0.323 (0.230)
Number of teachers per student	-0.000 (0.001)	-0.001 (0.001)	0.000 (0.001)
School size (nb. of 10 th graders)	1.4 (4.8)	5.9 (6.8)	-3.0 (6.5)
Distance to middle school (meters)	0.9 (78.4)	-90.8 (116.4)	75.6 (106.3)
N	5903	2937	2966
Average graduation rate	0.012** (0.004)	0.014* (0.007)	0.009 (0.006)
N	5732	2828	2904
Average graduation rate in science	0.020** (0.006)	0.024** (0.008)	0.015 (0.008)
N	5598	2762	2836
Average graduation rate in humanities	0.008 (0.004)	0.010 (0.007)	0.007 (0.006)
N	5525	2802	2723

Note: Same sample of students as in Table 1.1 Panel B. Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable. Row 1 to 6 correspond to high school teachers' characteristics, namely the proportion of female teachers (row 1), the proportion of female teachers in science (row 2), and the proportion of female teachers in humanities (row 3). Row 4 shows the proportion of teachers aged 40 or more, row 5 the mean age of science teachers, and row 6 the mean age of humanities teachers. Row 7 corresponds to the number of teachers per student, and row 8 to the number of students in 10th grade. Row 9 corresponds to the distance between students' middle and high schools. Rows 10 to 12 correspond to the students' high school graduation results. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A7 – Local average treatment effects – Type of school attended and characteristics of high school peers

	All (1)	Girls (2)	Boys (3)
<i>Panel A: Type of schools attended</i>			
Private high school	-0.072 (0.056)	-0.155 (0.086)	0.008 (0.066)
Non-Parisian high school	0.010 (0.032)	0.024 (0.046)	0.001 (0.045)
Middle school (grade repetition)	0.037 (0.043)	-0.001 (0.073)	0.084 (0.067)
Dropout (after 9 th grade)	0.013 (0.048)	-0.024 (0.053)	0.045 (0.076)
N	7573	3691	3882
<i>Panel B: Characteristics of high school peers</i>			
Peers' average (9 th grade) score	0.305** (0.060)	0.379** (0.084)	0.230** (0.082)
Students' rank in the distribution of peers' scores	-27.4** (4.2)	-30.3** (5.9)	-23.9** (5.7)
Stud.' prob. to fall in the bottom quartile of peers' scores	0.647** (0.089)	0.678** (0.107)	0.618** (0.131)
Proportion of girls among peers	-0.015 (0.023)	0.024 (0.029)	-0.048 (0.034)
Proportion of low-income students among peers	-0.017 (0.016)	0.013 (0.024)	-0.046* (0.023)
Average mark in Mathematics	0.300** (0.061)	0.329** (0.085)	0.265** (0.085)
Average mark in French	0.217** (0.052)	0.265** (0.069)	0.177* (0.071)
Average relative mark Mathematics/French	0.122** (0.029)	0.125** (0.038)	0.108* (0.044)
N	5903	2937	2966

Note: Same sample of students as in Table 1.1. Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable, namely enrollment at the higher ranked high school (row 1), enrollment at a private high school (row 2), enrollment at a non-Parisian high school (row 3), grade 9 repetition (row 4), and dropout after 9th grade (row 5). Row 6 corresponds to grade 10 peers' average ability. Row 7 corresponds to students' percentile rank within the distribution of their 10th grade high school peers, and row 8 shows students' probability to fall in the bottom quartile of their high school. For those three rows, ability is measured by the 9th grade score. Row 9 corresponds to the proportion of girls and row 10 to the proportion of low-income students among 10th grade peers. Rows 11 and 12 correspond to grade 10 peers' average ability in Mathematics (rows 11) and French (rows 12), and row 13 shows peers' relative ability in Mathematics and French, where relative ability is measured by the standardized ratio between peers' grades in Mathematics and French. For each dependent variable and each sample, the table reports the estimated effect of enrollment at the higher ranked high school instrumented by a dummy indicating eligibility for enrollment at this high school. Controls are the same as in Table 1.1. Estimates use a triangular kernel, and the bandwidth is computed following Calonico, Cattaneo and Titiunik (2014). Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A8 – Local average treatment effects – Characteristics of high schools attended

	All (1)	Girls (2)	Boys (3)
Prop. of female teachers	-0.034 (0.021)	-0.017 (0.028)	-0.045 (0.030)
Prop. of female teachers in science	-0.036 (0.030)	0.008 (0.041)	-0.074 (0.042)
Prop. of female teachers in humanities	0.024 (0.031)	0.017 (0.042)	0.031 (0.040)
Mean age of teachers	2.118** (0.548)	2.949** (0.738)	1.190 (0.760)
Mean age of science teachers	2.397** (0.656)	3.482** (0.882)	1.268 (0.920)
Mean age of humanities teachers	1.527* (0.661)	1.405 (0.908)	1.465 (1.031)
Number of teachers per student	-0.001 (0.004)	-0.005 (0.005)	0.002 (0.005)
School size (nb. of 10 th graders)	6.3 (21.8)	28.1 (31.9)	-13.6 (29.5)
Distance to middle school (meters)	4.1 (359.6)	-428.9 (533.9)	342.3 (486.4)
N	5903	2937	2966
Average graduation rate	0.055** (0.019)	0.065* (0.029)	0.042 (0.027)
N	5732	2828	2904
Average graduation rate in science	0.090** (0.023)	0.106** (0.034)	0.068* (0.034)
N	5598	2762	2836
Average graduation rate in humanities	0.036 (0.019)	0.044 (0.030)	0.029 (0.025)
N	5525	2802	2723

Note: Same sample of students as in Table 1.1 Panel B. Column 1 refers to the full sample, column 2 to the subsample of female students and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable. Row 1 to 6 correspond to high school teachers' characteristics, namely the proportion of female teachers (row 1), the proportion of female teachers in science (row 2), and the proportion of female teachers in humanities (row 3). Row 4 shows the proportion of teachers aged 40 or more, row 5 the mean age of science teachers, and row 6 the mean age of humanities teachers. Row 7 corresponds to the number of teachers per student, and row 8 to the number of students in 10th grade. Row 9 corresponds to the distance between students' middle and high schools. Rows 10 to 12 correspond to the students' high school graduation results. For each dependent variable and each sample, the table reports the estimated effect of enrollment at the higher ranked high school instrumented by a dummy indicating eligibility for enrollment at this high school. Controls are the same as in Table 1.1. Estimates use a triangular kernel, and the bandwidth is computed following Calonico, Cattaneo and Titiunik (2014). Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A9 – Local average treatment effects – Choice of major field of study and performance on high school exit exams

	All (1)	Girls (2)	Boys (3)
<i>Main field of study</i>			
Science	-0.189 (0.114)	-0.421** (0.152)	-0.007 (0.150)
Humanities	0.290* (0.119)	0.641** (0.178)	-0.030 (0.153)
Technical or vocational education	-0.065 (0.091)	-0.202 (0.126)	0.087 (0.133)
Dropout (after 10 th grade)	-0.090 (0.075)	-0.088 (0.101)	-0.106 (0.115)
<i>Graduation</i>			
Graduation in science	-0.129 (0.108)	-0.352* (0.138)	0.071 (0.140)
Graduation in science with honors	-0.160 (0.092)	-0.324** (0.120)	-0.014 (0.123)
Graduation in humanities	0.284* (0.120)	0.632** (0.180)	-0.034 (0.156)
Graduation in humanities with honors	0.078 (0.085)	0.207 (0.147)	-0.063 (0.110)
Graduation (any field)	0.155 (0.096)	0.280* (0.142)	0.038 (0.142)
Graduation with honors (any field)	-0.082 (0.103)	-0.117 (0.149)	-0.077 (0.141)
Selective undergraduate program in science	-0.003 (0.055)	-0.014 (0.041)	0.004 (0.098)
Selective undergraduate program in humanities	0.034 (0.048)	0.102 (0.080)	-0.010 (0.059)
Selective undergraduate program (any field)	0.030 (0.076)	0.088 (0.093)	-0.008 (0.113)
N	7573	3691	3882

Note: Same sample of students as in Table 1.1 Panel A. Column 1 refers to the full sample, column 2 to the female subsample and column 3 to the male subsample. Each row corresponds to a specific dependent variable. The first three rows correspond to dependent variables describing major fields of study at the end of grade 10, namely science (row 1), humanities (row 2) or technical/vocational education (row 3). The fourth row corresponds to a variable indicating whether students drop out from education after 10th grade. The next six rows correspond to variables indicating graduation in science at the end of grade 12 (row 5), graduation in science with honors (row 6), graduation in humanities (row 7), graduation in humanities with honors (row 8), graduation in general education (row 9), and graduation in general education with honors (row 10). The three last rows corresponds to dependent variables indicating whether students gain access to a selective program in science (row 11), in humanities (row 12), or in any such selective undergraduate program (row 13) after high school. For each dependent variable and each sample, the table reports the estimated effect of enrollment at the higher ranked high school instrumented by a dummy indicating eligibility for enrollment at this high school. Controls are the same as in Table 1.1. Estimates use a triangular kernel, and the bandwidth is computed following Calonico, Cattaneo and Titiunik (2014). Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A10 – Falsification test – Low-income students

	All (1)	Girls (2)	Boys (3)
<i>First stage</i>			
Enrollment at a higher ranked high school	-0.03 (0.02)	-0.04 (0.02)	-0.00 (0.03)
<i>Main field of study</i>			
Science	-0.01 (0.03)	0.00 (0.03)	-0.01 (0.04)
Humanities	-0.01 (0.03)	-0.01 (0.04)	-0.01 (0.04)
<i>Graduation</i>			
Graduation in science	-0.02 (0.02)	-0.00 (0.03)	-0.02 (0.04)
Graduation in humanities	0.01 (0.03)	0.00 (0.04)	0.01 (0.04)
Graduation (any field)	-0.01 (0.03)	0.00 (0.04)	-0.02 (0.04)
Selective undergraduate program in science	-0.00 (0.01)	-0.01 (0.01)	0.00 (0.01)
Selective undergraduate program in humanities	0.00 (0.01)	0.00 (0.01)	-0.01 (0.01)
Selective undergraduate program (any field)	-0.00 (0.01)	-0.00 (0.01)	-0.01 (0.02)
N	2157	1162	995

Note: The table refers to the sample of low-income middle school students who were in their last year of middle school (9th grade) in 2009 or 2010, and whose subsequent 9th grade score fell either just below or just above one of the admission thresholds estimated for non low-income students. Column 1 refers to the full sample, column 2 to the subsample of female students, and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable, and for each dependent variable the table shows the replication of the previous regression analysis. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A11 – Falsification test – Students from private middle schools

	All (1)	Girls (2)	Boys (3)
<i>First stage</i>			
Enrollment at a higher ranked high school	-0.00 (0.00)	-0.00 (0.00)	-0.00 (0.00)
<i>Main field of study</i>			
Science	0.01 (0.02)	0.01 (0.04)	0.01 (0.03)
Humanities	-0.01 (0.03)	-0.04 (0.04)	0.01 (0.03)
<i>Graduation</i>			
Graduation in science	0.02 (0.02)	0.03 (0.03)	0.01 (0.03)
Graduation in humanities	-0.01 (0.03)	-0.03 (0.04)	0.00 (0.03)
Graduation (any field)	0.01 (0.02)	0.00 (0.03)	0.02 (0.03)
Selective undergraduate program in science	0.01 (0.01)	0.01 (0.01)	0.01 (0.02)
Selective undergraduate program in humanities	-0.01 (0.01)	-0.01 (0.02)	-0.01 (0.02)
Selective undergraduate program (any field)	-0.00 (0.02)	-0.01 (0.02)	0.00 (0.03)
N	5144	2516	2628

Note: The table refers to the sample of middle school students who were in 9th grade in a private middle school in 2009 or 2010, and whose subsequent 9th grade score fell either just below or just above one of the admission thresholds estimated for students in public middle schools. Column 1 refers to the full sample, column 2 to the subsample of female students, and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable, and for each dependent variable the table shows the replication of the previous regression analysis. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A12 – Regression analysis by cohort

	All		Girls		Boys	
	2009 (1)	2010 (2)	2009 (3)	2010 (4)	2009 (5)	2010 (6)
<i>First stage</i>						
Enrollment at a higher ranked high school	0.144** (0.018)	0.194** (0.024)	0.136** (0.023)	0.202** (0.033)	0.153** (0.027)	0.182** (0.031)
<i>Main field of study</i>						
Science	-0.040 (0.031)	-0.027 (0.025)	-0.067 (0.039)	-0.074* (0.032)	-0.029 (0.038)	0.023 (0.036)
Humanities	0.059 (0.034)	0.043 (0.024)	0.108* (0.044)	0.113** (0.038)	0.023 (0.042)	-0.033 (0.033)
<i>Graduation</i>						
Graduation in science	-0.021 (0.029)	-0.025 (0.024)	-0.046 (0.035)	-0.072* (0.030)	-0.006 (0.035)	0.028 (0.035)
Graduation in humanities	0.045 (0.033)	0.054* (0.026)	0.094* (0.045)	0.124** (0.039)	0.008 (0.042)	-0.020 (0.035)
Graduation (any field)	0.024 (0.027)	0.029 (0.021)	0.047 (0.035)	0.052 (0.034)	0.003 (0.040)	0.008 (0.031)
Selective undergraduate program in science	0.007 (0.012)	-0.008 (0.015)	0.001 (0.011)	-0.005 (0.010)	0.008 (0.022)	-0.007 (0.025)
Selective undergraduate program in humanities	-0.007 (0.010)	0.015 (0.014)	-0.005 (0.015)	0.035 (0.022)	-0.001 (0.014)	-0.005 (0.016)
Selective undergraduate program (any field)	0.001 (0.016)	0.007 (0.021)	-0.005 (0.019)	0.030 (0.025)	0.006 (0.026)	-0.013 (0.029)
N	3784	3789	1794	1897	1990	1892

Note: Same working sample as in Table 1.1 Panel A. We divided each working sample into two subsamples according to whether students were in 9th grade in 2009 (columns 1 and 3), or in 2010 (columns 2 and 4). Each row corresponds to a specific dependent variable, and for each dependent variable the table shows the replication of the previous regression analysis separately on the two subsamples. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A13 – Heterogeneous effects

<i>Panel A: Heterogeneity by SES</i>				
	Girls		Boys	
	High SES (1)	Low SES (2)	High SES (3)	Low SES (4)
Enrollment	0.186** (0.029)	0.149** (0.025)	0.232** (0.028)	0.089** (0.026)
Science	-0.104** (0.036)	-0.043 (0.031)	-0.021 (0.037)	0.034 (0.036)
Humanities	0.122** (0.040)	0.098* (0.040)	0.013 (0.038)	-0.044 (0.039)
N	1935	1756	2219	1663

<i>Panel B: Heterogeneity by baseline ability</i>				
	Girls		Boys	
	High ability (1)	Low ability (2)	High ability (3)	Low ability (4)
Enrollment	0.213** (0.036)	0.151** (0.024)	0.196** (0.040)	0.165** (0.024)
Science	-0.136** (0.045)	-0.047 (0.030)	-0.012 (0.049)	-0.002 (0.031)
Humanities	0.130** (0.049)	0.109** (0.036)	-0.087 (0.051)	0.029 (0.033)
N	1375	2316	1064	2818

Note: Same working sample as in Table 1.1 Panel A. In this table, for Panel A, we divided each working sample into two subsamples according to whether students came from a high socio-economic environment (columns 1 and 3), or from a low socio-economic environment (columns 2 and 4). For Panel B, we divided each working sample into two subsamples according to whether the closest cut-off is located in the top tercile (columns 1 and 3), or the second or third terciles (columns 2 and 4). Each row corresponds to a specific dependent variable, and for each dependent variable the table shows the replication of the previous regression analysis. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.
 * significant at 5%. ** significant at 1%.

Table 1.A14 – Robustness checks – Imbens and Kalyanaraman optimal bandwidth

	All (1)	Girls (2)	Boys (3)
<i>First stage</i>			
Enrollment at a higher ranked high school	0.172** (0.015)	0.170** (0.020)	0.171** (0.020)
<i>Main field of study</i>			
Science	-0.031 (0.019)	-0.069** (0.023)	-0.000 (0.025)
Humanities	0.048* (0.020)	0.105** (0.028)	-0.005 (0.026)
<i>Graduation</i>			
Graduation in science	-0.022 (0.018)	-0.058** (0.022)	0.012 (0.023)
Graduation in humanities	0.047* (0.020)	0.104** (0.028)	-0.007 (0.026)
Graduation (any field)	0.025 (0.016)	0.046 (0.023)	0.006 (0.024)
Selective undergraduate program in science	-0.001 (0.009)	-0.004 (0.007)	0.001 (0.016)
Selective undergraduate program in humanities	0.006 (0.008)	0.018 (0.013)	-0.002 (0.010)
Selective undergraduate program (any field)	0.004 (0.013)	0.014 (0.015)	-0.002 (0.019)
N	8202	3978	4224

Note: Similar working sample as in Table 1.1 Panel A. The only difference is that we use a larger bandwidth. Column 1 refers to the full sample, column 2 to the subsample of female students, and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable, and for each dependent variable the table shows the replication of the previous regression analysis. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A15 – Robustness checks – No controls

	All (1)	Girls (2)	Boys (3)
<i>First stage</i>			
Enrollment at a higher ranked high school	0.177** (0.016)	0.175** (0.021)	0.176** (0.021)
<i>Main field of study</i>			
Science	-0.024 (0.022)	-0.076** (0.026)	0.015 (0.030)
Humanities	0.055* (0.023)	0.121** (0.031)	0.001 (0.030)
<i>Graduation</i>			
Graduation in science	-0.014 (0.021)	-0.065** (0.024)	0.028 (0.028)
Graduation in humanities	0.054* (0.023)	0.119** (0.030)	-0.000 (0.030)
Graduation (any field)	0.041* (0.018)	0.054* (0.026)	0.027 (0.026)
Selective undergraduate program in science	0.002 (0.010)	-0.003 (0.007)	0.006 (0.017)
Selective undergraduate program in humanities	0.008 (0.008)	0.019 (0.013)	-0.000 (0.010)
Selective undergraduate program (any field)	0.010 (0.013)	0.016 (0.016)	0.005 (0.020)
N	7640	3720	3920

Note: Same working sample as in Table 1.1 Panel A. Column 1 refers to the full sample, column 2 to the subsample of female students, and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable, and for each dependent variable the table shows the replication of the previous regression analysis. Each cell of the table corresponds to a specific regression where we control only for a full set of dummies indicating the high school that corresponds to the nearest cut-off (no controls for students' baseline characteristics). For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis.

* significant at 5%. ** significant at 1%.

Table 1.A16 – Robustness checks – Linear, quadratic and cubic spline functions

	All (1)	Girls (2)	Boys (3)
<i>Panel A: Linear spline function</i>			
Enrollment at a higher ranked high school	0.167** (0.016)	0.161** (0.020)	0.170** (0.020)
Specialization in science	-0.022 (0.017)	-0.055* (0.022)	0.004 (0.024)
Specialization in humanities	0.039* (0.020)	0.091** (0.027)	-0.008 (0.026)
<i>Panel B: Quadratic spline function</i>			
Enrollment at a higher ranked high school	0.184** (0.019)	0.186** (0.028)	0.175** (0.027)
Specialization in science	-0.051 (0.026)	-0.098** (0.035)	-0.012 (0.035)
Specialization in humanities	0.068** (0.026)	0.141** (0.038)	0.004 (0.034)
<i>Panel C: Cubic spline function</i>			
Enrollment at a higher ranked high school	0.182** (0.018)	0.181** (0.025)	0.178** (0.024)
Specialization in science	-0.044 (0.023)	-0.090** (0.030)	-0.006 (0.031)
Specialization in humanities	0.060* (0.024)	0.129** (0.035)	-0.003 (0.031)
N	7573	3691	3882

Note: Same working sample as in Table 1.1 Panel A. The only difference is that we control for a linear spline function (Panel A), or for a quadratic spline function (Panel B), or for a cubic spline function (Panel C) of the running variable with a uniform kernel. Column 1 refers to the full sample, column 2 to the subsample of female students, and column 3 to the subsample of male students. Each row corresponds to a specific dependent variable, and for each dependent variable the table shows the replication of the previous regression analysis. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold. Standard errors clustered at the school and cohort level are in parenthesis. Each cell of the table corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A17 – Candidate mechanism: rank consideration

	All (1)	Girls (2)	Boys (3)
Expected rank in science	-2.8** (1.0)	-2.9* (1.4)	-3.0* (1.3)
Mean of Y	27.9	27.7	28.0
Expected rank in humanities	-3.2** (1.1)	-4.1** (1.4)	-2.5 (1.5)
Mean of Y	44.5	46.5	42.4
Diff. in expected ranking (S-H)	0.4 (0.8)	1.2 (0.9)	-0.5 (1.1)
Mean of Y	-16.6	-18.8	-14.4
N	4810	2461	2349

Similar working sample as in Table 1.1 Panel B, restricted to students' who enrolled in a public high school where there is both a humanity and a science track. Each row corresponds to a specific dependent variable. Row 1 shows student's percentile rank within the distribution of their 11th grade high school peers in science, where ability is measured using standardized marks in 9th grade with similar weights as for the high school exit exam in science. Row 2 shows student's percentile rank within the distribution of their 11th grade high school peers in humanities, where ability is measured using standardized marks in 9th grade with similar weights as for the high school exit exam in humanity. Row 3 corresponds to students difference in expected ranking between the scientific and humanity specializations. For each dependent variable and each sample, the table shows the impact of falling just above the admission threshold as well as the mean of the dependent variable. Standard errors clustered at the school and cohort level are in parenthesis. Each cell corresponds to a specific regression. All regressions include the same controls as in Table 1.1.

* significant at 5%. ** significant at 1%.

Table 1.A18 – Candidate mechanism: comparative advantages and capacity constraints

	Q1 (1)	Q2 (2)	Q3 (3)	Q4 (4)
<i>Enrollment at a higher ranked high school</i>				
eligibility for enrollment	0.174** (0.034)	0.151** (0.035)	0.244** (0.035)	0.146** (0.034)
eligibility for enrollment#female students	0.036 (0.031)	-0.024 (0.028)	-0.043 (0.033)	-0.043 (0.032)
<i>Specialization in science</i>				
eligibility for enrollment	-0.021 (0.025)	0.030 (0.042)	0.002 (0.049)	-0.070 (0.042)
eligibility for enrollment#female students	-0.012 (0.026)	-0.048 (0.036)	-0.143** (0.045)	0.039 (0.047)
<i>Specialization in humanities</i>				
eligibility for enrollment	0.099 (0.051)	-0.023 (0.050)	0.010 (0.048)	0.015 (0.038)
eligibility for enrollment#female students	-0.009 (0.046)	0.087 (0.046)	0.149** (0.042)	-0.045 (0.053)
<i>Graduation in science</i>				
eligibility for enrollment	-0.012 (0.022)	0.040 (0.039)	0.025 (0.047)	-0.060 (0.043)
eligibility for enrollment#female students	-0.018 (0.023)	-0.067 (0.035)	-0.162** (0.043)	0.069 (0.047)
<i>Graduation in humanities</i>				
eligibility for enrollment	0.058 (0.050)	0.002 (0.049)	-0.031 (0.046)	0.040 (0.036)
eligibility for enrollment#female students	0.023 (0.046)	0.080 (0.045)	0.153** (0.043)	-0.043 (0.053)
<i>Graduation (any field)</i>				
eligibility for enrollment	0.046 (0.053)	0.042 (0.041)	-0.006 (0.037)	-0.020 (0.037)
eligibility for enrollment#female students	0.005 (0.046)	0.013 (0.037)	-0.009 (0.038)	0.026 (0.044)
<i>Selective undergraduate program in science</i>				
eligibility for enrollment	0.008 (0.008)	0.004 (0.017)	0.018 (0.024)	0.005 (0.034)
eligibility for enrollment#female students	-0.007 (0.010)	-0.025 (0.015)	-0.026 (0.022)	-0.016 (0.026)
<i>Selective undergraduate program in humanities</i>				
eligibility for enrollment	0.017 (0.014)	-0.014 (0.021)	-0.003 (0.020)	0.014 (0.021)
eligibility for enrollment#female students	-0.028 (0.016)	0.036 (0.021)	0.013 (0.025)	0.005 (0.025)
<i>Selective undergraduate program (any field)</i>				
eligibility for enrollment	0.025 (0.016)	-0.007 (0.025)	0.015 (0.031)	0.019 (0.039)
eligibility for enrollment#female students	-0.035* (0.018)	0.007 (0.025)	-0.013 (0.034)	-0.013 (0.032)
N	1745	2111	1880	1837

Note: Same working sample as in Table 1.1 Panel A. We divided the full working sample into four subsamples according students' standardized ratio between their grades in Mathematics and French. Column 1 corresponds to students whose standardized ratio lies in the bottom quartile of the distribution of relative ability, column 2 to students whose standardized ratio lies in the second quartile, column 3 to students whose standardized ratio lies in the third quartile, and column 4 to students whose standardized ratio lies in the top quartile. Each group of two rows corresponds to a specific dependent variable, and for each dependent variable the table shows the impact of falling just above the admission threshold as well as the impact of falling just above the admission threshold interacted with a dummy variable indicating female students. Controls are the same as in Table 1.1. Estimates use a triangular kernel, and the bandwidth is computed following Calonico, Cattaneo and Titiunik (2014). Standard errors clustered at the individual level are in parenthesis. Each group of cells corresponds to a specific regression.

* significant at 5%. ** significant at 1%.

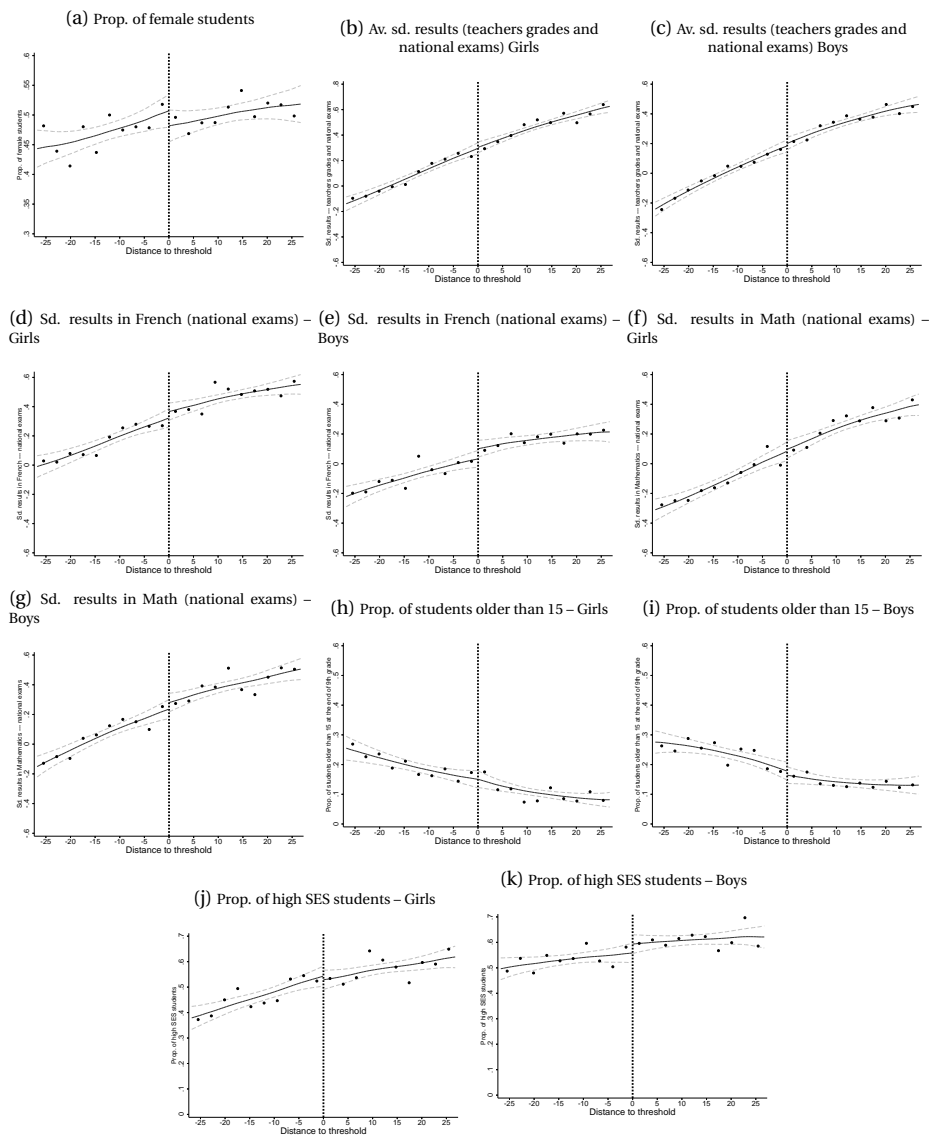


Figure 1.A1 – Students' pre-assignment characteristics

Note: Same sample as for Figure 1.1. The figure shows students' pre-assignment characteristics, plotted against the standardized distance to the threshold.

1.B Cut-off Scores

For each district and each cohort, we consider the sample of 9th grade students whose applications go through the standard assignment process, that is, they are not low-income and are not given any priority. Within each such sample, we partition the distribution of 9th grade scores into 1,000 equal size groups²² and we denote q_k ($k = 1, \dots, 999$) the corresponding quantiles. We then run 999 regressions for each high school j of the district, that is one regression per quantile q_k . For each school j and each k , the dependent variable is a dummy indicating whether student i is assigned to school j , and the independent variable is the dummy Q_{ik} indicating whether the score of student i is above the potential threshold q_k . We also control for a full set of fixed effects indicating students' middle schools of origin. For each school j , these 999 regressions provide us with 999 estimated R^2 and 999 estimated coefficients. For each school j , we denote $k(j)$ the k that corresponds to the maximum R^2 . For about half of schools, $q_j^* = q_{k(j)}$ is such that the estimated coefficient for $Q_{ik(j)}$ is positive, consistent with an increase in enrollment rates at the threshold. In all these cases, the estimated coefficient is also significant at standard level. In the remainder of the article, we assume that these schools were oversubscribed at the end of the first round of the assignment process, and that q_j^* corresponds to the minimum admission score of home-district applicants. In section 4, we show that for these schools the optimal q_j^* actually coincides with a very significant and discontinuous increase in the enrollment rate of middle school students coming from local middle schools. For the other schools, q_j^* is such that the estimated coefficient for $Q_{ik(j)}$ is non positive, namely there is no perceptible increase in enrollment at the main cut-off. We assume that these schools were undersubscribed at the end of the first round of the assignment process, and we drop them from our regression discontinuity analysis. We also drop three additional schools where we observe large proportion of out-of-district applicants among students eventually admitted (that is, more than 10% of out-of-districts students). Large proportion of out-of-districts students are suggestive that the school was either undersubscribed, or that it benefited from some derogatory rules²³.

²²We checked that results are very similar when we partition the sample in 2,000 equal size groups.

²³We checked that the vast majority of schools for which the estimated coefficient for Q_{ik} is non positive are also schools that admit large fraction of out-of-district students (that is, at least 10%), consistent with the assumption that they were undersubscribed at the end of the first round, and had to admit a significant number of students from other districts during the last two rounds.

Chapitre 2

Les redoublants nuisent-ils aux autres étudiants?

Le cas des classes préparatoires scientifiques

CO-ÉCRIT AVEC ÉRIC MAURIN

Résumé

Le redoublement est une pratique aujourd'hui largement répandue dans les premiers cycles de l'enseignement supérieur. Cette politique induit une augmentation de la taille des classes (effet de congestion) et modifie le profil des étudiants auxquels les professeurs doivent s'adresser (effet de composition). L'impact des redoublements sur le contexte dans lequel se déroulent les enseignements est particulièrement important dans les filières les plus sélectives, là où il est difficile de réussir et où les redoublants sont souvent d'un niveau supérieur à celui de la moyenne des nouveaux arrivants. Dans le contexte des classes préparatoires scientifiques, notre analyse ne détecte aucun effet de congestion, mais d'importants effets de composition : plus les redoublants sont d'un niveau élevé, moins les nouveaux arrivants réussissent. Les redoublants les plus forts ne semblent pas générer d'effets d'entraînement vertueux dans les classes, mais semblent en revanche induire les professeurs à produire des cours d'un niveau trop ambitieux pour la moyenne des nouveaux arrivants.

JEL : I21, I23, D62.

1 Introduction

Le redoublement est une pratique aujourd'hui très largement répandue dans les universités et les établissements d'enseignement supérieur du monde entier. Chaque année, dans la plupart des disciplines, un nombre important d'étudiants sont des redoublants, c'est-à-dire des étudiants ayant déjà bénéficié de l'enseignement proposé, mais ayant échoué à l'examen de validation ou échoué au concours préparé. En France, parmi les étudiants de première année d'université, on compte ainsi en moyenne près de 20% de redoublants (voir par exemple MESRI, 2013). Dans certaines disciplines, comme la médecine, la proportion de redoublants dans les classes de première année dépasse 30%. Les effets de cette politique restent pourtant mal connus, la recherche s'étant surtout intéressée aux effets du redoublement dans le primaire ou le secondaire.

Le redoublement affecte d'abord très directement ceux qui redoublent, même s'il n'est pas facile de prévoir exactement dans quel sens. Le redoublement peut en effet contribuer à décourager les étudiants et les conduire à abandonner leurs études. Mais il peut favoriser une assimilation plus profonde des enseignements et permettre aux étudiants concernés de continuer avec succès, et plus loin, dans la voie choisie. De fait, dans les filières les plus sélectives, celles où il est difficile de réussir du premier coup, nombre de redoublants comptent chaque année parmi les meilleurs élèves. Pour rester sur l'exemple de la médecine, les chances de passage en deuxième année sont environ quatre fois plus élevées pour les étudiants redoublant leur première année (43% en 2012) que pour les nouveaux arrivants (14%, voir MESRI, 2015).

En augmentant le nombre d'étudiants par classes et en modifiant la composition des classes, les redoublements modifient également le contexte dans lequel se déroulent les enseignements. Les redoublements n'affectent ainsi pas seulement les redoublants, ils affectent également indirectement l'ensemble des étudiants, redoublants ou non. Cet effet indirect n'est pas plus facile à prédire que l'effet direct sur les redoublants. D'un côté, l'augmentation de la taille des classes induit un effet a priori plutôt négatif sur les performances de l'ensemble des étudiants, qu'ils soient redoublants ou non. Mais les changements dans la composition des classes produisent des effets plus incertains. Dans les filières les plus sélectives, nombre de redoublants ont, dès le début d'année, un niveau académique supérieur à celui de la moyenne des nouveaux arrivants. Ils peuvent induire les enseignants à proposer des cours plus ambitieux, avec des effets difficiles à prédire sur les nouveaux arrivants.

L'objectif principal de cet article est d'éclairer ces questions et plus particulièrement, la question des effets indirects des redoublements. Nous explorons ce problème dans le contexte particulier des classes préparatoires aux écoles d'ingénieurs en France. Recrutés après un baccalauréat scientifique, les étudiants de ces classes doivent suivre deux années de préparation avant de pouvoir postuler aux écoles qu'ils convoitent. Il y a en France plus de 200 écoles d'ingénieurs, de prestige et renommée très variables, et l'accès à ces écoles est régi chaque année par une demi-douzaine de grands concours nationaux. En cas d'échec à ces concours, ou en cas de résultats trop en-deçà de leurs ambitions, les étudiants peuvent redoubler leur deuxième année (année dite de *Mathématiques Spéciales*) et tenter leur chance une seconde fois en sorte d'essayer d'être mieux classés et admis à rejoindre une école plus prestigieuse. Chaque année, un cinquième environ des étudiants décident de redoubler et reviennent l'année suivante grossir de nouveau les rangs des étudiants de *Mathématiques Spéciales*. Dans des classes comptant en moyenne environ trente-cinq étudiants, le nombre de redoublants dans une classe est en pratique très variable, pouvant passer de 3 ou 4 certaines années à plus de 12 d'autres années, en fonction des aléas des concours de l'année précédente. La question centrale sur laquelle nous allons nous pencher est de savoir la mesure dans laquelle un contingent plus ou moins important de redoublants (appelés "5/2" dans le jargon des classes préparatoires scientifiques) affecte les performances des primo-participants aux concours (appelés "3/2").

Le contexte des classes préparatoires est particulièrement propice à ce type d'étude pour au moins deux raisons. En premier lieu, les concours nationaux auxquels participent les étudiants de classes préparatoires fournissent un instrument de mesure des performances académiques très standardisé, stable dans le temps et indépendant des établissements où sont situées les classes préparatoires. En second lieu, le nombre d'étudiants de *Mathématiques Supérieures* admis en *Mathématiques Spéciales* se détermine chaque année, dans chaque établissement, de façon indépendante du nombre d'étudiants décidant de redoubler et des résultats aux concours obtenus par ces élèves. En pratique, les conseils de classe susceptibles de refuser le passage en *Mathématiques Spéciales* à certains étudiants de première année ont lieu début juin, soit bien avant que ne soient connus (par l'établissement) les flux de redoublements, lesquels ne peuvent commencer à être précisément évalués qu'à partir de début Août, quand commencent à tomber les résultats définitifs d'admission.

De fait, les données administratives dont nous disposons confirment que le nombre d'étudiants décidant de redoubler une année donnée, dans un lycée donné (et les résultats obtenus aux concours par ces redoublants) n'a aucun effet significatif sur le nombre et les

caractéristiques socio-démographiques des nouveaux arrivant en *Mathématiques Spéciales* au début de l'année suivante, dans ce même lycée. Autrement dit, la population d'étudiants arrivant en *Mathématiques Spéciales* varie d'une année sur l'autre d'une façon indépendante du nombre de redoublants observés en fin d'année précédente, largement indépendante également des résultats obtenus aux concours l'année précédente par ces redoublants, ces résultats pouvant être compris comme une mesure de leur niveau académique. Pour les nouveaux arrivants en *Mathématiques Spéciales*, la présence plus ou moins importante de redoublants s'apparente à un choc impossible à anticiper, mais contribuant très directement à modifier la taille et la composition des classes dans lesquelles ils peuvent préparer leurs concours.

Dans ce contexte, la question devient de savoir si (et comment) les changements de taille et de composition des classes induits par les redoublants (les 5/2) affectent les performances des nouveaux arrivants (les 3/2). L'analyse des résultats des 3/2 à l'un des concours nationaux les plus prestigieux et les plus importants (le concours Mines-Ponts) suggère que leurs performances ne sont en fait guère sensibles au nombre total de 5/2 et donc à la taille des classes, mais beaucoup plus sensibles au niveau académique des 5/2 et à la composition des classes. Plus précisément, les données dont nous disposons révèlent que les performances des 3/2 sont largement insensibles aux variations d'effectifs occasionnées par la présence plus ou moins grande de 5/2 ayant échoué au concours Mines-Ponts l'année précédente (aucune école obtenue), mais sont significativement moins bonnes lorsque leurs classes abritent un nombre important de 5/2 ayant réussi à être admis dans au moins une école du concours Mines-Ponts l'année précédente (et dont le redoublement est surtout motivé par l'espoir d'obtenir une admission dans une école plus prestigieuse encore). Cet effet négatif est particulièrement sensible dans les classes dites "étoilées", celles où les établissements concentrent traditionnellement leurs meilleurs 3/2. De fait, dans leur très grande majorité, ce sont dans ces classes étoilées que les 5/2 classés au concours Mines-Ponts effectuent leur deuxième année de *Mathématiques Spéciales*.

En définitive, nos résultats suggèrent que les 3/2 sont largement insensibles non seulement aux augmentations de taille des classes induites par les 5/2, mais également aux effets d'émulation et d'entraînement vertueux qu'on associe habituellement à la présence de très bons élèves dans les classes. Nos résultats suggèrent plutôt que c'est en induisant les professeurs à produire des cours trop ambitieux que les 5/2, et notamment les plus forts d'entre eux, affectent négativement les performances de leurs camarades 3/2. Les principaux bénéficiaires des redoublements seraient finalement les redoublants eux-mêmes. Nos données confirment que, dans leur très grande majorité, ils réussissent les

concours bien mieux lors de leur seconde participation que lors de la première. Les effets du redoublement se partageraient ainsi entre des effets directs positifs sur les redoublants eux-mêmes et des effets externes négatifs sur les autres élèves.

De manière générale, notre étude contribue à la littérature sur les effets du redoublement¹, et particulièrement sur les effets externes générées par les politiques de redoublement. S'agissant des effets externes, les quelques études existantes portent sur le lycée ou le collège² et, à notre connaissance, notre recherche est la première à analyser les externalités induites par les redoublements dans le contexte de l'enseignement supérieur, là où pourtant les redoublements restent massifs. L'une des spécificités du contexte institutionnel étudié est de permettre de séparer les effets de congestion liés aux redoublements (ils augmentent la taille des classes) des effets de composition des classes et de distorsion de l'enseignement (les redoublements modifient les interactions potentielles et induisent les professeurs à enseigner autrement).

Après avoir décrit le contexte institutionnel de notre étude (Section 2), nous détaillerons les données utilisées (Section 3) avant d'étudier le profil des redoublants et les effets directs du redoublement sur les redoublants (Section 4). Enfin, nous étudierons les effets indirects sur les non-redoublants (Section 5). Nous terminerons en analysant les conclusions qu'il est possible de tirer de cette étude (Section 6).

2 Contexte

Il existe en France plus de 200 écoles d'ingénieurs. Dans leur très grande majorité, les élèves de ces écoles sont recrutés sur concours à l'issue de deux ou trois années passées dans une classe préparatoire scientifique. Chaque année, environ 15,000 places sont ainsi offertes aux concours des écoles d'ingénieurs pour un total d'environ 25,000 candidats issus de classes préparatoires scientifiques³. Ces classes s'intègrent immédiatement après le baccalauréat. Elles recrutent traditionnellement parmi les meilleurs bacheliers des filières scien-

¹Les effets du redoublement sur les redoublants dans le primaire ou au collège ont donné lieu à de nombreuses contributions et de nombreux débats (voir par exemple Jacob et Lefgren, 2004, 2009; Manacorda, 2012; Schwerdt, West et Winters, 2017; Gary-Bobo, Goussé et Robin, 2016 ou Fruehwirth, Navarro et Takahashi, 2016. À notre connaissance, Tafreschi et Thiemann (2016) est le seul article proposant une évaluation des effets directs du redoublement dans l'enseignement supérieur.

²Dans une étude portant sur les collèges et lycées Israéliens, Lavy, Paserman et Schlosser (2012) mettent en évidence que les performances des élèves n'ayant pas de retard scolaire sont plutôt moins bonnes dans les cohortes comprenant une proportion plus importante d'élèves ayant un an de retard. Dans une étude portant sur un échantillon de lycéens américains, Hill (2014) identifie aussi que la probabilité de valider un cours de mathématiques du premier coup diminue quand augmente dans la classe la proportion d'élèves ayant déjà échoué à valider ce cours.

³Les données de cadrage sur les concours sont disponibles sur le site <https://www.scei-concours.fr/>.

tifiques⁴. La scolarité comprend une première année dite de *Mathématiques Supérieures* qui ne se redouble pas, suivie d'une année dite de *Mathématiques Spéciales* à l'issue de laquelle les étudiants passent les concours une première fois. En cas d'échec, les étudiants ont la possibilité de redoubler l'année de *Mathématiques Spéciales* et de passer les concours une seconde fois. Dans le jargon des classes préparatoires scientifiques, les redoublants sont désignés par le terme de "5/2" tandis que les nouveaux arrivant en classe de *Mathématiques Spéciales* sont désignés par le terme de "3/2". Chaque année, environ 25% des 3/2 décident de redoubler et de faire une année de 5/2.

2.1 Les concours

On compte une demi-douzaine de concours, chacun organisé par un groupe d'écoles particulier. Chacun de ces concours comporte typiquement trois ou quatre jours d'épreuves écrites, à l'issue desquelles une liste de candidats admissibles est établie. Ces candidats admissibles passent ensuite une semaine d'épreuves orales, à l'issue de laquelle une liste de candidats classés est établie (tous les admissibles ne sont pas nécessairement classés), chaque étudiant de cette liste recevant un rang selon son nombre total de points. Les candidats classés sont alors invités à ranger les écoles des concours dans lesquels ils sont classés par ordre décroissant de préférence. Un algorithme centralisé assigne ensuite une école à chaque candidat en fonction de son rang dans les différents concours et de ses préférences. À l'issue de cette procédure, on peut évaluer pour chaque école le rang du dernier admis, c'est-à-dire le rang minimal nécessaire pour une admission.

Les trois concours les plus difficiles et les plus prestigieux sont le concours X/ENS (sur lequel recrutent l'École Polytechnique et les Écoles Normales Supérieures); le concours Centrale-Supelec (sur lequel recrutent l'École CentraleSupelec et dix autres écoles) et le concours Mines-Ponts (sur lequel recrutent l'École Nationale Supérieure des Mines de Paris et l'École Nationale des Ponts et Chaussées ainsi que sept autres écoles). À ces trois concours s'ajoutent le concours Mines-Télécom (une douzaine d'écoles), le Concours Commun Polytechniques (dit CCP, une soixantaine d'écoles) et le concours E3A (une soixantaine d'écoles également). Certaines écoles sont relativement faciles d'accès (notamment au sein des concours CCP et E3A), mais d'autres sont en revanche très sélectives et prestigieuses, figurant même en bonne place dans les palmarès mondiaux d'écoles d'ingénieurs, notamment ceux basés sur la réputation des écoles auprès des employeurs.

Les épreuves écrites commencent en général vers la fin Avril et les épreuves orales

⁴Voir MESRI (2012)

s'achèvent fin Juillet. Le processus centralisé d'assignation des candidats classés aux écoles des différents concours se déroule en plusieurs tours durant le mois d'Août. Les organisateurs des épreuves des différents concours se coordonnent de façon à ce qu'un élève puisse passer l'ensemble des principaux concours. Il s'agit de concours nationaux, chacun disposant de nombreux centres d'examens répartis sur l'ensemble du territoire.

Dans la suite, notre analyse portera essentiellement sur les résultats des élèves au concours Mines-Ponts. Ce concours attire chaque année environ 15,000 candidats de classes préparatoires (pour environ 1,200 places offertes), ce qui en fait l'un des plus importants. Dans les lycées de notre échantillon de travail, environ 80% des étudiants s'inscrivent au concours Mines-Ponts. Ce concours détermine le recrutement de dix écoles d'ingénieurs parmi les plus prestigieuses⁵. Il s'agit de l'un des concours les plus difficiles et les plus sélectifs : un peu plus du quart seulement des candidats sont classés et admis à rejoindre une des dix écoles organisatrices.

2.2 Les classes préparatoires

Dans leur très grande majorité les candidats aux concours des écoles ingénieurs sont issus d'une classe préparatoire scientifique. Bien que relevant de l'enseignement supérieur, ces classes sont situées dans les lycées de l'enseignement secondaire. Les étudiants sont regroupés en classes de 30 à 50 élèves qui suivent les principaux cours ensemble, tout au long de l'année. Les principales matières étudiées sont les mathématiques, la physique, l'informatique et les sciences de l'ingénieur. Il existe toutefois plusieurs types de classes préparatoires en fonction de l'importance relative donnée à ces différentes matières. (pour une vue d'ensemble du système des classes préparatoire scientifiques, voir la Figure 2.A1 en Annexe). Les professeurs sont des fonctionnaires de l'éducation nationale comme les professeurs de lycée. Ils sont choisis dans chaque discipline parmi les professeurs agrégés, généralement parmi les mieux classés au concours de l'agrégation. Enseigner en classe préparatoire leur permet d'accéder au corps des professeurs de chaire supérieure, l'un des plus prestigieux de l'éducation nationale.

⁵Il s'agit de l'École Nationale Supérieure des Mines de Paris, de l'École Nationale des Ponts et Chaussées (ENPC), de l'Institut Supérieur de l'Aéronautique (ISAE-Supaero), de l'École Nationale Supérieure des Techniques Avancées, de l'École Nationale Supérieure des Télécommunications, de l'École Nationale de la Statistique et de l'Administration Économique (ENSAE), de l'École Nationale Supérieure des Mines de Nancy, de l'École Nationale Supérieure des Mines de Saint-Etienne et de l'École Nationale Supérieure des Télécommunication de Bretagne. Ces écoles se rangent traditionnellement par ordre décroissant de sélectivité et de prestige selon : Mines de Paris > Ponts et Chaussées > Supaero ≈ Techniques Avancées ≈ Télécom Paris > ENSAE > Mines de Nancy > Mines de Saint-Etienne > Télécom Bretagne. D'une année sur l'autre, les rangs des derniers admis dans ces écoles respectent quasi systématiquement l'ordre ci-dessus.

En première année (classes dites de *Mathématiques Supérieures*), les principales filières offertes aux nouveaux bacheliers sont une filière à dominante mathématiques (classes de type *MPSI*, environ 8,000 étudiants chaque année), une filière à programme renforcé en physique chimie (classes de type *PCSI*, environ 8,000 élèves) ainsi qu'une filière à programme renforcé en sciences de l'ingénieur (*PTSI*, environ 3,000 élèves)⁶. Comme mentionné plus haut, cette première année ne se redouble pas. Environ 20% des élèves abandonnent toutefois avant la fin de l'année de *Mathématiques Supérieures*, la plupart en début d'année scolaire, ce qui leur permet de rejoindre d'autres cursus⁷.

En deuxième année (classes dites de *Mathématiques Spéciales*), la majorité des étudiants issus de *MPSI* poursuivent dans une filière où la dominante mathématiques est renforcée en rejoignant des classes de type *MP*/MP*. Certains ont toutefois la possibilité de revenir à un programme plus équilibré, avec davantage de sciences de l'ingénieur en rejoignant des classes de type *PSI*/PSI*. De même les étudiants issus de *PCSI* se partagent entre ceux qui poursuivent dans des classes renforcées en physique-chimie (classes *PC*/PC*) et ceux qui rejoignent des classes de type *PSI*/PSI*. Dans leur immense majorité, les étudiants des classes *PTSI* rejoignent des classes de types *PT*/PT* à programme renforcé en sciences de l'ingénieur.

À de rares exceptions près, les différentes écoles d'ingénieurs offrent des places et recrutent des candidats dans chacune des grandes filières *MP*/MP*, *PC*/PC*, *PSI*/PSI* ou *PT*/PT*. Chacun des grands concours se déclinent donc en autant de série d'épreuves qu'il y a de grande filière de *Mathématiques Spéciales*. Par exemple, pour le concours Mines-Ponts, il y a une série d'épreuves pour les élèves issus des classes *MP*/MP*, une série pour les élèves issus des classes *PSI*/PSI*, etc. Il existe ensuite autant de classements (puis de procédures d'affectation) qu'il y a de filières.

2.3 Classes étoilées

Dans chacune des filières coexistent deux types de classes de *Mathématiques Spéciales*, les classes dites "étoilées" (i.e., classes *MP**, *PSI**, *PC** ou *PT**) et les classes dites "non étoilées" (i.e., classes *MP*, *PSI*, *PC* ou *PT*). Les classes étoilées regroupent dans chaque lycée et filière les élèves dont le potentiel est jugé le plus élevé en fin de *Mathématiques Supérieures*. Ces élèves représentent un tiers environ du total des élèves de *Mathématiques*

⁶Il existe également une filière à programme renforcé en science de la vie et de la terre préparant spécifiquement aux concours d'entrées des écoles d'ingénieurs agronomes (classes *BPCST*, environ 3,000 élèves).

⁷Cette proportion peut être estimée à l'aide du panel des bacheliers 2008 (voir *MENESR*, 2013).

Supérieures.

Aucun concours n'est réservé aux élèves de classes étoilées et, chaque année, la quasi-totalité des écoles admet des élèves issus des deux types de classes. Par exemple, en 2016, sur un total d'environ 5,400 candidats au concours Mines-Ponts filière MP/MP*, on comptait 65% d'élèves issus de classes MP et 35% issus de classes MP*. Sélectionnés parmi les meilleurs élèves de *Mathématiques Supérieures* et préparés aux exercices les plus difficiles durant leur année de *Mathématiques Spéciales*, les élèves issus des classes étoilées ont toutefois des résultats en moyenne nettement meilleurs que ceux des élèves issus des classes non étoilées. Toujours lors du concours Mines-Ponts de 2016, environ 63% des candidats issus de classes MP* ont réussi à être classés (ce qui situe leur résultats parmi les 30% les meilleurs à ce concours) contre seulement 14% des élèves de classes MP⁸. De même, la proportion d'élèves étoilés parmi les admis aux différentes écoles s'accroît au fur et à mesure qu'on considère des écoles plus prestigieuses. La Figure 2.A2 en Annexe retrace la proportion d'élèves issus de classes étoilées parmi les admis aux différentes écoles du concours Mines-Ponts : elle est très majoritaire dans les dix écoles, mais elle n'en demeure pas moins près de 20 points plus élevés parmi les admis à l'école de Mines de Paris (la plus prestigieuse des dix écoles) que parmi les admis à Télécom Bretagne (la moins prestigieuse des dix).

3 Données utilisées

Les données administratives utilisées proviennent du service statistique du ministère de l'éducation nationale ainsi que du service statistique du concours Mines-Ponts.

3.1 Les données du concours Mines-Ponts

Les données du concours Mines-Ponts dont nous disposons recensent l'ensemble des élèves inscrits à ce concours dans la filière MP/MP* ainsi que dans la filière PSI/PSI*, pour l'ensemble des sessions entre 2012 et 2016. Pour chacun des élèves inscrits à ces sessions, le service statistique du concours Mines-Ponts fournit l'établissement d'origine. Il indique également s'il provient d'une classe étoilée et s'il passe le concours en tant que 3/2 ou 5/2. S'agissant des performances au concours lui-même, ces données précisent pour chaque élève s'il a été admissible à l'issue des écrits et le cas échéant son classement final, permettant ainsi de reconstituer le nombre d'écoles obtenues pour chaque élève⁹. À partir de ces informations, nous avons constitué une base de données décrivant chaque année, pour chaque type

⁸Ces informations sont disponibles sur le site du concours Mines-Ponts, <https://mines-ponts.fr>.

⁹Les rangs des derniers appelés de chaque école du concours Mines-Ponts sont disponibles sur <https://www.scei-concours.fr/>.

de classe (étoilée, non étoilée) de chaque filière (MP/MP* ou PSI/PSI*) de chaque établissement, les performances des 3/2 au concours, telles que mesurées par le nombre d'admissibles, le nombre d'admis à au moins une école ou le nombre d'admis à au moins l'une des cinq écoles les plus prestigieuses. Ces variables de performances constitueront nos principales variables dépendantes.

3.2 Les données du ministère

Les données du ministère de l'éducation nationale recensent pour chaque année scolaire entre 2011-2012 et 2015-2016 l'ensemble des élèves scolarisés en première ou seconde année de classe préparatoire scientifique dans l'ensemble des établissements français, publics et privés. Ces données sont anonymes, mais contiennent un identifiant élève permettant de suivre les élèves d'une année scolaire à l'autre. Il est ainsi possible de distinguer les élèves scolarisés pour la première fois en *Mathématiques Spéciales (3/2)* et les élèves redoublant leur année de *Mathématiques Spéciales (5/2)*. Ces données contiennent également pour chaque élève des informations socio-démographiques (date de naissance, sexe, statut de boursier), des informations sur les établissements et les classes fréquentées (filière, classe étoilée ou non) et des informations sur les résultats au baccalauréat.

À partir de ces informations nous avons construit une base de données décrivant pour chaque année (entre 2011-2012 et 2015-2016), et chaque type de classe (étoilée, non étoilée) de chaque filière (MP/MP* ou PSI/PSI*) de chaque établissement : les effectifs de 5/2 et de 3/2, les effectifs de 3/2 de sexe féminin, les effectifs de 3/2 boursiers, et les notes moyennes des 3/2 au baccalauréat. Nous disposons également d'une variable décrivant (pour chaque année entre 2011-2012 et 2014-2015, chaque lycée et chaque filière) les effectifs de 3/2 décidant de redoubler l'année suivante, ce qui permettra de constituer notre principale variable indépendante.

3.3 Identification des étudiants qui redoublent après avoir été classés

Nous avons apparié la base de données issue du service du concours et celle issue des données du ministère en utilisant les identifiants établissement, filière, type de classe et année disponibles dans chacun d'eux. Les différentes variables d'effectifs présentes dans ces bases ont été rapportées au nombre de classes de l'établissement dans la filière et le type de classe correspondants.

Ce panel apparié permet d'étudier les liens existants (au niveau établissements \times filières) entre le nombre de 3/2 par classe décidant de redoubler à la fin d'une année donnée d'une

part, et, d'autre part, le nombre de 5/2 par classe, la taille des classes et les performances des 3/2 l'année suivante. Il ne permet toutefois pas de tester si l'impact du nombre de redoublants par classe dépend du niveau académique de ces redoublants, tel que mesuré notamment par leurs performances lors de leur première participation aux concours. Pour pouvoir éclairer cette question, nous avons complété notre panel d'établissement d'une information sur les effectifs d'étudiants redoublant après avoir été classés au concours Mines-Ponts. Pour construire cette variable, nous avons apparié les données issues du concours Mines-Ponts et les données issues du ministère au niveau individuel, en utilisant le détail des informations individuelles (scolaires et socio-démographiques) présentes dans les deux sources. Cet appariement au niveau individuel permet de reconstituer pour 94% des étudiants inscrits au concours une variable décrivant s'ils ont redoublé après avoir été classés au concours. Après agrégation au niveau établissement, filière et année, ces données permettent de reconstituer une variable décrivant pour chaque établissements \times filières et chaque année, les effectifs d'élèves décidant de redoubler alors qu'ils ont été classés au concours Mines-Ponts et une variable décrivant les effectifs décidant de redoubler alors qu'ils ont échoué à être classés à ce même concours. Notons également que le panel de données appariées au niveau individuel nous permettra de donner une idée très précise de l'impact direct du redoublement sur les redoublants, en comparant les performances des étudiants redoublant lors de leur première et de leur seconde participation aux concours.

3.4 Échantillon de travail

Ainsi construit et complété, le panel issu de l'appariement des sources ministère et concours comprend un total de 258 établissements \times filières observés sur la période comprise entre 2011-2012 et 2015-2016. Dans la suite, pour être en mesure d'identifier séparément les effets de congestion et de composition induits par les redoublements¹⁰, nous nous concentrerons sur le panel de 65 établissements \times filières pour lesquels on observe à la fois des classes étoilées et non étoilées. Nous écarterons également de notre échantillon de travail les deux établissements \times filières qui ne sont pas présents sur l'ensemble de la période ainsi que les quatre établissements dont le nombre de classes a changé au cours de la période. En définitive, notre échantillon de travail comportera les 59 établissements \times filières observés cinq années consécutives, sans modification de nombre de classes et comprenant à la fois des classes étoilées et non étoilées. Ces 59 établissements \times filières représentent 23% de l'échantillon initial d'établissements \times filières, mais scolarisent environ 45% des élèves de l'échantillon initial.

¹⁰Comme nous le verrons dans la section suivante, cette identification séparée réclame en effet de pouvoir comparer les effets induits par les redoublants dans les classes étoilées et dans les classes non étoilées.

Les Tables 2.A1 à 2.A2 en Annexe présentent des statistiques décrivant les principales variables dans l'échantillon initial ainsi que dans l'échantillon de travail.

4 Les redoublants : profils et performances

La première année de classe préparatoire (*Mathématiques Supérieures*) ne se redouble pas, sauf circonstance exceptionnelle, comme une maladie de longue durée. En revanche, la seconde année (*Mathématiques Spéciales*) peut se redoubler. C'est l'année à la fin de laquelle les élèves passent les concours et les élèves qui les passent pour la première fois (les "3/2") peuvent décider de redoubler, en particulier s'ils n'obtiennent aucune admission ou s'ils estiment pouvoir obtenir une école plus prestigieuse après une année de préparation supplémentaire.

Chaque année, à l'issue de leur première participation aux concours, 25% environ des élèves 3/2 de notre échantillon de travail décident de redoubler l'année de *Mathématiques Spéciales* (soit environ 8 redoublements par classe comptant en moyenne une trentaine de 3/2). Sans surprise, la probabilité de redoubler est plus forte pour les élèves n'ayant obtenu aucune admission au concours Mines-Ponts (30%), même si elle est loin d'être négligeable pour les élèves ayant pourtant réussi à être admis à au moins une école (14%). Parmi les élèves ayant obtenu une admission dans une école du concours Mines-Ponts, la probabilité de redoubler est d'autant plus forte que cette école est moins prestigieuse. Ainsi, environ 29% des 3/2 admis à Télécom Bretagne redoublent, contre moins de 5% des 3/2 admis aux Mines de Paris.

Dans près de neuf cas sur dix, le redoublement s'effectue dans le même établissement. Il arrive toutefois qu'il s'effectue dans un autre lycée. Quand un élève issu d'une classe préparatoire relativement peu prestigieuse obtient de bons résultats aux concours, il peut en effet obtenir un redoublement dans une classe préparatoire d'un autre lycée, plus prestigieux, avec l'espoir de se préparer plus spécifiquement aux concours les plus difficiles.

4.1 Les effets directs du redoublement sur les redoublants

Les élèves qui redoublent après avoir été classés au concours Mines-Ponts représentent, l'année suivante, un noyau d'élèves d'autant plus forts académiquement que le redoublement leur est bien souvent très bénéfique. Non seulement se recrutent-ils parmi les bons élèves de l'année précédente, mais le redoublement les renforce.

Ainsi, quand on se concentre sur les élèves redoublant après avoir été classés à l'issue de

leur première participation, on peut mesurer qu'ils sont admis en moyenne à 2.8 écoles de plus en fin de 5/2, ce qui représente un gain de plus de 300 places en moyenne. De fait, quand on compare leur classement en fin de 5/2 à leur classement en fin de 3/2, on peut constater que 80% obtiennent un meilleur classement¹¹.

Les progrès sont particulièrement significatifs pour les redoublants admis à une des 5 écoles du groupe B lors de leur première participation¹² (lesquels représentent 75% des redoublants classés). Dans leur immense majorité (97%) ces redoublants repassent le concours Mines-Ponts l'année suivante et 61% de ceux qui le repassent obtiennent une admission dans une école du groupe A, c'est-à-dire l'une des 5 écoles les plus sélectives. Environ 34% ré-obtiennent une école du groupe B et seulement environ 5% n'arrivent pas à obtenir d'école. En d'autres termes, le redoublement permet à une large majorité de redoublants classés d'obtenir un gain très significatif en terme de prestige et ne conduit à un échec (au concours Mines-Ponts) que pour une toute petite minorité d'entre eux. La nette amélioration des performances des redoublants lors de leur seconde participation est d'autant plus remarquable qu'ils sont lestés (lors de leur dernière participation) d'un handicap initial de points à peu près équivalent à 200 places.

4.2 Le niveau académique des redoublants classés et non classés

De façon générale, nos données permettent de distinguer deux grands types de redoublants, à savoir ceux qui redoublent alors qu'ils ont été classés au concours Mines-Ponts lors de leur première participation (redoublants "classés") et ceux qui redoublent alors qu'ils n'ont pas été classés à ce concours (redoublants "non classés"). Dans notre échantillon de travail, on compte environ un redoublant classé par classe et sept redoublants non classés.

Lors de leur première participation, les premiers sont admis en moyenne à 3,2 écoles du concours Mines-Ponts, un résultat qui les situe au-dessus du résultat moyen des élèves de leur lycée passant le concours pour la première fois dans une classe non étoilée (0,3 école) et à peu près au même niveau que la moyenne de ceux passant le concours pour la première fois dans une classe étoilée (2,8 écoles). Au début de l'année suivante, ces redoublants représenteront un petit noyau d'élèves dont le niveau sera d'emblée au-dessus

¹¹On vérifie que 94% des 3/2 qui redoublent après avoir été classés repassent le concours l'année suivante. Pour les 6% qui ne le repassent pas, nous avons appliqué la convention de les considérer comme non classés (c'est-à-dire comme étant classé à un rang au-delà du rang du dernier classé), en sorte que nos estimations représentent une sous-estimation de l'effet des redoublements sur les rangs potentiels des redoublants.

¹²Dans le groupe B, on convient ici de placer les 5 écoles les moins sélectives, à savoir l'ENSAE, les Mines de Nancy, les Mines de Saint-Etienne, Télécom Bretagne ainsi que le campus niçois de Télécom Paris.

de la moyenne des nouveaux élèves de *Mathématiques Spéciales* de leur lycée, en classe non étoilée, mais également en classe étoilée (leur niveau de départ étant d'emblée supérieur à celui que les 3/2 des classes étoilées n'atteignent en moyenne qu'en fin d'année). À l'issue de leur seconde année de *Mathématiques Spéciales*, ces élèves seront finalement admis en moyenne à près de 6 écoles du concours Mines-Ponts, des résultats largement au-dessus de ceux obtenus en moyenne par les 3/2.

Le deuxième type de redoublants (les non classés) n'est par construction admis à aucune école lors de sa première participation. Leur impact sur le niveau général des classes dans lesquelles ils redoublent l'année suivante est plus incertain. En fin de seconde année de *Mathématiques Spéciales*, on peut constater qu'ils sont admis à un peu moins d'une école Mines-Ponts en moyenne, soit 1,7 écoles quand ils redoublent en classe étoilée, et 0,4 école quand ils redoublent en classe non étoilée. Leur niveau de fin d'année apparaît très proche de celui de leur classe et on peut spéculer qu'ils n'induisent pas de modification fondamentale dans le niveau des classes, juste une modification de la taille des classes.

5 L'influence des redoublants sur les non-redoublants

Chaque année, parmi les élèves achevant leur première année de *Mathématiques Spéciales*, près de 25% décident de redoubler et de revenir l'année suivante grossir les rangs des préparatoires. Dans des classes comptant en moyenne une quarantaine d'élèves, le nombre d'élèves décidant de redoubler peut toutefois beaucoup varier d'une année sur l'autre : il y en a parfois deux ou trois, d'autres fois plus de quinze (voir Figure 2.A3 en Annexe). Certaines années, ce contingent est essentiellement composé d'étudiants qui n'ont rien eu aux concours l'année précédente, tandis que d'autres années, ce contingent compte des étudiants qui auraient pu intégrer une très bonne école, mais espèrent améliorer encore leur performance. La question posée est de savoir si la taille et la composition de ce contingent de redoublants a une influence sur les performances de la cohorte d'élèves arrivant l'année suivante dans ces mêmes classes de *Mathématiques Spéciales*. D'une année sur l'autre, si le nombre d'étudiants décidant de redoubler vient à s'accroître, comment cela se répercute-t-il sur les conditions d'études et les performances des 3/2, l'année suivante dans le même lycée? Les effets sont-ils les mêmes dans les classes étoilées et non étoilées? Sont-ils identiques lorsque l'afflux de redoublants correspond à des étudiants à fort potentiel, qui étaient déjà au niveau des meilleures écoles l'année précédente, ou au contraire à des étudiants qui n'ont rien obtenu de très prestigieux à leur première participation aux concours?

Pour éclairer ces questions, nous allons analyser la mesure dans laquelle les performances aux concours Mines-Ponts des étudiants de 3/2 une année t donnée sont affectées par le nombre d'étudiants ayant décidé de redoubler dans le même lycée et la même filière à la fin de l'année précédente $t - 1$. Plus précisément, en notant Y_{jft} les performances moyennes au concours Mines-Ponts des élèves 3/2 l'année t , dans la filière f et le lycée j , notre objectif est d'estimer un modèle de régression en forme réduite du type :

$$Y_{jft} = \alpha R_{jft-1} + \beta X_{jft} + u_{jft}$$

où R_{jft-1} représente le nombre de redoublants par classe en fin de $t - 1$ dans le lycée j et la filière f tandis que X_{jft} représente un ensemble de variables de contrôle incluant des effets fixes lycées, filières, dates ainsi que des variables décrivant le profil socio-démographique des élèves de 3/2 en t (nombre de filles et de boursiers par classes, résultats moyens au baccalauréat). La variable u_{jft} représente quant à elle les facteurs explicatifs non observés de la performance des 3/2 l'année t .

Dans cette analyse, le paramètre d'intérêt est le paramètre α . L'hypothèse d'identification est que la variable R_{jft-1} est non corrélée avec la variable u_{jft} . Cela revient à postuler que les facteurs explicatifs de la réussite des 3/2 l'année t n'ont rien de commun avec les facteurs expliquant les décisions de redoubler des élèves de la génération précédente, à la fin de l'année $t - 1$. Pour tester la crédibilité de cette hypothèse, la première colonne de panel A du Tableau 2.1 retrace les résultats obtenus en régressant les variables décrivant le nombre et les caractéristiques des 3/2 de l'année t sur le nombre d'élèves décidant de redoubler à la fin de $t - 1$, tel que mesuré par la variable R_{jft-1} . Il s'agit par exemple de savoir si les conditions d'admission des 3/2 se durcissent quand le nombre de 5/2 potentiels est plus élevé. De façon réconfortante, cette analyse ne révèle aucun effet significatif du nombre d'élèves décidant de redoubler à la fin de l'année $t - 1$ dans un lycée et une filière donné sur le nombre et les caractéristiques des 3/2 observés l'année suivante en *Mathématiques Spéciales* dans le même lycée et la même filière. Les deuxième et troisième colonnes répètent le même exercice séparément sur les étudiants de classes étoilées et non étoilées. De nouveau, le nombre de redoublants en $t - 1$ dans un lycée et une filière donnés n'apparaît avoir aucun effet sur le nombre et les caractéristiques des 3/2 observés en t , ni dans les classes étoilées ni dans les classes non étoilées. Ces différents résultats sont cohérents avec notre hypothèse d'identification, c'est-à-dire cohérent avec l'idée que les flux d'entrées en 3/2 pour l'année t se décident indépendamment des décisions de redoublements en fin d'année $t - 1$, une des raisons fondamentales étant que les entrées en 3/2 se décident pour l'essentiel avant les redoublements.

5.1 Effets des redoublements sur la taille des classes l'année suivante

De cette indépendance entre décisions de redoublements en $t - 1$ et flux d'entrées en 3/2 en t , il découle que l'effet le plus direct des redoublements décidés en $t - 1$ est de gonfler la taille des classes dans laquelle les 3/2 de l'année t effectuent leur première année de *Mathématiques Spéciales* et, plus précisément, de la gonfler de 5/2 issus du lycée, c'est-à-dire d'élèves ayant déjà une année de préparation derrière eux dans le lycée considéré (effet de congestion). Pour le vérifier, nous avons régressé le nombre de 5/2 par classe observé l'année t (ainsi que le nombre total d'élèves par classe observé l'année t) sur le nombre R_{jft} d'élèves décidant de redoubler à la fin de l'année $t - 1$ (voir panel B du Tableau 2.1). Ces régressions confirment que chaque redoublant supplémentaire en fin d'année $t - 1$ se traduit l'année suivante par un accroissement de quasi une unité du nombre de 5/2 et du nombre total d'élèves par classe. Ce lien très direct entre le nombre de redoublements l'année $t - 1$ et le nombre de 5/2 l'année suivante est par ailleurs quasi aussi fort dans les classes étoilées et non étoilées.

Notons que si l'augmentation du nombre de redoublants ne se répercute pas parfaitement sur le nombre de 5/2 l'année suivante (i.e., on n'a pas exactement un 5/2 supplémentaire par redoublant supplémentaire), c'est simplement, comme nous avons déjà pu le mentionner plus haut, qu'une petite partie des élèves redoublent dans un autre lycée. Inversement, chaque lycée peut accueillir chaque année des élèves de 5/2 en provenance d'autres lycées. Comme le montre la deuxième ligne du panel B, on vérifie toutefois que le nombre de redoublants l'année $t - 1$ dans un lycée et une filière donnés n'a aucun effet sur le nombre de 5/2 en provenance d'un autre lycée accueilli l'année suivante dans ce même lycée et cette même filière. Les décisions d'admission de 5/2 en provenance de lycées extérieurs portent sur un très petit nombre d'élèves et apparaissent donc elles-aussi comme largement indépendantes des flux de redoublements R_{jft} qui se décident en $t - 1$.

5.2 Effets des 5/2 sur les performances des 3/2

À la rentrée qui suit une année où les redoublements sont relativement nombreux, les élèves arrivant en 3/2 se retrouvent dans des classes plus nombreuses, mais avec des camarades en moyenne plus aguerris et préparés. A priori, une taille des classes plus élevée contribue à les pénaliser dans leur préparation (effet de congestion). Le fait d'avoir un noyau plus important de camarades expérimentés et déjà avancés dans leur préparation a un effet plus incertain : cela peut contribuer à une émulation tirant la classe vers le haut,

mais cela peut également conduire le professeur à produire un cours trop ambitieux pour les 3/2.

Pour évaluer lequel de ces différents mécanismes domine, nous avons estimé l'effet du nombre de redoublements décidés à la fin d'une année donnée sur un ensemble de variables décrivant (dans le même lycée et la même filière) les performances des 3/2 au concours Mines-Ponts l'année suivante, à savoir (i) les effectifs par classe de 3/2 inscrits au concours Mines-Ponts, (ii) les effectifs par classe de 3/2 admissibles à ce même concours (rappelons que cela revient à être dans les 30% des meilleurs candidats aux épreuves écrites), (iii) les effectifs par classe de 3/2 admis à au moins une école à l'issue des épreuves orales (iv) le nombre de 3/2 par classe admis dans au-moins une des 5 écoles les plus prestigieuses l'année suivante (Mines, Ponts, Supaero, Ensta, Télécom Paris, ce qui revient à peu près à être dans les 15% les mieux classés du concours). De nouveau, les modèles de régression utilisés pour cette analyse incluent un ensemble complet d'effets fixes lycées, filières et dates, en sorte que les coefficients de régression présentés dans le tableau peuvent se comprendre comme l'effet des variations du nombre d'étudiants par classe décidant de redoubler sur les performances observées l'année suivante dans la nouvelle cohorte.

Les résultats sont reportés dans le panel C du Tableau 2.1 et ils sont dépourvus d'ambiguïté : un nombre important de redoublants en fin d'année $t - 1$ ne s'accompagne l'année suivante t chez les 3/2 d'aucune variation significative du nombre d'inscrits au concours Mines-Ponts ni d'aucune variation significative du nombre d'admissibles ou du nombre d'admis aux écoles de ce concours. Le même constat reste valable que l'on considère l'ensemble des classes ou que l'on considère séparément les classes étoilées et non étoilées. L'effet tend à être négatif dans les classes étoilées, mais non significatif aux seuils usuels.

5.3 Redoublants classés vs. non classés

Les résultats du Tableau 2.1 suggèrent que les variations de taille des classes générées par les 5/2 n'ont en elles-mêmes pas d'effets déprimants majeurs sur les performances des 3/2. Pour explorer si le niveau académique des 5/2 a une influence, nous avons répliqué l'analyse précédente en distinguant parmi les redoublants de l'année $t - 1$ ceux qui redoublent après avoir échoué au concours Mines-Ponts et ceux – moins nombreux – qui redoublent en dépit du fait qu'ils ont réussi à être classés à ce même concours.

Cette analyse est présentée dans le Tableau 2.2. Pour commencer, le panel A du Tableau permet de vérifier que les effectifs et caractéristiques des 3/2 ne sont corrélés ni avec le nombre de redoublants classés ni avec celui de redoublants non classés observés l'année

précédente. Ces résultats se situent dans la lignée de ceux du panel A du Tableau 2.1. Ils sont cohérents avec l'hypothèse de travail qui est la nôtre, à savoir que les flux d'entrée en 3/2 se déterminent de façon indépendante des flux de redoublants, qu'il s'agisse de redoublants classés ou non classés.

Le panel B analyse l'impact des deux types de redoublants, classés et non classés, sur le nombre de 5/2 et la taille des classes l'année suivante. Il confirme que les deux types de redoublants ont un effet très direct sur la taille des classes observées en moyenne l'année suivante. Qu'ils soient classés ou non, quand le nombre de redoublants augmente d'une unité, la taille des classes augmente en moyenne de quasi une unité l'année suivante. Mais le tableau révèle également que les redoublants classés contribuent essentiellement à augmenter la taille des classes étoilées, pas du tout celle des classes non étoilées¹³, tandis que les redoublants non classés ont un effet similaire sur la taille des classes étoilées et non étoilées. L'asymétrie de l'effet des redoublants classés sur la taille des classes étoilées et non étoilées reflète que les redoublants classés sont dans leur immense majorité invités à faire leur deuxième année de *Mathématiques Spéciales* dans les classes étoilées de leur lycée.

Pour finir, le panel C du Tableau 2.2 analyse l'impact des deux types de redoublants sur les performances des 3/2 au concours. Il révèle que les redoublants non classés n'ont aucun effet sur les performances des 3/2, ni en classe étoilées ni en classe non étoilées, en dépit de leur impact très direct sur les effectifs des deux types de classes. Mais le même panel C révèle également que les redoublants classés ont un effet significativement négatif sur les performances des 3/2, qu'il s'agisse de leur probabilité d'être admissible à l'issue des écrits ou de leur probabilité d'être admis dans une école à l'issue des oraux. En outre, de façon cohérente avec le panel B, ces effets négatifs sont essentiellement perceptibles en classe étoilée, c'est-à-dire dans les classes que les redoublants classés contribuent spécifiquement à grossir. Pour être précis, si le nombre de redoublants classés s'accroît d'un élève par classe en $t - 1$ dans une filière donnée (ce qui correspond à environ 70% d'un écart type de cette variable) alors nos estimations suggèrent que le nombre de 3/2 de cette filière admis à rejoindre au moins une école du concours baisse l'année suivante de -0,79 élèves par classe en moyenne¹⁴ (soit environ -11% d'un écart type de cette variable).

¹³Quand le nombre de redoublants classés augmente d'un élève par classe, on n'observe l'année suivante aucune hausse significative dans les classes non étoilées, mais une hausse de plus d'un élève par classe dans les classes étoilées, puisque les redoublants classés vont quasi exclusivement dans ces classes.

¹⁴En guise de test placebo, nous avons vérifié que le nombre de redoublants dans une filière n'avait pas d'effet sur les performances des 3/2 de l'autre filière l'année suivante. Nos résultats ne sont donc pas l'expression de facteurs inobservés propres au lycée (ou proviseur) qui affecteraient conjointement les probabilités de redoubler et les performances dans l'ensemble des filières.

De façon générale, cette analyse confirme que les augmentations de taille des classes générées par les redoublements n'ont sans doute pas d'effets majeurs sur les performances des 3/2, mais elle suggère aussi que quelques élèves supplémentaires de très bon niveau peuvent suffire à induire les professeurs à produire un cours trop ambitieux, cette distorsion se faisant au détriment des élèves de 3/2.

S'agissant finalement des stratégies susceptibles d'être suivies par les établissements, elles demandent de comparer les effets directs et indirects induits par les redoublements. Un établissement ayant chaque année en moyenne deux redoublants classés dans chacune de ses classes étoilées et qui déciderait d'interdire ce type de redoublements générerait par exemple deux types d'effets : d'une part, une amélioration des performances moyennes des élèves de 3/2 en classes étoilées que nos estimations permettent d'évaluer à environ un élève 3/2 de plus admis dans une école du top 5 du concours Mines-Ponts dans chacune de ces classes (où $1 = 2 \times 0.5$), mais d'autre part une baisse mécanique du nombre de 5/2 admis dans les meilleures écoles de l'ordre de -1 admis dans des écoles du top 5 dans chaque classe étoile chaque année (puisque la probabilité d'admission dans le top 5 passe de 0.25 à 0.6 entre la 3/2 et la 5/2 pour les élèves classés). Si en outre l'établissement tire parti de la baisse du nombre de redoublants pour accroître dans des proportions équivalentes la taille des cohortes et le nombre d'élèves admis en 3/2 par classe (en sorte de raisonner à nombre d'élève-année constant), il obtiendra en moyenne 0.05 admissions supplémentaires dans des écoles du top 5 chaque année, pour un bilan total finalement très équilibré.

6 Conclusion

Le redoublement est une pratique aujourd'hui largement répandue dans l'enseignement supérieur. Dans cet article nous en étudions les effets sur les non-redoublants, dans le contexte particulier des classes préparatoires aux écoles d'ingénieurs en France.

À l'issue de notre travail, il apparaît que les non-redoublants tendent à être significativement moins performants quand ils doivent partager leurs professeurs et leur classe avec un nombre important de redoublants à fort potentiel académique, qui étaient déjà parmi les bons élèves à l'issue de leur première année et dont le redoublement est motivé non pas par l'échec à être admis dans une bonne école, mais par le désir d'obtenir une admission dans l'une des toutes meilleures écoles. En revanche les performances des non-redoublants apparaissent largement insensibles à la présence plus ou moins forte de redoublants dont le potentiel académique se situe dans la moyenne de leur classe.

Pris ensemble, ces résultats suggèrent que les non-redoublants réussissent tout aussi bien qu'il y ait ou non beaucoup de redoublants venant grossir leurs rangs et augmenter la taille de leurs classes. Si les redoublants ont une influence sur les non-redoublants, ce n'est pas parce qu'ils contribuent à surcharger les classes, mais parce qu'ils en modifient le profil, le paradoxe étant que leur influence apparaît d'autant plus problématique qu'ils sont eux-mêmes à fort potentiel académique. Ce résultat est en contradiction avec l'hypothèse généralement admise selon laquelle les bons élèves suscitent des effets d'entraînement vertueux, bénéfiques à l'ensemble des élèves de leurs classes. C'est un mécanisme très différent qui semble ici à l'œuvre : un surcroît de bons élèves semble plutôt induire les professeurs à produire des cours trop ambitieux pour les nouveaux arrivants. Cette hypothèse permet d'interpréter tout à la fois l'effet direct très positif du redoublement sur les redoublants et les effets externes négatifs des redoublants sur les non-redoublants dans les classes où sont concentrés les redoublants les plus forts. De nouvelles recherches devront toutefois être conduites pour tester cette hypothèse plus directement que nous n'avons pu le faire, et dans d'autres contextes que celui des classes préparatoires aux écoles d'ingénieurs.

TABLEAU 2.1 – Effet du nombre de redoublants sur la composition et les performances des 3/2 l'année suivante

Variables dépendantes	Toutes classes confondues	Classes étoilées	Classes non étoilées
<i>Panel A : Effectifs et caractéristiques des 3/2</i>			
Nb. de 3/2 par classe	0.099 (0.084)	0.121 (0.093)	0.093 (0.131)
Nb. de 3/2 de sexe féminin par classe	0.013 (0.047)	0.019 (0.084)	0.028 (0.080)
Nb. de 3/2 boursiers par classe	0.036 (0.059)	0.030 (0.053)	0.046 (0.080)
Résultats standardisés des 3/2 au bac	0.007* (0.003)	0.006 (0.005)	0.007 (0.004)
<i>Panel B : Nombre de 5/2 et tailles des classes</i>			
Nb. de 5/2 par classe	0.845*** (0.051)	0.906*** (0.080)	0.807*** (0.089)
Nb. de 5/2 d'un autre lycée par classe	-0.027 (0.024)	-0.023 (0.037)	-0.024 (0.022)
Taille des classes	0.912*** (0.092)	0.974*** (0.123)	0.895*** (0.119)
<i>Panel C : Résultats des 3/2 au concours Mines-Ponts</i>			
Nb. de 3/2 inscrits par classe	-0.031 (0.076)	0.032 (0.096)	-0.063 (0.112)
Nb. de 3/2 admissibles par classe	-0.023 (0.067)	-0.032 (0.105)	0.001 (0.049)
Nb. de 3/2 par classe reçus à au moins une école	-0.068 (0.065)	-0.102 (0.111)	-0.024 (0.041)
Nb. de 3/2 par classe reçus à une des 5 écoles les plus sélectives	-0.052 (0.045)	-0.106 (0.087)	0.011 (0.025)
N	236	236	236

Note : Panel des 59 établissements × filières observés entre 2012 et 2016 sans changement de nombre de classes sur la période. La première colonne concerne l'ensemble des classes, alors que la deuxième colonne se focalise sur les classes étoilées et la troisième colonne sur les classes non étoilées.

Chaque ligne correspond à une variable dépendante distincte, à savoir le nombre de 3/2 par classe (ligne 1), le nombre de 3/2 de sexe féminin par classe (ligne 2), le nombre de 3/2 boursiers par classe (ligne 3), et les résultats moyens standardisés des 3/2 au baccalauréat (ligne 4). Les lignes 5 à 7 décrivent le nombre total de 5/2 par classe (ligne 5), le nombre de 5/2 provenant d'un autre établissement par classe (ligne 6), et le nombre d'élèves par classe (ligne 7). Enfin, les lignes 8 à 11 correspondent aux résultats des 3/2 au concours Mines-Ponts.

Chaque cellule correspond à une régression spécifique et montre l'impact du nombre 3/2 par classe ayant pris la décision de redoubler à la fin de l'année. Chaque régression inclut un ensemble d'effets fixes établissements interagis avec la filière considérée (MP ou PSI), des effets fixes années et des contrôles pour le nombre par classe de 3/2 de sexe féminin, le nombre par classe de 3/2 boursiers, et pour les résultats moyens standardisés des 3/2 au baccalauréat (excepté pour les lignes 3 à 5 où la variable dépendante n'est pas incluse dans les contrôles). Les erreurs types sont indiquées entre parenthèses.

* significatif à 10%. ** significatif à 5%. * significatif à 1%.

TABLEAU 2.2 – Effet du nombre de redoublants classés et non classés sur la composition et les performances des 3/2 l'année suivante

Variables dépendantes	Toutes classes confondues	Classes étoilées	Classes non étoilées
<i>Panel A : Effectifs et caractéristiques des 3/2</i>			
Nb. de 3/2 par classe :			
Nb. de redoublants classés par classe	-0.103 (0.230)	-0.070 (0.257)	-0.273 (0.370)
Nb. de redoublants non classés par classe	0.124 (0.089)	0.145 (0.101)	0.139 (0.138)
Nb. de 3/2 de sexe féminin par classe :			
Nb. de redoublants classés par classe	-0.071 (0.138)	-0.321 (0.195)	0.174 (0.237)
Nb. de redoublants non classés par classe	0.023 (0.048)	0.061 (0.087)	0.010 (0.089)
Nb. de 3/2 boursiers par classe :			
Nb. de redoublants classés par classe	0.018 (0.248)	-0.012 (0.225)	0.034 (0.342)
Nb. de redoublants non classés par classe	0.039 (0.061)	0.035 (0.059)	0.047 (0.082)
Résultats standardisés des 3/2 au bac :			
Nb. de redoublants classés par classe	0.001 (0.009)	0.003 (0.012)	-0.002 (0.012)
Nb. de redoublants non classés par classe	0.007* (0.004)	0.006 (0.005)	0.008 (0.005)
<i>Panel B : Nombre de 5/2 et tailles des classes</i>			
Nb. de 5/2 par classe :			
Nb. de redoublants classés par classe	0.886*** (0.104)	1.551*** (0.189)	0.360 (0.242)
Nb. de redoublants non classés par classe	0.840*** (0.054)	0.826*** (0.084)	0.862*** (0.095)
Nb. de 5/2 d'un autre lycée par classe :			
Nb. de redoublants classés par classe	0.004 (0.056)	0.089 (0.062)	-0.083 (0.093)
Nb. de redoublants non classés par classe	-0.031 (0.026)	-0.036 (0.040)	-0.017 (0.025)
Taille des classes :			
Nb. de redoublants classés par classe	0.831*** (0.250)	1.548*** (0.337)	0.097 (0.374)
Nb. de redoublants non classés par classe	0.922*** (0.095)	0.903*** (0.132)	0.993*** (0.120)
<i>Panel C : Résultats des 3/2 au concours Mines-Ponts</i>			
Nb. de 3/2 inscrits par classe :			
Nb. de redoublants classés par classe	-0.129 (0.256)	-0.135 (0.279)	-0.178 (0.404)
Nb. de redoublants non classés par classe	-0.019 (0.083)	0.053 (0.102)	-0.049 (0.121)
Nb. de 3/2 admissibles par classe :			
Nb. de redoublants classés par classe	-0.377** (0.184)	-0.722** (0.301)	-0.101 (0.140)
Nb. de redoublants non classés par classe	0.021 (0.072)	0.053 (0.116)	0.013 (0.050)
Nb. de 3/2 par classe reçus à au moins une école :			
Nb. de redoublants classés par classe	-0.392** (0.186)	-0.793** (0.310)	-0.065 (0.130)
Nb. de redoublants non classés par classe	-0.027 (0.067)	-0.017 (0.116)	-0.019 (0.042)
Nb. de 3/2 par classe reçus à une des 5 écoles les plus sélectives :			
Nb. de redoublants classés par classe	-0.224 (0.147)	-0.533* (0.271)	0.053 (0.107)
Nb. de redoublants non classés par classe	-0.030 (0.040)	-0.053 (0.081)	0.005 (0.019)
N	236	236	236

Note : Cette table se concentre sur le même échantillon que la table 2.1. La première colonne concerne l'ensemble des classes, alors que la deuxième colonne se focalise sur les classes étoilées et la troisième colonne sur les classes non étoilées.

Chaque ligne correspond à une variable dépendante distincte, à savoir le nombre de 3/2 par classe (ligne 1), le nombre de 3/2 de sexe féminin par classe (ligne 2), le nombre de 3/2 boursiers par classe (ligne 3), et les résultats moyens standardisés des 3/2 au baccalauréat (ligne 4). Les lignes 5 à 7 décrivent le nombre total de 5/2 par classe (ligne 5), le nombre de 5/2 provenant d'un autre établissement par classe (ligne 6), et le nombre d'élèves par classe (ligne 7). Enfin, les lignes 8 à 11 correspondent aux résultats des 3/2 au concours Mines-Ponts.

Chaque cellule correspond à une régression spécifique et montre l'impact du nombre par classe de 3/2 classés (1^{er} interligne) et non classés (2^{ème} interligne) au concours Mines-Ponts ayant pris la décision de redoubler à la fin de l'année. Chaque régression inclut un ensemble d'effets fixes établissements interagis avec la filière considérée (MP ou PSI), des effets fixes années et des contrôles pour le nombre par classe de 3/2 de sexe féminin, le nombre par classe de 3/2 boursiers, et pour les résultats moyens standardisés des 3/2 au baccalauréat (excepté pour les lignes 3 à 5 où la variable dépendante n'est pas incluse dans les contrôles). Les erreurs types sont indiquées entre parenthèses.

* significatif à 10%. ** significatif à 5%. * significatif à 1%.

Annexe

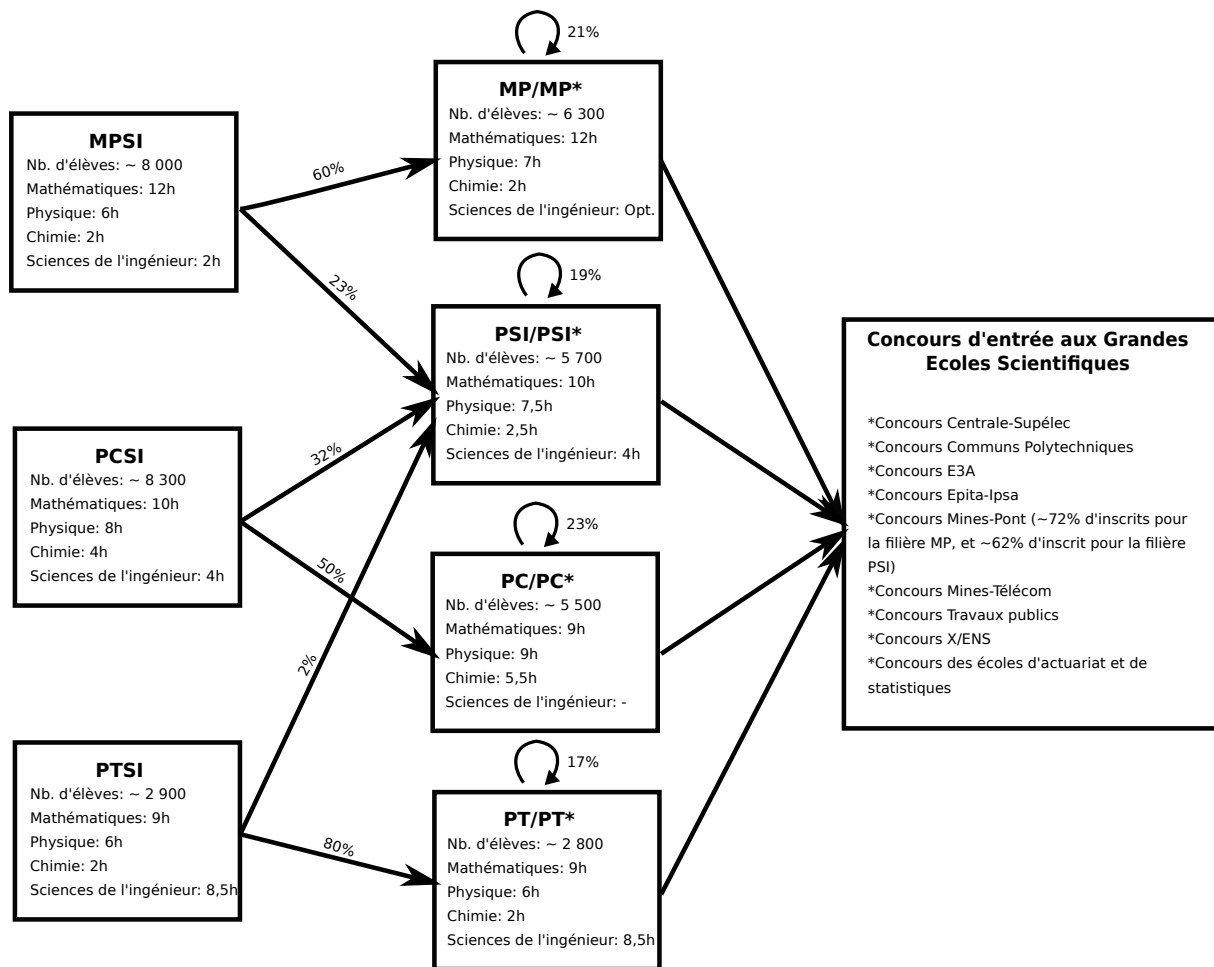


FIGURE 2.A1 – Classes préparatoires scientifiques

Sources : Panels des élèves du secondaires, années scolaires 2011-2012 à 2014-2015, service statistique du ministère de l'éducation nationale.

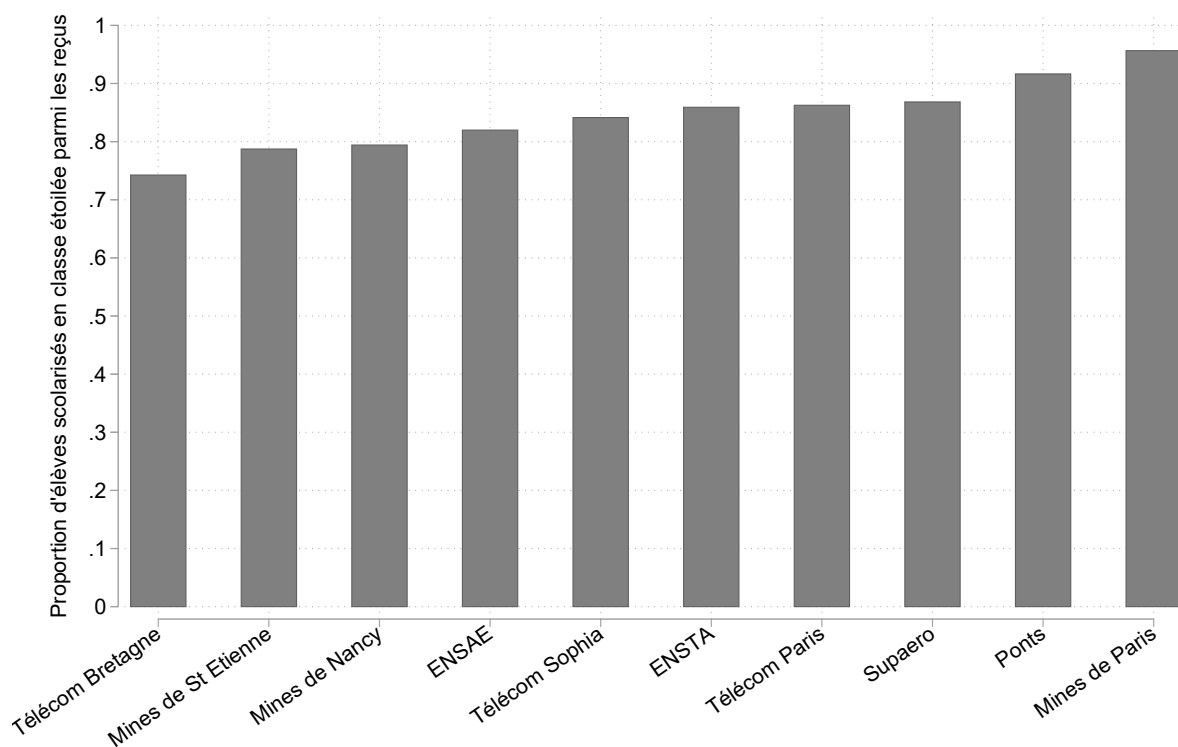


FIGURE 2.A2 – Proportion d’élèves provenant d’une classe étoilée parmi les reçus des écoles du concours Mines-Ponts

Note : Cette figure se concentre sur l’ensemble des élèves ayant passé le concours Mines-Ponts MP ou PSI entre 2012 et 2016. Pour chaque école de ce concours, cette figure indique la proportion d’élèves scolarisés en classe étoilée parmi les élèves ayant obtenu un classement inférieur au seuil d’admission de l’école considérée.

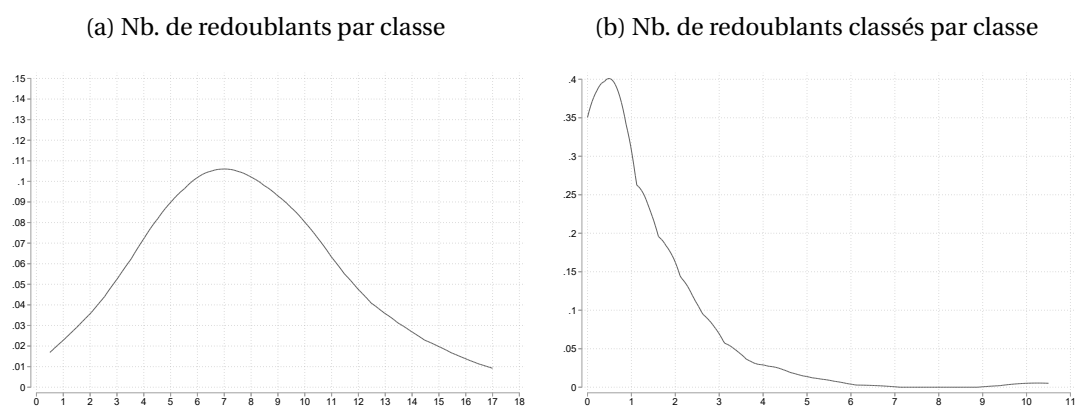


FIGURE 2.A3 – Distribution du nombre de redoublants et du nombre de redoublants classés par classe

Note : Ces deux figures se concentrent sur le même échantillon que la table 2.1. La figure 2.A3a montre la distribution du nombre de 3/2 par classe ayant pris la décision de redoubler, et la figure 2.A3b représente la distribution du nombre par classe de 3/2 classés au concours Mines-Ponts ayant pris la décision de redoubler.

TABLEAU 2.A1 – Statistiques descriptives — Échantillon complet

Variables dépendantes	Toutes classes confondues	Classes étoilées	Classes non étoilées
<i>Panel A : Effectifs et caractéristiques des 3/2</i>			
Nb. de 3/2 par classe	25.8 [8.7]	29.0 [7.4]	25.7 [9.1]
Nb. de 3/2 de sexe féminin par classe	5.6 [2.8]	5.5 [2.8]	5.9 [3.2]
Nb. de 3/2 boursiers par classe	2.6 [3.9]	2.9 [3.8]	2.7 [4.1]
Résultats standardisés des 3/2 au bac	0.9 [0.5]	1.4 [0.4]	0.8 [0.4]
<i>Panel B : Nombre de 5/2 et tailles des classes</i>			
Nb. de 5/2 par classe	7.1 [3.8]	7.4 [3.9]	7.1 [4.0]
Nb. de 5/2 d'un autre lycée par classe	0.6 [1.0]	0.6 [1.3]	0.6 [1.1]
Taille des classes	33.4 [9.7]	36.9 [7.1]	33.3 [10.2]
<i>Panel C : Résultats des 3/2 au concours Mines-Ponts</i>			
Nb. de 3/2 inscrits par classe	17.0 [10.7]	26.9 [8.8]	15.0 [9.9]
Nb. de 3/2 admissibles par classe	4.9 [8.1]	15.1 [11.5]	1.8 [3.8]
Nb. de 3/2 par classe reçus à au moins une école	4.6 [7.8]	14.4 [11.2]	1.7 [3.7]
Nb. de 3/2 par classe reçus à une des 5 écoles les plus sélectives	2.3 [5.3]	7.9 [8.9]	0.5 [1.5]
N	896	355	789

Note : Panels des 243 établissements × filières observés pendant au moins deux années consécutives entre 2012 et 2016. La première colonne concerne l'ensemble des classes, alors que la deuxième colonne se focalise sur les classes étoilées et la troisième colonne sur les classes non étoilées.

Chaque ligne correspond à une variable distincte, à savoir le nombre de 3/2 par classe (ligne 1), le nombre de 3/2 de sexe féminin par classe (ligne 2), le nombre de 3/2 boursiers par classe (ligne 3), et les résultats moyens standardisés des 3/2 au baccalauréat (ligne 4). Les lignes 5 à 7 décrivent le nombre total de 5/2 par classe (ligne 5), le nombre de 5/2 provenant d'un autre établissement par classe (ligne 6), et le nombre d'élèves par classe (ligne 7). Enfin, les lignes 8 à 11 correspondent aux résultats des 3/2 au concours Mines-Ponts.

Chaque cellule indique la moyenne de la variable précisée à gauche dans l'échantillon considéré. Les écarts types sont rapportés entre crochet.

* significatif à 10%. ** significatif à 5%. * significatif à 1%.

TABLEAU 2.A2 – Statistiques descriptives — Échantillon de travail

Variables dépendantes	Toutes classes confondues	Classes étoilées	Classes non étoilées
<i>Panel A : Effectifs et caractéristiques des 3/2</i>			
Nb. de 3/2 par classe	29.6 [6.2]	27.9 [7.0]	31.2 [6.7]
Nb. de 3/2 de sexe féminin par classe	6.5 [2.2]	5.2 [2.6]	7.7 [3.1]
Nb. de 3/2 boursiers par classe	3.5 [4.1]	2.8 [3.7]	4.0 [4.8]
Résultats standardisés des 3/2 au bac	1.186 [0.340]	1.418 [0.311]	1.018 [0.382]
<i>Panel B : Nombre de 5/2 et tailles des classes</i>			
Nb. de 5/2 par classe	7.8 [3.3]	7.7 [3.9]	7.8 [4.0]
Nb. de 5/2 d'un autre lycée par classe	0.5 [0.9]	0.5 [1.3]	0.5 [1.1]
Taille des classes	37.8 [6.0]	36.1 [7.0]	39.5 [6.7]
<i>Panel C : Résultats des 3/2 au concours Mines-Ponts</i>			
Nb. de 3/2 inscrits par classe	24.2 [8.2]	26.5 [7.9]	22.5 [9.5]
Nb. de 3/2 admissibles par classe	8.3 [6.8]	14.8 [9.8]	2.7 [4.5]
Nb. de 3/2 par classe reçus à au moins une école	7.9 [6.5]	14.1 [9.6]	2.5 [4.3]
Nb. de 3/2 par classe reçus à une des 5 écoles les plus sélectives	3.8 [4.7]	7.3 [7.4]	0.7 [1.8]
N	236	236	236

Note : Cette table se concentre sur le même échantillon que la table 2.1. La première colonne concerne l'ensemble des classes, alors que la deuxième colonne se focalise sur les classes étoilées et la troisième colonne sur les classes non étoilées.

Chaque ligne correspond à une variable distincte, à savoir le nombre de 3/2 par classe (ligne 1), le nombre de 3/2 de sexe féminin par classe (ligne 2), le nombre de 3/2 boursiers par classe (ligne 3), et les résultats moyens standardisés des 3/2 au baccalauréat (ligne 4). Les lignes 5 à 7 décrivent le nombre total de 5/2 par classe (ligne 5), le nombre de 5/2 provenant d'un autre établissement par classe (ligne 6), et le nombre d'élèves par classe (ligne 7). Enfin, les lignes 8 à 11 correspondent aux résultats des 3/2 au concours Mines-Ponts.

Chaque cellule indique la moyenne de la variable précisée à gauche dans l'échantillon considéré. Les écarts types sont rapportés entre crochet.

* significatif à 10%. ** significatif à 5%. * significatif à 1%.

Chapter 3

From Employment to Engagement? Stable Jobs, Temporary Jobs, and Cohabiting Relationships

Abstract

Using French data on the work and family history of large samples of young adults, this paper provides evidence that access to permanent jobs has a much stronger impact than access to temporary jobs on the probability of entering a first cohabiting relationship as well as on the probability of having a first child. Men's and women's age at cohabiting relationship and at first child have strongly increased during the second half of the last century. According to our estimates, about 25% of the increase in age at first cohabitation and about 40% of the increase in age at first child can be explained by the rise in unemployment and in the share of temporary jobs among young workers.

JEL: J12, J64, C32.

1 Introduction

Over the past decades, it has become increasingly difficult for young men and women to enter the labor market, especially with a permanent contract (*OECD.Stat*). In this paper, we provide evidence that this rise in job insecurity likely contributed to increase the age at which individuals start cohabiting relationships as well as the age at which they have their first child. Using detailed information on the work and family history of large representative samples of young French adults, we show that access to permanent jobs has actually a much stronger impact on the probability of entering a cohabiting relationship and forming a family than access to temporary jobs. According to our estimates, about 25% of the increase in age at first cohabiting relationship and about 40% of the increase in age at first birth observed over the recent decades can be explained by the decline in the share of permanent labor contracts among younger workers.

Our findings contribute to a long-standing strand of research that explores the relationship between employment and family formation¹. Several papers have highlighted the importance of job security for fertility decisions (See Pailhé and Solaz, 2012; Prifti and Vuri, 2013 or Modena, Rondinelli and Sabatini, 2014 for instance). Much less is known however on the differential impact of long-term and short-term contracts for cohabiting relationships or marriage, and our main contribution is to show that access to permanent job positions has different implications for cohabitation than access to temporary job positions. Using both an event study and a timing of events methodology (Abbring and van den Berg, 2003, we are also able to investigate whether access to employment impacts fertility decisions directly or only through cohabiting relationships. We also investigate whether the links between employment and cohabitation changed across cohorts born in the second half of the last century.

Few papers have provided causal evidence on the links between employment and marriage or cohabiting relationships. The two basic reasons for this lack of evidence are reverse causality and selection issues. Individuals who find a job sooner are likely to exhibit unobserved characteristics that may also influence the timing of their cohabiting relationship and both transitions may impact each other.

The few papers that account for these issues are suggestive that access to employment has a positive impact on cohabiting relationships. For instance, building on an RCT, Mamun (2008) finds that worked hours and earnings increase women's probability to marry, but have

¹See, e.g., Ekert-Jaffé and Solaz (2001); Prioux (2003) or Vergauwen, Neels and Wood (2016) on the links between employment status and couple formation and Adsera (2005); Goldstein et al. (2013) or Huttunen and Kellokumpu (2016) for employment and fertility decisions.

no effects on men. In another setting, Black, McKinnish and Sanders (2003) find that the coal boom that happened in the seventies in the US increased women's marriage rate.

Relative to these articles, our paper does not focus on the overall probability to marry, but on the timing of cohabiting relationships. As it happens, it is mainly the timing of cohabitations which changed a lot over the recent decades, not the overall rate of cohabitation. In France, cohabitation rates have remained completely flat over the recent years, whereas the average age at first cohabitation increased significantly (Prioux, 2003). These delays in age at first cohabiting relationship likely have important implications for fertility decisions. Postponements in cohabiting relationships are likely to generate postponements in birth that mechanically result in either an overall decrease in fertility or in shorter intervals between births. Evidence are mixed regarding the links between birth postponements and completed fertility². There exist however a few papers documenting the negative consequences of shorter birth spacings for children's outcomes (see e.g., Buckles and Munnich, 2012). Additionally, birth postponements may have adverse public health consequences as mothers' and children's health hazards increase with maternal age.

Most papers on the links between employment and the timing of marriages (or cohabiting relationships) use duration models, with or without frailty (See, e.g., Ekert-Jaffé and Solaz, 2001, 2002; Kalmijn, 2011; Vergauwen, Neels and Wood, 2016). One feature of these duration models is that they do not account for the potential correlation between unobservable characteristics impacting employment and cohabitation transitions. They cannot account for reverse causality either³. To overcome these issues, we develop a timing of events analysis, using the model developed by Abbring and van den Berg (2003). Assuming that individuals cannot predict in advance the exact date of occurrence of the events under consideration or do not act upon this information before the events happen (the no-anticipation assumption), this model enables to assess the causal relationships between these events, even when their timings are affected by common unobserved factors. This analysis is suggestive that stable jobs have an impact on the probability of entering cohabiting relationships which is much stronger than the impact of temporary jobs. First stable jobs multiply men's and women's instantaneous probability to enter a first cohabiting relationship by 3.8 and 3.4 respectively. By contrast, first temporary jobs multiply women's

²Pailhé and Solaz (2012) for instance, find that employment uncertainty tends to delay first parenthood but has a limited impact on completed fertility in France. In other European countries however, d'Albis, Greulich and Ponthière (2017) show that women who postpone childbirth for economic reasons are less likely to become mothers. Also, Goldstein et al. (2013) find that OECD countries that were hit hard by the recession experienced a stronger decrease in fertility rates.

³An exception is Olsen and Farkas (1990). Using a duration model with participation in a program aimed at helping disadvantaged youths on the job market, they find that better economic opportunities accelerate the path to marriage.

instantaneous probability to enter a first cohabiting relationship by 1.1 only and they have no significant impact for men.

Using the same model, we further investigate the links between employment, cohabitation and fertility. We provide evidence that access to stable employment affects positively the probability of having a first child both indirectly (through its effect on cohabitation) and directly. We also show that the links between employment and cohabitation are not exactly the same for older cohorts and for more recent cohorts. In particular, women's access to employment used to be negatively affected by cohabitation, whereas it does not appear to be the case anymore for more recent cohorts.

Abbring and van den Berg's methodology shares common features with Lillard's simultaneous hazard model (Lillard, 1993). This methodology has been used by Lillard and Waite (2000), Aassve et al. (2006), and Niedergesäss (2013) to study the relationship between employment, cohabitation or marriage and fertility. These three papers find a positive association between employment status and the timing of marriage or cohabitation for men, and a negative association for women (except for Aassve et al., 2006 who find a positive association for both genders). They also find no relationship between marriage or cohabitation and transition into employment for men, but they find a negative association for women. Relative to this literature, the main contribution of our paper is to focus on the differential impact of stable and temporary jobs. The proportion of temporary jobs has increased dramatically in most countries over the past decades. One reason for this increase is the idea that flexibility may be a remedy to unemployment, and especially to youth unemployment. Understanding whether temporary jobs have (or not) similar implications as stable jobs for cohabiting trajectories and fertility decisions may shed some light on the indirect consequences of such policies.

The paper is organized as follows. Section 2 focuses on cohorts born in or after 1954 using the French survey *Families and Employers*. This section first describes the survey and the main variables of interest. Section 2 then develops an event study and a timing of events analysis to investigate the links between stable jobs, temporary jobs and cohabiting relationships. This section further develops a second timing of events analysis to include fertility decisions. Section 3 focuses on cohorts born before 1954 using the French survey *Household Budget*. Similarly to Section 2, Section 3 describes the data and develops an event study and a timing of events analysis on employment and cohabiting relationships. Section 4 discusses the policy implications of the estimation results, and Section 5 concludes.

2 Stable Jobs, Temporary Jobs, and Cohabiting Relationships

2.1 Data

To investigate whether stable and temporary jobs have similar implications for family formation, this paper uses the French survey *Families and Employers – FE* hereafter – conducted jointly in 2004-2005 by the French National Institute for Demographic Studies (INED) and the French National Institute for Statistical and Economic Studies (INSEE). This survey provides detailed retrospective information on the work and family history of 9,547 individuals representative of the French population born between 1954 and 1985.

The *FE* survey provides detailed employment calendars where individuals indicate for each year starting from age 18 their employment status. These calendars distinguish year-periods during which individuals had positions that lasted less than 6 months only (hereafter temporary job positions) and year period during which they had at least one job that lasted more than 6 months (hereafter, stable job positions). We use these retrospective information to construct the two following variables: t_{sj} and t_{tj} . The variable t_{sj} records when individuals started their first stable job i.e., t_{sj} corresponds to the first year when individuals indicate that they were employed for more than six months, and t_{tj} records when individuals started their first temporary position i.e., t_{tj} corresponds to the first year when individuals indicate that they were employed or unemployed for less than six months⁴.

With respect to family formation, individuals are asked to indicate the year when they started their first cohabiting relationship as well as the months and years of birth of all their children. We use these information to construct the three following variables: t_r , t_{ch} and t_{co} . The variable t_r indicates the year when individuals started their first cohabiting relationship, t_{ch} records the year when they had their first child, and t_{co} is a proxy for the year of conception of individuals' first child⁵.

Noticeably, t_{sj} and t_{tj} are left censored at age 18 and t_{sj} , t_{tj} , t_r , t_{ch} and t_{co} are right censored at the time of the survey. These five variables are also interval censored: the

⁴Unfortunately, the *FE* survey does not distinguish between periods when individuals were unemployed for less than six months in-between two long-term statuses (employment, education, inactivity or education/training) from years when individuals were employed for less than six months in-between long-term or short-term employment or unemployment spells. Similarly to Pailhé and Solaz (2012) we consider these unstable positions as temporary jobs.

⁵ $t_{co} = t_{ch} - 1$ for children born between January 1st and September 30th and $t_{co} = t_{ch}$ for children born between October 1st and December 31st.

survey records during which years the events of interest happened but the exact timings are unknown. Both the event study and the timing of events analysis account for these features.

For control variables in the timing of events analysis, the *FE* survey provides information on whether and when individuals finished their initial schooling, on individuals' highest diploma, on their religious beliefs and on age at residential independence. Additionally, for years 1968 to 2004 we use publicly available information from the National Institute for Statistical and Economic Studies (INSEE) to control for yearly unemployment rates at the national level. For previous years, we use information from the Research and Expertise on the World Economy website⁶.

Regarding sample selection, this paper focuses separately on men and women who have completed their initial schooling and have complete information regarding their retrospective calendars (schooling, employment and family). These two samples represent about 93% of the initial sample and Table 3.A1 in the Appendix provides descriptive statistics for these samples. Additionally, the event study is restricted to subsamples of individuals who experienced the event of interest by the time of the survey and after age 18. By contrast, the timing of events analysis includes individuals with right censored information but the no-anticipation assumption requires to drop individuals who experienced the events of interest during the same year. Also, to avoid left censoring the timing of events analysis focuses on individuals who experienced every event of interest after age 18, and the analysis starts at age 18 and ends at the time of the survey or at age 35⁷. Tables 3.A2 and 3.A3 in the Appendix provide descriptive statistics on the subsamples used for the timing of events analysis. Table 3.A4 and 3.A5 in the Appendix show the characteristics of individuals excluded from the timing of events analysis, namely individuals who experienced at least two events during the same year.

The different samples described in Tables 3.A1 to 3.A3 underline that the vast majority of individuals had held a stable job by the time of the survey (89% to 92% for women depending on the specification, and 94% to 96% for men). A sizable proportion of individuals had also held a temporary job by the time of the survey (42% to 50% depending on the specification). Half of the men had entered a first stable job by age 20, half of the women by age 21, and half of the men and women had entered a first temporary job by age 20. Most individuals had also

⁶The National Institute for Statistical and Economics Studies also provides information on yearly unemployment rates by gender and age groups from the year 1968 onwards. We check that the results are robust to using these more detailed unemployment rates that are not available for previous years.

⁷In the initial sample, for each event, more than 95% of individuals who had experienced the event before the survey had experienced it by age 35.

lived in a cohabiting relationship (84% to 90% for women and to 76% to 83% for men) and had a child (67% to 77% for women and 55% to 66% for men). More women than men had lived in a cohabiting relationship and had a child as they entered cohabiting relationships and parenthood younger (about two years before men).

2.2 Event Study

This subsection documents sharp changes in individuals' probability to have entered a cohabiting relationship just around the year when they entered their first job, and in particular their first stable job as opposed to their first temporary job.

Methodology

Using the subsamples described in Table 3.A1 in the Appendix, we construct a panel where each individual i is observed every year s between her year of birth and the year of the survey minus one (the year of the survey is not observed from January to December contrary to other years). With this panel, we define three variables, t^{e_x} with $e_x = \{sj, tj, r\}$, indicating time-distance to event e_x such that $t^{e_x} = 0$ for the year during which event e_x happens (event e_x being either individuals' first stable job, first temporary job or first cohabiting relationship). For each event $e_y = \{sj, tj, r\}$ we also define $Y_{i,s,t}^{e_y}$ which equals one for years after event e_y happened and equals zero otherwise (including during the year when event e_y happened). We estimate the following two sets of equations separately on subsamples of men and women for which event e_x happened between age 18 and the year of the survey minus one:

$$Y_{i,s,t^{sj}}^r = \sum_{\substack{j=-5 \\ j \neq 0}}^{10} \alpha_j^{sj,r} \mathbb{1}[j = t^{sj}] + \sum_k \beta_k^{sj,r} \mathbb{1}[k = age_{i,s}] + \sum_l \gamma_l^{sj,r} \mathbb{1}[l = s] + v_{i,s,t}^{sj,r} \quad (3.1)$$

$$Y_{i,s,t^{tj}}^r = \sum_{\substack{j=-5 \\ j \neq 0}}^{10} \alpha_j^{tj,r} \mathbb{1}[j = t^{tj}] + \sum_k \beta_k^{tj,r} \mathbb{1}[k = age_{i,s}] + \sum_l \gamma_l^{tj,r} \mathbb{1}[l = s] + v_{i,s,t}^{tj,r}$$

$$Y_{i,s,t^r}^s = \sum_{\substack{j=-5 \\ j \neq 0}}^{10} \alpha_j^{r,sj} \mathbb{1}[j = t^r] + \sum_k \beta_k^{r,sj} \mathbb{1}[k = age_{i,s}] + \sum_l \gamma_l^{r,sj} \mathbb{1}[l = s] + v_{i,s,t}^{r,sj} \quad (3.2)$$

$$Y_{i,s,t^r}^t = \sum_{\substack{j=-5 \\ j \neq 0}}^{10} \alpha_j^{r,tj} \mathbb{1}[j = t^r] + \sum_k \beta_k^{r,tj} \mathbb{1}[k = age_{i,s}] + \sum_l \gamma_l^{r,tj} \mathbb{1}[l = s] + v_{i,s,t}^{r,tj}$$

Each regression includes a full set of event-year dummies, age dummies and year dummies to control non parametrically for life-cycle trends and time trends, and the event time-distance $t^{e_x} = 0$ is the reference. Figures 3.1a to 3.1d show the parameters of interest

$\alpha_j^{e_x, e_y}$ for each specification estimated separately for men and women. For event $e_y = r$ the main specifications consider a non-balanced panel of individuals observed between 5 years before their first stable or temporary job and either 2003-2004 or 10 years later to include as many observations as possible. Similarly, for events $e_y = \{sj, tj\}$ the main specifications consider a non-balanced panel of individuals observed either between 5 years before their first cohabiting relationship or age 18 and either 2003-2004 or 10 years later⁸. Figures 3.A1a to 3.A1d in the Appendix show the results on balanced panels of individuals observed every years between 5 years before event e_x and 10 years later.

In this setting, identification relies on the assumption that the timing of e_x is not determined by the outcome, here the event e_y . Under this assumption, conditional on age and year, there should be no discontinuity in $v_{i,s,t}^{e_x, e_y}$ around the year when event e_x happens and the short-term impact of event e_x on $Y_{i,s,t}^{e_y}$ is obtained by comparing $Y_{i,s,t}^{e_y=0}$ to $Y_{i,s,t}^{e_y=1}$. At this stage, two points are worth noticing. First, identification of Equations (3.1) requires that first stable and temporary jobs are not determined by first cohabiting relationships, and identification of Equations (3.2) requires that first cohabiting relationships are not determined by first stable and temporary jobs. This means that either first jobs and first cohabiting relationships are mutually independent or that only one set of equations is identified. Figures 3.1a to 3.1d lend support to the idea that first jobs and first cohabiting relationships are not mutually independent and that only Equations (3.1) are identified. Second, identification requires that either unobserved characteristics are constant over time or that there are not correlated with the timing of e_x . If there was another event e_{x2} whose timing was correlated with e_x it would bias the results. In particular, if the timing of individuals' first stable job was correlated with the timing of their first temporary job this event study could wrongly find a significant impact of first temporary jobs on first cohabiting relationships that would be driven by individuals' first stable job. The timing of events analysis in the next subsection accounts for those two issues.

Results

Figures 3.1a and 3.1b show the estimated parameters of interest for Equations (3.1) ($\alpha_{j=-5,\dots,10}^{sj,r}$ and $\alpha_{j=-5,\dots,10}^{tj,r}$) and the corresponding 95% confidence intervals estimated separately for men and women⁹. These figures first show that men's and women's probability to have entered a cohabiting relationship does not vary significantly with respect to the time-distance to first stable or temporary jobs before these events have actually happened. This gives credit to the assumption that first stable and temporary jobs

⁸ $Y_{i,s,t}^{sj}$, $Y_{i,s,t}^{tj}$ and $Y_{i,s,t}^r$ are right censored, and $Y_{i,s,t}^{sj}$ and $Y_{i,s,t}^{tj}$ are also left censored.

⁹Standard errors are clustered at the individual level.

are not determined by first cohabiting relationships. Second, Figures 3.1a and 3.1b show that men's and women's probability to have entered a cohabiting relationship increases significantly just after their first stable or temporary job. These figures also show that stable jobs have significantly larger impacts for cohabiting relationships than temporary jobs. For men and women, the probability to have entered a cohabiting relationship is respectively 6.2 and 8.6 percentage points higher once they have started their first stable job, compared to an increase of respectively 1.2 and 3.0 percentage points for temporary jobs.

Figures 3.1c and 3.1d show the estimated parameters of interest for Equations (3.2) ($\alpha_{j=-5,\dots,10}^{r,sj}$ and $\alpha_{j=-5,\dots,10}^{r,tj}$). Consistent with Figures 3.1a and 3.1b, Figures 3.1c and 3.1d show that men's and women's probability to have started a stable job increases significantly with respect to the time-distance to first cohabiting relationships before this event happened. Because individuals are more likely to start a cohabiting relationship after their first stable job, the closer they get to their first cohabiting relationship the more likely they are to have already started their stable job. As a result, it is problematic to interpret $Y_{i,s,t^{er}=1}^{es,j} - Y_{i,s,t^{er}=0}^{sj}$ as the impact of cohabitation on stable employment chances. Regarding temporary employment, Figures 3.1c and 3.1d do not show much variation in individuals' probability to have entered a temporary job around the time of first cohabiting relationship. This could either reflect independence between both events, or a smaller impact of temporary jobs on cohabitation chances.

Overall, Figures 3.1a to 3.1d provide evidence for the existence of a causal link running from stable employment to cohabiting relationships. Figures 3.A1a to 3.A1d in the Appendix show that the results are robust to using a balanced panel for each regression.

The next subsection develops a second identification strategy that enables to estimate the impact of cohabitation on employment even when the timing of cohabitation is endogenous to the timing of employment. This methodology also enables to account for the potential correlation between the timings of first stable and temporary jobs.

2.3 Timing of Events Analysis

Methodology

To study the links between stable jobs, temporary jobs and cohabiting relationships, this subsection develops a timing of events approach (Abbring and van den Berg, 2003) that accounts for the potential correlation between the three different timings. This analysis focuses on the subsamples of men and women defined in Subsection 2.1 and described in

Table 3.A2 in the Appendix. With respect to the event study results, this analysis includes individuals with right-censored information but excludes individuals with simultaneous transitions. Also, to avoid issues related to left-censoring, this analysis focuses on individuals who did not enter a cohabiting relationship before age 18, and the panel starts at age 18 instead of birth¹⁰.

We define $\theta_{i,r}$ as individuals' instantaneous probability to enter a first cohabiting relationship, $\theta_{i,sj}$ as individuals' instantaneous probability to start a first stable job and $\theta_{i,tj}$ as individuals' instantaneous probability to start a first temporary job¹¹. The three hazard rates depend on the duration t elapsed since age 18 (the initial date), time-constant observed characteristics $(x_{i,r}, x_{i,sj}, x_{i,tj})$, time-variant observed characteristics $(x_{i,r,t}, x_{i,sj,t}, x_{i,tj,t})$, time-constant unobserved characteristics $(v_{i,r}, v_{i,sj}, v_{i,tj})$, and the timings of individuals' first stable job $(t_{i,sj})$, first temporary job $(t_{i,tj})$, and first cohabiting relationship $(t_{i,r})$. Hazard rates are assumed to have a standard mixed proportional hazard specification:

$$\begin{aligned}\theta_{i,r}(t|x_{i,r}, v_{i,r}, t_{i,sj}, t_{i,tj}) &= \lambda_r(t) e^{(\beta_r x_{i,r} + \gamma_r x_{i,r,t} + \delta_{sj}^r \mathbb{1}(t > t_{i,sj}) + \delta_{tj}^r \mathbb{1}(t > t_{i,tj}) + v_{i,r})} \\ \theta_{i,sj}(t|x_{i,sj}, v_{i,sj}, t_{i,r}, t_{i,tj}) &= \lambda_{sj}(t) e^{(\beta_{sj} x_{i,sj} + \gamma_{sj} x_{i,sj,t} + \delta_r^{sj} \mathbb{1}(t > t_{i,r}) + \delta_{tj}^{sj} \mathbb{1}(t > t_{i,tj}) + v_{i,sj})} \\ \theta_{i,tj}(t|x_{i,tj}, v_{i,tj}, t_{i,r}, t_{i,sj}) &= \lambda_{tj}(t) e^{(\beta_{tj} x_{i,tj} + \gamma_{tj} x_{i,tj,t} + \delta_r^{tj} \mathbb{1}(t > t_{i,r}) + \delta_{sj}^{tj} \mathbb{1}(t > t_{i,sj}) + v_{i,tj})}\end{aligned}\quad (3.3)$$

The main parameters of interest are δ_{sj}^r , δ_{tj}^r , δ_r^{sj} , and δ_r^{tj} .

The analysis includes schooling, residential independence and yearly unemployment rates as time-variant control variables. For time-constant control variables, each specification uses whether individuals graduated from higher education or whether they graduated from high school only (as opposed to lower diploma), as well as a dummy variable indicating strong religious beliefs. For the baseline hazard rates $(\lambda_r(t), \lambda_{sj}(t), \text{ and } \lambda_{tj}(t))$ we use piecewise constant functions. Regarding the joint distribution for the unobserved characteristics we assume that they follow a discrete distribution with two points of support and unrestricted mass point locations.

Abbring and van den Berg (2003) show that under the no-anticipation assumption Model (3.3) enables to sort out causality from selection. Intuitively, the idea is to estimate whether individuals who find their first stable or temporary job later enter their cohabiting

¹⁰For ending date, we use the minimum of year at age 35 and year of the survey minus one. As mentioned in Subsection 2.1, in the initial sample, for each event more than 95% of individuals who had experienced the event before the survey had experienced it by age 35.

¹¹This paper focuses on first jobs and first cohabiting relationships so as to avoid making assumptions regarding the independence between the timings of individuals' several stable jobs, several temporary jobs or several cohabiting relationships.

relationship as fast afterward as those who had found a first stable or temporary job earlier (and similarly for the impact of first cohabiting relationships on first stable and temporary jobs). As Abbring and van den Berg (2004) underline, whether one transition happens systematically just after another provides evidence that there is a causal link running from the first transition in time to the second, as selection per se would imply a strong correlation between both timings but not a quick succession of events.

Abbring and van den Berg's empirical model is identified under the no-anticipation assumption i.e., under the assumption that individuals either do not know the exact year when their first stable job, temporary job and cohabiting relationship will happen or do not act upon this information more than a few months before the event happens. In our context where the analysis is at the year level and excludes individuals with simultaneous transitions, the effect of event e_x on event e_y would not be identified if e_x was impacting e_y one year or more before its occurrence (in anticipation). Essentially, the no-anticipation assumption holds if disclosing whether event e_x will happen during year t did not change individuals' instantaneous probability to experience event e_y in year $t - 1$. Alternatively, if event e_x was postponed from year t to year $t + 1$, it should have no impact on individuals' instantaneous probability to experience event e_y during $t - 1$. Reassuringly, the event study reveals that the probability to have entered a cohabiting relationship is not statistically different two years before the first stable or temporary job compared to one year before. Similarly, Figures 3.1c and 3.1d show no jump in the probability to have found a first stable or temporary job one year before the first cohabiting relationship.

In our context the tighter the labor or housing market the more likely it is that the no-anticipation assumption holds. In particular, for younger cohorts facing a high unemployment rate, finding a first stable job the exact year when individuals start to search is less likely than for older cohorts. As such, finding similar results for older and younger cohorts may provide evidence that the results are not driven by anticipatory behaviors. More in general, it seems reasonable to assume that individuals do not anticipate the exact date when they will start their first stable or temporary job or when they will move in with their partner more than a few months in advance or at least do not act upon this information more than a few months in advance as these events are uncertain until housing or work contracts are signed.

As a robustness test, we check that the main results hold when we also exclude individuals who experienced two events during consecutive years to account for transitions happening during different years but with only a few months in-between.

Lastly, Table 3.A4 provides descriptive statistics on individuals with simultaneous transitions. By definition, individuals who experienced simultaneous transitions are more

likely to have found a first stable or temporary job and a first cohabiting relationship. Coherently, they are also more educated, less religious and they entered their first cohabiting relationship and had a first child younger¹².

Identification also requires that observed and unobserved characteristics are independent. This is a common assumption in duration models. The main interest of Abbring and van den Berg's methodology compared to standard Cox duration models with frailty is that this model allows for correlation between the timings of the different events. This feature is also helpful compared to the event study: Abbring and van den Berg's methodology enables to estimate the impact of cohabitation on employment even when the timing of cohabitation is not exogenous to employment.

Lastly, identification requires that unobserved characteristics impacting each transition are constant over time. In other words, the model is identified provided there is no unobserved event that jointly determines the transitions of interest. This is the main reason why we estimate the links between stable jobs, temporary jobs and cohabiting relationships jointly rather than estimating separately the links between stable jobs and cohabiting relationships on the one hand and the links between temporary jobs and cohabiting relationships on the other.

Time is continuous in Model (3.3) while the *FE* survey provides interval-censored information. Using Monte Carlo simulations Gaure, Røed and Zhang (2007) show that it is feasible to recover the parameters of Model (3.3) in this context provided that the likelihood function takes into account the discrete nature of the available data. Therefore, we compute the sample likelihood using the discrete-time version of Model (3.3):

$$\begin{aligned}\theta_{i,r}^D(t_k | x_{i,r}, v_{i,r}, t_{i,sj}, t_{i,tj}) &= 1 - e^{-e^{(\beta_r x_{i,r} + \gamma_r x_{i,r,t} + \delta_{sj}^r \mathbb{1}(t > t_{i,sj}) + \delta_{tj}^r \mathbb{1}(t > t_{i,tj}) + v_{i,r} + \phi_k^r)}} \\ \theta_{i,sj}^D(t_k | x_{i,sj}, v_{i,sj}, t_{i,r}, t_{i,tj}) &= 1 - e^{-e^{(\beta_{sj} x_{i,sj} + \gamma_{sj} x_{i,sj,t} + \delta_r^{sj} \mathbb{1}(t > t_{i,r}) + \delta_{tj}^{sj} \mathbb{1}(t > t_{i,tj}) + v_{i,sj} + \phi^s j_k)}} \\ \theta_{i,tj}^D(t_k | x_{i,tj}, v_{i,tj}, t_{i,r}, t_{i,sj}) &= 1 - e^{-e^{(\beta_{tj} x_{i,tj} + \gamma_{tj} x_{i,tj,t} + \delta_r^{tj} \mathbb{1}(t > t_{i,r}) + \delta_{sj}^{tj} \mathbb{1}(t > t_{i,sj}) + v_{i,tj} + \phi^t j_k)}}\end{aligned}$$

where $\theta_{i,j}^D$ indicates the instantaneous probability that the event j happens during interval $[t_{k-1}, t_k)$. Also, $\phi_k^c = \ln\left(\int_{t_{k-1}}^k \lambda_c(t) dt\right)$, $\phi^s j_k = \ln\left(\int_{t_{k-1}}^k \lambda_{sj}(t) dt\right)$, and $\phi^t j_k = \ln\left(\int_{t_{k-1}}^k \lambda_{tj}(t) dt\right)$.

We estimate jointly the parameters of interest using maximum likelihood.

¹²It is problematic to estimate Model (3.3) for these individuals as we do not know which transition happened first when they happened during the same year.

Results

Table 2.A1 presents the estimated parameters of interest when estimating Model (3.3) separately on men and women subsamples.

The upper part of this table shows that first stable jobs have a significant positive impact on the hazard rate of first cohabiting relationships for both men and women. Estimated effects are strong and similar across gender. A first stable job multiplies women's instantaneous probability to enter a first cohabiting relationship by 3.4, and by 3.8 for men. By contrast, first temporary jobs have a smaller impact on women's instantaneous probability to enter a first cohabiting relationship (3 times lower) and no significant impact for men.

The second and third parts of Table 2.A1 show that first cohabiting relationships have a positive impact on men's and women's instantaneous probability to enter a first stable job ($\times 1.4$ and $\times 1.5$ respectively), but no significant impact for temporary jobs. These findings are consistent with Figures 3.1a and 3.1b which show an increase in men's and women's probability to have entered a stable job after cohabitation and no variation for temporary jobs.

Lastly, this table shows that temporary jobs have a significant positive impact for entry into stable jobs while stable jobs decrease individuals' instantaneous probability to enter a temporary job. This is consistent with the idea that temporary jobs may provide a first step towards stable employment, and that individuals with stable employment usually do not switch back to unstable positions.

Table 3.A6 in the Appendix show the detailed results for Model (3.3). Unsurprisingly, this table shows that individuals are more likely to enter a first stable or temporary job once they are no longer students. Schooling also decreases women's probability to start a cohabiting relationship but not men's, and non-residential independence decreases men's probability of cohabitation but not women's.

Additionally, Table 3.A7 shows the estimated distributions for unobserved characteristics impacting first cohabiting relationships, first stable jobs and first temporary jobs (respectively N_r , N_{sj} , and N_{tj}). This table shows that the timings of this three events are indeed correlated. This explains why estimations that do not account for these correlations find biased estimates (see Table 3.A8).

As a robustness check, Table 3.A9 in the Appendix show the parameters of interest for Model (3.3) when controlling for yearly unemployment rates by age groups and gender. Comfortingly, the results with this specification are very similar to the results when estimating Model (3.3) with aggregated yearly unemployment rates (Table 2.A1).

Regarding the no-anticipation assumption, Table 3.A10 in the Appendix shows the results

on subsamples of men and women who did not experience their first stable job, temporary job or cohabiting relationship during consecutive years. This table also confirms that stable employment has a stronger positive impact for cohabitation than temporary employment.

2.4 Stable Jobs, Temporary Jobs, Cohabiting Relationships and Fertility

The event study and timing of events results suggest that stable employment has important implications for cohabiting relationships. As cohabiting relationships are often a first step before childbirth, stable employment is likely to also have implications for fertility decisions. It is however an open question as to whether employment impacts fertility only through cohabiting relationships or also directly. This subsection develops a timing of event analysis to investigate the links between stable jobs, temporary jobs, cohabiting relationships and childbirths¹³.

To study the links between stable jobs, temporary jobs, cohabiting relationships and child-births, we estimate jointly a discrete version of the following model on the samples described in Table 3.A3:

$$\begin{aligned}
 \theta_{i,r}(t|x_{i,r}, v_{i,r}, t_{i,sj}, t_{i,tj}, t_{co}) &= \lambda_r(t) e^{(\beta_r x_{i,r} + \gamma_r x_{i,r,t} + \delta_{sj}^r \mathbb{1}(t > t_{i,sj}) + \delta_{ij}^r \mathbb{1}(t > t_{i,tj}) + \delta_{co}^r \mathbb{1}(t > t_{i,co}) + v_{i,r})} \\
 \theta_{i,co}(t|x_{i,co}, v_{i,co}, t_{i,sj}, t_{i,tj}, t_r) &= \lambda_{co}(t) e^{(\beta_{co} x_{i,co} + \gamma_{co} x_{i,co,t} + \delta_{sj}^{co} \mathbb{1}(t > t_{i,sj}) + \delta_{ij}^{co} \mathbb{1}(t > t_{i,tj}) + \delta_r^{co} \mathbb{1}(t > t_{i,r}) + v_{i,co})} \\
 \theta_{i,sj}(t|x_{i,sj}, v_{i,sj}, t_{i,r}, t_{i,tj}, t_{co}) &= \lambda_{sj}(t) e^{(\beta_{sj} x_{i,sj} + \gamma_{sj} x_{i,sj,t} + \delta_r^{sj} \mathbb{1}(t > t_{i,r}) + \delta_{ij}^{sj} \mathbb{1}(t > t_{i,tj}) + \delta_{co}^{sj} \mathbb{1}(t > t_{i,co}) + v_{i,sj})} \\
 \theta_{i,tj}(t|x_{i,tj}, v_{i,tj}, t_{i,r}, t_{i,sj}, t_{co}) &= \lambda_{tj}(t) e^{(\beta_{tj} x_{i,tj} + \gamma_{tj} x_{i,tj,t} + \delta_r^{tj} \mathbb{1}(t > t_{i,r}) + \delta_{ij}^{tj} \mathbb{1}(t > t_{i,sj}) + \delta_{co}^{tj} \mathbb{1}(t > t_{i,co}) + v_{i,tj})}
 \end{aligned} \tag{3.4}$$

where t_{co} is a proxy for the year of conception of individuals' first child¹⁴. Here, we use year of conception instead of year of birth for the first child as it is likely that individuals know rather precisely when they will become parents more than a few months in advance and act upon this information.

Table 2.A2 shows the estimated parameters of interest for Model (3.4) estimated separately for men and women.

This table shows that first stable jobs have an indirect impact on men's and women's probability to have a first child as they impact positively cohabiting relationships and

¹³Figures 3.A2a to 3.A2b in the Appendix suggest that nor the timing of the first stable job is exogenous to the first child or the timing of the first child is exogenous to the first job for women. In this context, it would be difficult to interpret the results of an event study

¹⁴ $t_{co} = t_{ch} - 1$ for children born between January 1st and September 30th and $t_{co} = t_{ch}$ for children born between October 1st and December 31st.

cohabiting relationships impact positively fertility decisions. This table further shows that stable jobs also have a strong direct positive impact on men's and women's instantaneous probability to have a first child ($\times 3.3$ for women and $\times 2.0$ for men). By contrast, first temporary jobs have a smaller indirect impact on fertility decisions (about 3 times smaller) and they do not have any direct positive impact. For women, first temporary jobs even have a small direct negative impact as they divide women's instantaneous probability to have a first child by 1.2.

Noticeably, Table 2.A2 stresses a significant difference across gender regarding the impact of childbirth on employment. While first children have no impact on men's employment chances they decrease women's instantaneous probability to start a first stable job. This is consistent with the literature on gender and child penalty (see, e.g., Kleven, Landais and Sogaard (2018)).

Table 3.A11 and 3.A12 in the Appendix provide the detailed results for Model (3.4). Additionally, Table 3.A13 show the results when estimating Model (3.4) on subsamples of men and women who did not enter their first stable job, first temporary job, first cohabiting relationship or had a first child during consecutive years. This table also shows that stable employment has both a direct and an indirect impact on fertility decisions. In both cases and for both gender stable employment has a stronger impact than temporary employment.

The next section investigates whether the impact of employment on cohabitation and conversely were similar across gender for cohorts born before the 1950s.

3 Gender, Employment and Cohabiting Relationships before the 1950s

3.1 Data

To investigate whether the links between employment and family formation were similar across gender for older cohorts, this paper uses a second survey: *Household Budget – HB* hereafter. This survey was conducted in 2000-2001 by INSEE and also provides retrospective information on cohabiting relationships, fertility events, and labor market events for 17,039 individuals born between 1900 and 1983.

Regarding labor market events, in the *HB* survey individuals indicate their employment status in the year they finished their initial schooling as well as each subsequent change to

an employment status that lasted at least one year. For initial and subsequent employment statuses this survey distinguishes between eleven different statuses. In particular, the *HB* survey distinguishes permanent or temporary contracts of more than a year from three more unstable situations: interim jobs, internships, and sequences of short-term statuses (interim jobs, unemployment, temporary contracts...). As a consequence, with this survey the variable t_{sj} corresponds to the first year when individuals indicate that they were employed with a permanent or temporary contract for more than a year and t_{tj} corresponds to the first year when individuals record that they spend more than a year holding an interim job, doing an internship, or switching from unemployment to interim jobs, internships or temporary contracts of less than a year. Contrary to the *FE* survey, the definition of stable and temporary jobs are more restrictive. In the *FE* survey stable jobs correspond to jobs that lasted more than six months while they correspond to jobs that lasted at least one year in the *HB* survey, and to be recorded as a period of temporary employment individuals' interim jobs or internships have to last more than a year or they have to spend more than a year switching from unemployment to interim jobs or internships or temporary contracts of less than a year. By contrast, the *FE* survey records every period when individuals held a job even when it lasted only a few months between periods of long-term employment or unemployment.

In the *HB* survey cohabiting individuals are asked to indicate the year when their current cohabiting relationship started as well as the year when their previous cohabiting relationship started (if applicable). Individuals who were not in a cohabiting relationship at the time of the survey are asked to indicate the year when their last cohabiting relationship started (if applicable). Consequently, with this survey t_r records years when individuals started their previous or last cohabiting relationship. Using the *FE* survey – which provides information about individuals' total number of cohabiting relationships – we estimate that in the *HB* survey t_r indicates the year when individuals' started their first cohabiting relationship in at least 94% of cases. In the *HB* survey about 46% of individuals are born between 1954 and 1983. In this context, it is possible to compare the results obtained with the *FE* survey with results for cohorts born in or after 1954 in the *HB* survey to assess the robustness of the information provided in the *HB* survey.

Regarding fertility events, in the *HB* survey individuals are only asked about their children's years of birth. As a consequence, t_{co} is not defined in the *HB* survey.

For control variables, the *HB* survey provides information on individuals' highest diploma and on their fathers' occupation, and the sample selection is the same as in the previous section except for left censoring. In the *HB* survey, t_{sj} and t_{tj} are left censored at school

leaving age. We also use the same information as in the previous section to control for yearly unemployment rates.

Table 3.A14 in the Appendix provides information on the main samples from which the different subsamples for the event study are drawn, and Table 3.A15 in the Appendix provides information on the subsamples for the timing of events analysis. Columns 3 and 4 in Tables 3.A14 and 3.A15 compared to columns 1 and 2 in Tables 3.A1 and 3.A2 confirm that the two surveys provide comparable information. The only noticeable difference concerns the definition of temporary jobs. As mentioned above, the definition of temporary jobs is more restrictive in the *HB* survey.

Additionally, Table 3.A15 provides information on the sample of individuals excluded from the timing of events analysis, namely individuals who entered their first stable job and their first cohabiting relationship during the same year.

The event study and the timing of events analysis in the following subsections focus on older cohorts and provides robustness checks by comparing the results obtained with the *FE* survey to results obtained with younger cohorts from the *HB* survey. Older and younger cohorts are therefore defined relative to year 1954.

3.2 Event Study

This subsection develops the same methodology as Subsection 2.2. The only difference concerns left-censoring: in the *HB* survey, $Y_{i,s,tex}^{sj}$ and $Y_{i,s,tex}^{tj}$ are left censored at school leaving age instead of age 18. Consequently, the analysis focuses on individuals who have experienced the events e_x between school leaving age and 1999-2000, and for Equations (3.2) individuals' first observation is either five years before their cohabiting relationship or the year they finished schooling.

For men born before 1954, Figures 3.2b and 3.2d exhibit similar patterns as Figures 3.1a and 3.1b: men's probability to have entered a cohabiting relationship is flat before their first stable or temporary job and increases significantly just after the first stable job but not after the first temporary job. Consistently, they are also more likely to have started a stable job as they get closer to their first cohabiting relationship, while their probability to have started a temporary job is flat around the time of first cohabiting relationship.

For women born before 1954, Figures 3.2a and 3.2c exhibit differences with respect to either Figures 3.1a and 3.1c or Figures 3.2b and 3.2d. For older cohorts of women, first stable and temporary jobs have a similar and very small impact on the probability to have entered

a cohabiting relationship. In particular, the impact of a first stable job is significantly lower for women born before 1954 than for men born before 1954, while this impact is similar if not higher for women born in or after 1954 compared to men born in or after 1954 in the *FE* survey. Figure 3.2c also shows that women’s probability to have started a stable job is quite flat before their first cohabiting relationship, and starts decreasing afterward.

Overall, the event study suggests that while the impact of finding a stable job on cohabitation is about the same for men and women in younger cohorts, it used to be stronger for men than women in older cohorts. This analysis also suggests that the employment chances of women in older cohorts used to be negatively impacted by cohabitation.

Figures 3.A3a to 3.A3c in the Appendix compare the results obtained on younger cohorts from the *HB* survey with results obtained with the *FE* survey. These figures suggest that the gender differences obtained with older cohorts from the *HB* survey are not driven by differences in the definition of first stable or temporary jobs or first cohabiting relationships.

3.3 Timing of Events Analysis

Table 2.A3 presents the results when estimating the following model on older cohorts from the *HB* survey¹⁵:

$$\begin{aligned}\theta_{i,r}(t|x_{i,r}, v_{i,r}, t_{i,sj}, t_{i,tj}) &= \lambda_r(t) e^{(\beta_r x_{i,r} + \gamma_r x_{i,r,t} + \delta_{sj}^r \mathbb{1}(t > t_{i,sj}) + v_{i,r})} \\ \theta_{i,sj}(t|x_{i,sj}, v_{i,sj}, t_{i,r}, t_{i,tj}) &= \lambda_{sj}(t) e^{(\beta_{sj} x_{i,sj} + \gamma_{sj} x_{i,sj,t} + \delta_r^{sj} \mathbb{1}(t > t_{i,r}) + v_{i,sj})}\end{aligned}\quad (3.5)$$

Similarly to Table 2.A1, Table 2.A3 shows that first stable jobs have a significant positive impact on cohabitation. This table also shows that first stable jobs used to have a significantly stronger impact on cohabitation for men than women. For cohorts born before 1954, first stable jobs multiplies the instantaneous probability to enter a first cohabiting relationship by 2.7 for men and by 1.8 for women. By contrast, Table 2.A1 does not show a similar gender difference for the impact of stable employment on cohabiting relationships. Importantly, Table 3.A17 in the Appendix replicates the same analysis on younger cohorts from the *HB* survey and does not show either a smaller impact of stable jobs on cohabitation for women compared to men. Additionally, Table 3.A18 replicates the three-events analysis on younger cohorts from the *BF* survey and also shows a similar impact of employment on cohabitation across gender.

¹⁵As very few individuals born before 1954 held a temporary position, it is difficult to estimate the impact of these positions on the hazard rate of a first cohabiting relationship and conversely.

Table 2.A3 further shows that cohabitation used to have different implications by gender on stable employment. For cohort born before 1954, cohabiting relationships decreased women's employment chances. For younger cohorts, Table 2.A1 reveals no such gender differences for the impact of cohabiting relationships on stable employment.

Table 3.A19 and 3.A20 in the Appendix present the detailed results for Model (3.5) estimated on older cohorts from the *HB* survey.

Regarding the no-anticipation assumption, Table 3.A21 in the Appendix show the results when estimating Model (3.5) on subsamples of men and women who did not enter their first stable job and their first cohabiting relationship during consecutive years. Comfortingly, this table also shows a significant positive impact of stable employment for cohabiting relationships. Additionally, Table 2.A3 show a similar pattern of results as Table 2.A1 or Table 3.A17 for men, while it has become more uncertain to succeed in entering stable employment the exact year when individuals start looking for a job. This gives some credit to the idea that the results are not driven by anticipatory behaviors.

This set of results underlines that stable employment used to have a stronger impact on cohabiting relationships for men than women, and that women's employment chances used to be negatively impacted by cohabitation. These gender differences seem to have reduced: for cohorts born in or after 1954 Section 2 finds no gender differences for the links between employment and cohabitation. Nevertheless, Subsection 2.4 underlines that the links between employment and fertility are still different across gender for younger cohorts. Unfortunately, the *HB* survey does not unable to investigate whether these gender differences were more pronounced for older cohorts¹⁶.

4 Discussion

Men and women born in the early seventies entered their first cohabiting relationship and had their first child older than cohorts born in the mid-fifties. In the samples used for the timing of events analysis, women and men entered their first cohabiting relationship respectively 0.93 years and 0.66 years later on average than women and men born in the mid-fifties, and they had their first child 1.50 and 1.15 years later. One commonly advanced explanation for these delays would be changes in social norms which culminated in May '68 and had major legal consequences with – for example – the legalization of the birth-control pill in 1967 and of abortion in 1975. As a consequence, men's and women's demand for education also increased resulting in delays in school leaving age. In our

¹⁶The *HB* survey does not provide children's months of birth to compute their time of conception, and using children's year of birth would likely violate the no-anticipation assumption.

working samples, women and men born in the early seventies finished initial schooling about 1.6 years later than women and men born in the mid-fifties. Additionally, the proportion of women holding a higher education degree increased by about 10 percentage points between cohorts born in the mid-fifties and cohorts born in the early seventies, and by about 8 percentage points for men.

In this section, we study to which extent the increasing difficulties to enter stable employment also plays a role in explaining delays in family formation. In our working samples, for cohorts born in the early seventies compared to cohorts born in the mid-fifties, the mean duration between school leaving age and age at first stable job has increased by 0.56 years on average for women and men holding a high school degree. For women and men with lower education attainments, this increase reaches 0.66 years. By contrast, individuals with a higher education degree only spend about 0.06 years longer searching for a first stable job. This delayed entry into stable employment are likely driven by increases in unemployment rates (+ 6.2 percentage points on average) and by the rise of temporary contracts among young workers. In our working samples, about 25% of men and women born in the mid-fifties had held a temporary position by age 30 compared to 50% for men and women born in the early seventies.

We use the point estimates in Tables 3.A6 and 3.A11 to document the consequences of labor market entry conditions for family formation. First, we compute expected age at first cohabitation and at first child for average men and women born in the mid-fifties. In a second step, we compute expected age at first cohabitation and at first child for similar men and women in terms of schooling length and achievements, residential independence or religious beliefs, but different in terms of labor market entry conditions. For each gender and education group, we replace the average duration between school leaving age and age at first stable or temporary job by averages obtained for cohorts born in the early seventies. These computations yield an increase in age at first cohabitation of about 0.16 years for women and 0.24 years for men, and an increase in age at first child of about 0.73 years for women and 0.36 years for men.

Compared to actual increases, this analysis therefore suggests that the increases in unemployment and in temporary labor contract explain about 17% of the increase in average age at first cohabiting relationship for women and about 36% for men. Similarly, we estimate that the increase in unemployment rates and in temporary labor contracts explains about 49% of the increase in age at first child for women, and about 31% for men.

For policy implications, an important question is whether these delays in family formation

result in a decreased total fertility. Anecdotally, we find that men and women who found their first stable job after age 30 are about 15 percentage points less likely to have a child at the time of the survey, but this estimate likely suffers from selection biases. However, if delayed family formation had no impact for completed fertility, it would likely decrease birth spacings and Buckles and Munnich (2012) found a negative impact of shorter birth spacings for children's educational outcomes.

5 Conclusion

This paper develops an event study and a timing of events analysis to investigate the links between stable employment, temporary employment and cohabiting relationships.

These analyses reveal that the impact of employment on cohabiting relationships depends on whether the job position under consideration is stable or not. First stable jobs multiplies men's and women's instantaneous probability to enter a first cohabiting relationship by respectively 3.8 and 3.4. By comparison, the impact of first temporary jobs is much smaller ($\times 1.1$ for women and no significant impact for men).

Secondly, this paper shows that stable jobs indirectly impact fertility decisions through cohabiting relationships but also have a direct impact. First stable jobs direct multiplicative impact on men's and women's instantaneous probability to have a first child is respectively $\times 2.0$ and $\times 3.3$. By contrast, temporary jobs have smaller indirect impacts and no direct impact for fertility decisions. This analysis also reveals that childbirth has different implications by gender for entry into stable employment: a first child decreases women's stable employment chances but not men's.

Lastly, this paper finds that the links between employment and cohabiting relationships used to be different across gender. For cohorts born before 1954, first stable jobs had a stronger impact on cohabiting relationships for men than women and women's employment chances used to be negatively impacted by cohabitation.

More research is needed to properly assess why stable employment impacts family formation differently than temporary employment and in particular to assess to which extent the housing market plays a role for our results.

All in all, the results reported in this paper suggest that the increasing difficulties to enter the labor market with a permanent contract plays an important role in addition to changes in social norms or in schooling attainments to explain the delays in family formation observed over the past decades. We estimate that the rise in youth unemployment and in the share of temporary contracts among young workers explains about 25% of the increase in average age at first cohabiting relationship and about 40% of the delays in fertility

decisions.

As temporary jobs do not have similar implications as stable jobs for cohabiting relationships and fertility decisions, this paper suggests that policies favoring temporary jobs at the expense of stable jobs may incidentally delay individuals' cohabiting relationships and fertility decisions.

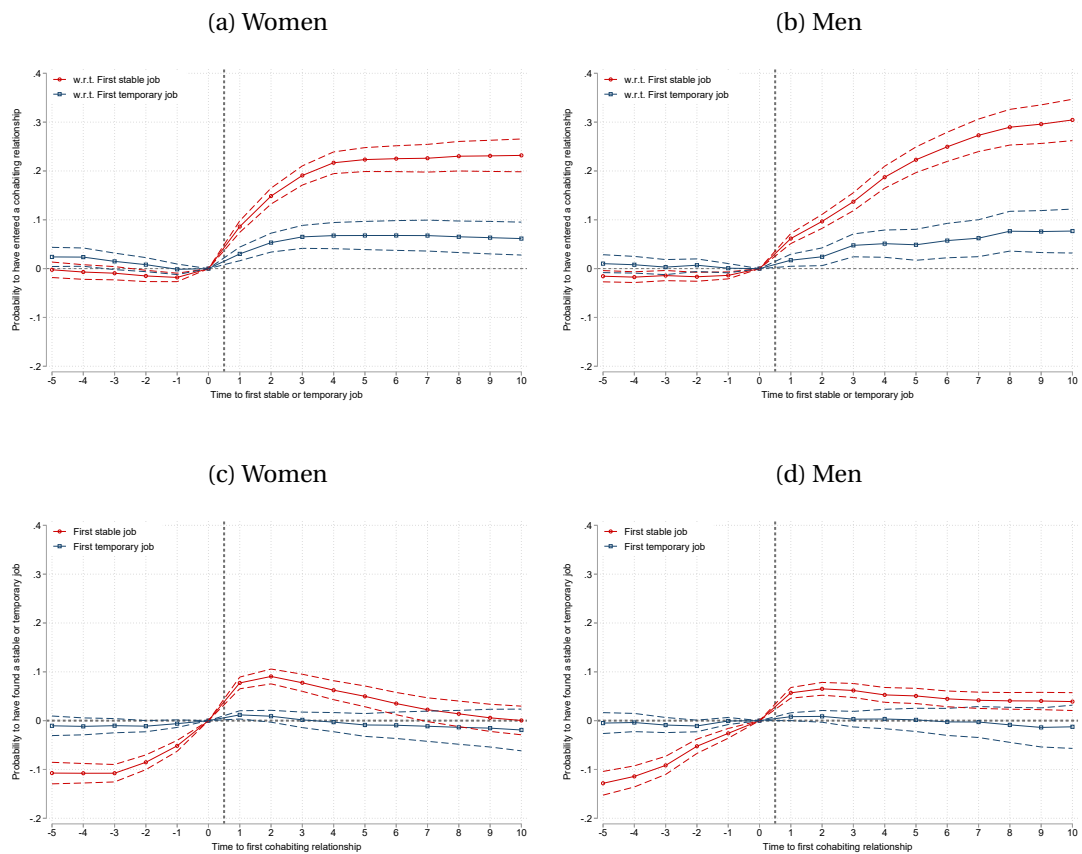


Figure 3.1 – First stable job, first temporary job and first cohabiting relationship

Note: The figures refer to the samples of men and women from the *FE* survey who completed initial schooling before the survey and experienced the event used for the horizontal axis between age 18 and 2003 or 2004. Figures 3.1a and 3.1b show the event time coefficients estimated for Equations (3.1), and Figures 3.1c and 3.1d show the event time coefficients estimated for Equations (3.2). Dotted lines represent 95% confidence intervals computed with robust standard errors clustered at the individual level.

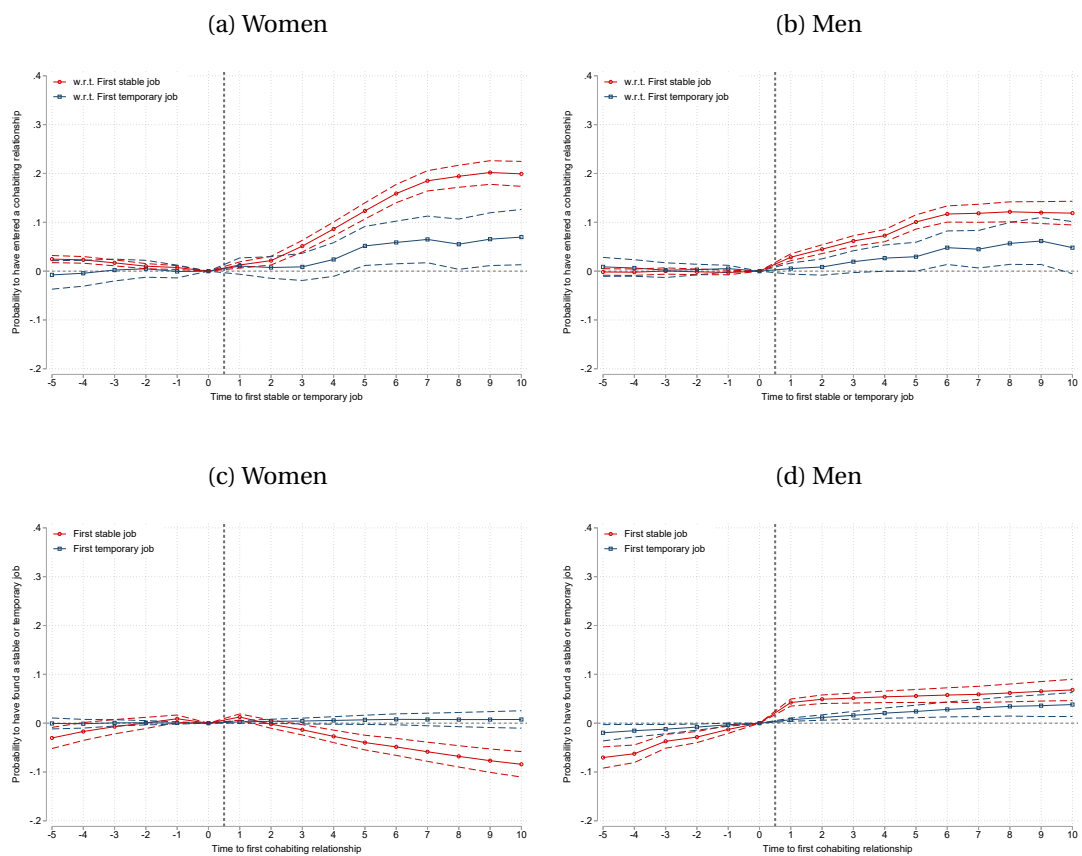


Figure 3.2 – First stable job, first temporary job and first cohabiting relationship – Older cohorts

Note: The figures refer to the samples of men and women from the *HB* survey born before 1954 who completed initial schooling before the survey and experienced the event used for the horizontal axis between school leaving age and 1999 or 2000. Figures 3.2a and 3.2b show the event time coefficients estimated for Equations (3.1), and Figures 3.2c and 3.2d show the event time coefficients estimated for Equations (3.2). Dotted lines represent 95% confidence intervals computed with robust standard errors clustered at the individual level.

Table 2.A1 – Timing of events analysis – Model (3.3)

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	1.213*** (0.075)	1.327*** (0.084)
First temporary job	0.133*** (0.047)	0.027 (0.045)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	0.409*** (0.093)	0.320*** (0.089)
First temporary job	0.388*** (0.065)	0.191*** (0.064)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	0.086 (0.106)	0.076 (0.116)
First stable job	-0.773*** (0.086)	-0.769*** (0.101)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	3551	3495
Sample log-likelihood	-21241.8	-19842.8

Note: The table refers to the samples of men and women from the FE survey who finished initial schooling before the survey, did not start their first stable job, their first temporary job or their first cohabiting relationship during the same year, and experienced each event either between age 18 and the minimum of year at age 35 and 2003-2004 or never experienced it. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 2.A2 – Timing of events analysis – Model (3.4)

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	1.193*** (0.081)	1.171*** (0.077)
First temporary job	0.131*** (0.050)	0.151** (0.061)
First child	0.273*** (0.102)	2.647*** (0.215)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first child</i>		
First stable job	1.191*** (0.156)	0.703*** (0.139)
First temporary job	-0.152*** (0.058)	-0.105 (0.078)
First cohabiting relationship	2.578*** (0.098)	3.286*** (0.118)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	0.392*** (0.094)	0.181** (0.089)
First child	-0.666*** (0.127)	-0.201 (0.182)
First temporary job	0.398*** (0.075)	0.145** (0.065)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	0.012 (0.120)	0.083 (0.128)
First child	-0.201 (0.124)	-0.079 (0.152)
First stable job	-0.822*** (0.105)	-0.786*** (0.101)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	3018	3148
Sample log-likelihood	-23334.7	-22739.6

Note: The table refers to the samples of men and women from the FE survey who finished initial schooling before the survey, did not start their first stable job, their first temporary job, their first cohabiting relationship or conceive their first child during the same year, and experienced each event either between age 18 and the minimum of year at age 35 and 2003-2004 or never experienced it. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship or a first fertility event and the impact of a first cohabiting relationship or a first fertility event on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 2.A3 – Timing of events analysis – Model (3.5) – Older cohorts

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	0.564*** (0.049)	0.976*** (0.071)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	-1.198*** (0.119)	0.748*** (0.141)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	4401	3870
Sample log-likelihood	-17952.6	-16830.2

Note: The table refers to the samples of men and women from the *HB* survey born before 1954 who finished initial schooling before the survey, did not start their first stable job and their first cohabiting relationship during the same year, and experienced each event either between schooling age and the minimum of 20 years later and 1999-2000 or never experienced it. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Appendix

Table 3.A1 – Descriptive statistics

	Women (1)	Men (2)
Stable employment	0.92 [0.28]	0.96 [0.21]
Med age at first stable job	21 [3.62]	20 [2.99]
Temporary employment	0.50 [0.50]	0.49 [0.50]
Med. age at first temp. job	20 [5.31]	20 [4.72]
Cohabiting relationship	0.90 [0.30]	0.83 [0.38]
Med. age at first cohabiting relationship	22 [3.78]	24 [3.97]
Children	0.77 [0.42]	0.66 [0.48]
Med. age at first child	24 [4.42]	27 [4.51]
Residential independence	0.96 [0.21]	0.90 [0.30]
Med. age at residential independence	21 [3.34]	23 [3.82]
Med. school leaving age	19 [3.15]	19 [3.32]
Higher education	0.33 [0.47]	0.28 [0.45]
Secondary education	0.18 [0.39]	0.16 [0.36]
Strong religious beliefs	0.30 [0.46]	0.21 [0.41]
N	4707	4156

Note: The table refers to the samples of men and women from the *FE* survey who finished initial schooling before the survey. The first column refers to the subsample of women and the second column to the subsample of men. Each row corresponds to a specific variable. Rows 1 and 3 show the employment rate over the life-course for stable jobs and temporary jobs respectively. Row 2 shows the median age at first stable job, and row 4 shows the same figure for temporary jobs. Row 5 corresponds to the proportion of individuals who lived in a cohabiting relationship before the survey, and row 8 to the proportion of individuals who had a child. Rows 6 and 7 show median age at first cohabiting relationship and at first child respectively. Row 9 indicates the proportion of individuals who no longer lived with their parents at the time of the survey, and row 10 indicates the median age at residential independence. Row 11 gives the median school leaving age. Row 12 shows the proportion of men and women who graduated from higher education and row 13 from high school. Row 14 shows the proportion of individuals who indicated strong religious beliefs. For each variable and each sample, the table reports the mean or the median of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 3.A2 – Descriptive statistics – Timing of events analysis – Model (3.3)

	Women (1)	Men (2)
Stable employment	0.89 [0.31]	0.95 [0.22]
Med age at first stable job	21 [3.25]	20 [2.96]
Temporary employment	0.43 [0.49]	0.42 [0.49]
Med. age at first temp. job	20 [4.26]	20 [3.75]
Cohabiting relationship	0.86 [0.35]	0.79 [0.41]
Med. age at first cohabiting relationship	22 [3.43]	24 [3.44]
Children	0.72 [0.45]	0.60 [0.49]
Med. age at first child	25 [3.95]	27 [3.62]
Residential independence	0.94 [0.23]	0.89 [0.31]
Med. age at residential independence	21 [3.44]	23 [3.92]
Med. school leaving age	19 [3.19]	19 [3.37]
Higher education	0.33 [0.47]	0.26 [0.44]
Secondary education	0.18 [0.38]	0.15 [0.36]
Strong religious beliefs	0.31 [0.46]	0.22 [0.41]
N	3551	3495

Note: The table refers to the same samples as Table 2.A1. Each row corresponds to a specific variable. Rows 1 and 3 show the employment rate over the life-course for stable jobs and temporary jobs respectively. Row 2 shows the median age at first stable job, and row 4 shows the same figure for temporary jobs. Row 5 corresponds to the proportion of individuals who lived in a cohabiting relationship before the survey, and row 8 to the proportion of individuals who had a child. Rows 6 and 7 show median age at first cohabiting relationship and at first child respectively. Row 9 indicates the proportion of individuals who no longer lived with their parents at the time of the survey, and row 10 indicates the median age at residential independence. Row 11 gives the median school leaving age. Row 12 shows the proportion of men and women who graduated from higher education and row 13 from high school. Row 14 shows the proportion of individuals who indicated strong religious beliefs. For each variable and each sample, the table reports the mean or the median of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 3.A3 – Descriptive statistics – Timing of events analysis – Model (3.4)

	Women (1)	Men (2)
Stable employment	0.90 [0.30]	0.94 [0.23]
Med age at first stable job	21 [3.17]	20 [2.94]
Temporary employment	0.43 [0.49]	0.43 [0.50]
Med. age at first temp. job	20 [4.10]	20 [3.66]
Cohabiting relationship	0.84 [0.37]	0.76 [0.43]
Med. age at first cohabiting relationship	22 [3.39]	24 [3.42]
Children	0.67 [0.47]	0.55 [0.50]
Med. age at first child	25 [3.66]	27 [3.34]
Residential independence	0.94 [0.24]	0.88 [0.32]
Med. age at residential independence	22 [3.46]	23 [3.97]
Med. school leaving age	20 [3.17]	19 [3.34]
Higher education	0.36 [0.48]	0.27 [0.45]
Secondary education	0.19 [0.39]	0.16 [0.36]
Strong religious beliefs	0.29 [0.46]	0.21 [0.41]
N	3018	3148

Note: The table refers to the same samples as Table 2.A2. Each row corresponds to a specific variable. Rows 1 and 3 show the employment rate over the life-course for stable jobs and temporary jobs respectively. Row 2 shows the median age at first stable job, and row 4 shows the same figure for temporary jobs. Row 5 corresponds to the proportion of individuals who lived in a cohabiting relationship before the survey, and row 8 to the proportion of individuals who had a child. Rows 6 and 7 show median age at first cohabiting relationship and at first child respectively. Row 9 indicates the proportion of individuals who no longer lived with their parents at the time of the survey, and row 10 indicates the median age at residential independence. Row 11 gives the median school leaving age. Row 12 shows the proportion of men and women who graduated from higher education and row 13 from high school. Row 14 shows the proportion of individuals who indicated strong religious beliefs. For each variable and each sample, the table reports the mean or the median of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 3.A4 – Descriptive statistics – Timing of events analysis – Individuals with simultaneous transitions – Model (3.3)

	Women (1)	Men (2)
Stable employment	0.98 [0.13]	0.99 [0.09]
Med age at first stable job	21 [2.58]	21 [2.69]
Temporary employment	0.66 [0.47]	0.73 [0.44]
Med. age at first temp. job	20 [3.31]	20 [3.12]
Cohabiting relationship	0.98 [0.15]	0.94 [0.23]
Med. age at first cohabiting relationship	21 [2.50]	22 [2.92]
Children	0.80 [0.40]	0.69 [0.46]
Med. age at first child	24 [3.68]	25 [3.68]
Residential independence	0.99 [0.10]	0.96 [0.20]
Med. age at residential independence	20 [2.44]	22 [2.95]
Med. school leaving age	20 [2.81]	20 [3.00]
Higher education	0.36 [0.48]	0.36 [0.48]
Secondary education	0.22 [0.41]	0.17 [0.38]
Strong religious beliefs	0.27 [0.44]	0.19 [0.39]
N	997	645

Note: The table refers to similar samples as Table 2.A1 restricted to individuals who experienced at least two events during the same year. Each row corresponds to a specific variable. Rows 1 and 3 show the employment rate over the life-course for stable jobs and temporary jobs respectively. Row 2 shows the median age at first stable job, and row 4 shows the same figure for temporary jobs. Row 5 corresponds to the proportion of individuals who lived in a cohabiting relationship before the survey, and row 8 to the proportion of individuals who had a child. Rows 6 and 7 show median age at first cohabiting relationship and at first child respectively. Row 9 indicates the proportion of individuals who no longer lived with their parents at the time of the survey, and row 10 indicates the median age at residential independence. Row 11 gives the median school leaving age. Row 12 shows the proportion of men and women who graduated from higher education and row 13 from high school. Row 14 shows the proportion of individuals who indicated strong religious beliefs. For each variable and each sample, the table reports the mean or the median of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 3.A5 – Descriptive statistics – Timing of events analysis – Individuals with simultaneous transitions – Model (3.4)

	Women (1)	Men (2)
Stable employment	0.94 [0.24]	0.99 [0.09]
Med age at first stable job	21 [2.94]	21 [2.86]
Temporary employment	0.59 [0.49]	0.61 [0.49]
Med. age at first temp. job	20 [3.79]	20 [3.51]
Cohabiting relationship	0.98 [0.14]	0.96 [0.19]
Med. age at first cohabiting relationship	21 [2.93]	22 [3.29]
Children	0.86 [0.35]	0.79 [0.40]
Med. age at first child	23 [3.72]	25 [3.65]
Residential independence	0.99 [0.10]	0.97 [0.17]
Med. age at residential independence	21 [2.74]	22 [3.16]
Med. school leaving age	19 [2.96]	19 [3.26]
Higher education	0.30 [0.46]	0.30 [0.46]
Secondary education	0.19 [0.39]	0.16 [0.36]
Strong religious beliefs	0.30 [0.46]	0.22 [0.42]
N	1496	980

Note: The table refers to similar samples as Table 2.A2 restricted to individuals who experienced at least two events during the same year. Each row corresponds to a specific variable. Rows 1 and 3 show the employment rate over the life-course for stable jobs and temporary jobs respectively. Row 2 shows the median age at first stable job, and row 4 shows the same figure for temporary jobs. Row 5 corresponds to the proportion of individuals who lived in a cohabiting relationship before the survey, and row 8 to the proportion of individuals who had a child. Rows 6 and 7 show median age at first cohabiting relationship and at first child respectively. Row 9 indicates the proportion of individuals who no longer lived with their parents at the time of the survey, and row 10 indicates the median age at residential independence. Row 11 gives the median school leaving age. Row 12 shows the proportion of men and women who graduated from higher education and row 13 from high school. Row 14 shows the proportion of individuals who indicated strong religious beliefs. For each variable and each sample, the table reports the mean or the median of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 3.A6 – Timing of events analysis – Model (3.3) – Detailed results

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	1.213*** (0.075)	1.327*** (0.084)
First temporary job	0.133*** (0.047)	0.027 (0.045)
Schooling	-0.213*** (0.065)	-0.054 (0.080)
Non residential independence	-0.002 (0.047)	-0.242*** (0.044)
Higher education	0.102* (0.052)	0.156*** (0.053)
Secondary education	0.022 (0.055)	-0.053 (0.060)
Strong religious beliefs	-0.049 (0.042)	-0.166*** (0.048)
Unemployment rate	-0.005 (0.010)	0.008 (0.011)
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	0.409*** (0.093)	0.320*** (0.089)
First temporary job	0.388*** (0.065)	0.191*** (0.064)
Schooling	-1.575*** (0.054)	-1.307*** (0.045)
Non residential independence	0.072 (0.078)	-0.140** (0.068)
Higher education	-1.008*** (0.083)	-1.197*** (0.071)
Secondary education	-0.197** (0.082)	-0.556*** (0.070)
Strong religious beliefs	-0.096* (0.053)	-0.168*** (0.052)
Unemployment rate	-0.127*** (0.011)	-0.089*** (0.009)
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	0.086 (0.106)	0.076 (0.116)
First stable job	-0.773*** (0.086)	-0.769*** (0.101)
Schooling	-0.146* (0.086)	-0.276*** (0.079)
Non residential independence	0.186* (0.104)	0.102 (0.103)
Higher education	0.393*** (0.076)	0.180** (0.078)
Secondary education	0.361*** (0.082)	0.245*** (0.081)
Strong religious beliefs	-0.086 (0.061)	-0.130* (0.068)
Unemployment rate	0.107*** (0.014)	0.110*** (0.013)
N	3551	3495
Sample log-likelihood	-21241.8	-19842.8
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2

Note: The table refers to the same samples as Table 2.A1. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A7 – Unobserved heterogeneity distribution – Model (3.3)

	Women	Men
p_{111}	0.793	0.013
p_{112}	0.000	0.049
p_{121}	0.037	0.910
p_{122}	0.030	0.000
p_{211}	0.000	0.002
p_{212}	0.000	0.026
p_{221}	0.128	0.000
p_{222}	0.012	0.000
$v_{i,r}^1$	-3.001*** (0.118)	-3.918*** (0.141)
$v_{i,r}^2$	-1.387*** (0.135)	-1.871*** (0.294)
$v_{i,js}^1$	1.280*** (0.124)	-1.801*** (0.160)
$v_{i,js}^2$	-2.494*** (0.154)	1.134*** (0.101)
$v_{i,tj}^1$	-3.160*** (0.178)	-2.937*** (0.167)
$v_{i,tj}^2$	-0.738*** (0.286)	-1.429*** (0.232)
$corr(N_r, N_{sj})$	-0.790	-0.534
$corr(N_r, N_{tj})$	0.092	0.552
$corr(N_{sj}, N_{tj})$	-0.411	-0.904

Note: The table refers to the same samples as Table 2.A1. The two columns show the estimated parameters for the distribution of unobserved characteristics impacting transitions into stable employment, temporary employment and cohabiting relationships. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A8 – Timing of events analysis – Model (3.3) without unobserved heterogeneity

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	0.674*** (0.052)	1.128*** (0.069)
First temporary job	0.119*** (0.043)	0.072* (0.043)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	No	No
Number of mass points	1	1
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	-0.418*** (0.067)	0.031 (0.071)
First temporary job	0.259*** (0.043)	-0.283*** (0.046)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	No	No
Number of mass points	1	1
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	0.063 (0.105)	0.131 (0.113)
First stable job	-0.834*** (0.082)	-1.053*** (0.088)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	No	No
Number of mass points	1	1
N	3551	3495
Sample log-likelihood	-21900.2	-20217.6

Note: The table refers to the same samples as Table 2.A1. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed characteristics, yearly unemployment rates and duration variables but no unobserved heterogeneity. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A9 – Timing of events analysis – Model (3.3) – Robustness checks (1)

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	1.222*** (0.075)	1.325*** (0.084)
First temporary job	0.133*** (0.047)	0.030 (0.045)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	0.414*** (0.094)	0.302*** (0.090)
First temporary job	0.385*** (0.064)	0.171*** (0.063)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	0.091 (0.107)	0.083 (0.116)
First stable job	-0.772*** (0.089)	-0.772*** (0.101)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	3551	3495
Sample log-likelihood	-21241.8	-19842.8

Note: The table refers to the same samples as Table 2.A1. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates by age groups and gender, and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A10 – Timing of events analysis – Model (3.3) – Robustness checks (2)

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	0.549*** (0.073)	1.424*** (0.103)
First temporary job	-0.114* (0.062)	-0.126** (0.057)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	-0.537*** (0.123)	-0.220** (0.104)
First temporary job	0.137 (0.116)	-0.282*** (0.078)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	-0.184 (0.149)	-0.025 (0.150)
First stable job	-0.708*** (0.142)	-0.734*** (0.134)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	2409	2705
Sample log-likelihood	-13773.2	-14453.0

Note: The table refers to the same samples as Table 2.A1 restricted to individuals who did not experience the different events during two consecutive years. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A11 – Timing of events analysis – Model (3.4) – Detailed results

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	1.193*** (0.081)	1.171*** (0.077)
First temporary job	0.131*** (0.050)	0.151** (0.061)
First child	0.273*** (0.102)	2.647*** (0.215)
Schooling	-0.192*** (0.071)	-0.092 (0.085)
Non residential independence	-0.026 (0.051)	-0.252*** (0.052)
Higher education	0.162*** (0.057)	0.133** (0.064)
Secondary education	0.082 (0.059)	-0.100 (0.071)
Strong religious beliefs	-0.023 (0.046)	-0.249*** (0.059)
Unemployment rate	0.009 (0.011)	0.015 (0.013)
<i>Hazard rate of a first child</i>		
First stable job	1.191*** (0.156)	0.703*** (0.139)
First temporary job	-0.152*** (0.058)	-0.105 (0.078)
First cohabiting relationship	2.578*** (0.098)	3.286*** (0.118)
Schooling	-0.663*** (0.131)	-0.438*** (0.160)
Non residential independence	-0.459*** (0.092)	-0.379*** (0.107)
Higher education	-0.221*** (0.058)	-0.362*** (0.074)
Secondary education	-0.105 (0.065)	-0.269*** (0.086)
Strong religious beliefs	0.201*** (0.051)	0.253*** (0.071)
Unemployment rate	-0.056*** (0.015)	-0.044** (0.018)
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	0.392*** (0.094)	0.181** (0.089)
First child	-0.666*** (0.127)	-0.201 (0.182)
First temporary job	0.398*** (0.075)	0.145** (0.065)
Schooling	-1.607*** (0.058)	-1.313*** (0.047)
Non residential independence	-0.026 (0.083)	-0.131* (0.070)
Higher education	-1.143*** (0.088)	-1.222*** (0.075)
Secondary education	-0.343*** (0.085)	-0.554*** (0.072)
Strong religious beliefs	-0.115** (0.058)	-0.144*** (0.054)
Unemployment rate	-0.138*** (0.012)	-0.092*** (0.009)
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	0.012 (0.120)	0.083 (0.128)
First child	-0.201 (0.124)	-0.079 (0.152)
First stable job	-0.822*** (0.105)	-0.786*** (0.101)
Schooling	-0.134 (0.095)	-0.238*** (0.082)
Non residential independence	0.180 (0.113)	0.104 (0.108)
Higher education	0.404*** (0.084)	0.165** (0.082)
Secondary education	0.373*** (0.090)	0.260*** (0.085)
Strong religious beliefs	-0.067 (0.067)	-0.148** (0.072)
Unemployment rate	0.096*** (0.015)	0.105*** (0.014)
N	3018	3148
Sample log-likelihood	-23334.7	-22739.6
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2

Note: The table refers to the same samples as Table 2.A2. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship or a first fertility event and the impact of a first cohabiting relationship or a first fertility event on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A12 – Unobserved heterogeneity distribution – Model (3.4)

	Women	Men
p_{1111}	0.000	0.007
p_{1112}	0.812	0.002
p_{1121}	0.038	0.009
p_{1122}	0.031	0.107
p_{1211}	0.000	0.069
p_{1212}	0.000	0.000
p_{1221}	0.011	0.000
p_{1222}	0.015	0.723
p_{2111}	0.000	0.000
p_{2112}	0.000	0.000
p_{2121}	0.007	0.000
p_{2122}	0.011	0.000
p_{2211}	0.000	0.003
p_{2212}	0.000	0.008
p_{2221}	0.000	0.013
p_{2222}	0.075	0.059
$v_{i,r}^1$	-3.180*** (0.133)	-3.759*** (0.149)
$v_{i,r}^2$	-1.668*** (0.191)	-6.918*** (0.364)
$v_{i,co}^1$	-4.595*** (0.224)	-7.765*** (1.083)
$v_{i,co}^2$	-2.350*** (0.199)	-5.084*** (0.242)
$v_{i,sj}^1$	1.573*** (0.141)	-1.798*** (0.176)
$v_{i,sj}^2$	-1.947*** (0.182)	1.143*** (0.106)
$v_{i,tj}^1$	-1.090*** (0.292)	-1.536*** (0.203)
$v_{i,tj}^2$	-3.095*** (0.195)	-2.933*** (0.181)
$corr(N_r, N_{sj})$	-0.666	0.041
$corr(N_r, N_{tj})$	0.030	-0.089
$corr(N_r, N_{co})$	0.749	-0.114
$corr(N_{co}, N_{sj})$	-0.698	-0.024
$corr(N_{co}, N_{tj})$	0.081	-0.036
$corr(N_{sj}, N_{tj})$	-0.508	-0.813

Note: The table refers to the same samples as Table 2.A2. The two columns show the estimated parameters for the distribution of unobserved characteristics impacting transitions into stable employment, temporary employment, cohabiting relationships and fertility. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A13 – Timing of events analysis – Model (3.4) – Robustness checks

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	0.910*** (0.110)	1.596*** (0.125)
First temporary job	-0.106 (0.078)	-0.142** (0.069)
First child	-0.692*** (0.197)	0.142 (0.254)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first child</i>		
First stable job	0.664*** (0.206)	0.692*** (0.257)
First temporary job	-0.284*** (0.097)	-0.125 (0.085)
First cohabiting relationship	2.314*** (0.137)	3.589*** (0.204)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	-0.305** (0.139)	-0.093 (0.130)
First child	-0.707*** (0.189)	-0.283 (0.327)
First temporary job	0.083 (0.156)	-0.332*** (0.084)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	-0.462** (0.198)	-0.410** (0.184)
First child	0.265 (0.215)	-0.021 (0.236)
First stable job	-0.399** (0.198)	-0.830*** (0.148)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	1561	2044
Sample log-likelihood	-11167.0	-13424.0

Note: The table refers to the same samples as Table 2.A2 restricted to individuals who did not experience the different events during two consecutive years. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship or a first fertility event and the impact of a first cohabiting relationship or a first fertility event on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A14 – Descriptive statistics – *HB* Survey

	Older cohorts		Younger cohorts	
	Women (1)	Men (2)	Women (3)	Men (4)
Stable employment	0.84 [0.36]	0.92 [0.28]	0.90 [0.31]	0.95 [0.22]
Med. age at first stable job	17 [6.99]	18 [5.35]	20 [3.99]	20 [3.78]
Temporary employment	0.10 [0.30]	0.17 [0.37]	0.24 [0.43]	0.25 [0.43]
Med. age at first temp. job	17 [11.71]	15 [11.00]	20 [5.69]	19 [5.48]
Cohabiting relationship	0.95 [0.22]	0.96 [0.20]	0.92 [0.27]	0.91 [0.29]
Med. age at first cohabiting relationship	22 [5.23]	24 [5.50]	22 [4.02]	24 [3.87]
Children	0.88 [0.32]	0.88 [0.32]	0.77 [0.42]	0.74 [0.44]
Med. age at first child	24 [4.33]	26 [4.92]	25 [4.29]	27 [4.28]
Med. school leaving age	16 [3.14]	16 [3.72]	18 [2.99]	18 [3.18]
Higher education	0.12 [0.33]	0.16 [0.37]	0.29 [0.45]	0.27 [0.44]
Secondary education	0.10 [0.30]	0.10 [0.30]	0.16 [0.36]	0.13 [0.34]
High SES	0.29 [0.45]	0.29 [0.45]	0.34 [0.47]	0.35 [0.48]
N	4671	4207	4109	3428

Note: The table refers to the samples of men and women from the *HB* survey who finished initial schooling before the survey. The first column refers to the subsample of women born before 1954, the second column to the subsample of women born in or after 1954, the third column to the subsample of men born before 1954 and the fourth column to the subsample of men born in or after 1954. Each row corresponds to a specific variable. Rows 1 and 3 show the employment rate over the life-course for stable jobs and temporary jobs respectively. Row 2 shows the median age at first stable job, and row 4 shows the same figure for temporary jobs. Row 5 corresponds to the proportion of individuals who lived in a cohabiting relationship before the survey, and row 8 to the proportion of individuals who had a child. Rows 6 and 7 show median age at first cohabiting relationship and at first child respectively. Row 9 gives the median school leaving age. Row 10 indicates the proportion of men and women who graduated from higher education and row 11 from high school. Row 12 shows the proportion of individuals whose father held a high SES position. For each variable and each sample, the table reports the mean or the median of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 3.A15 – Descriptive statistics – *HB* Survey – Timing of events analysis – Model (3.5)

	Older cohorts		Younger cohorts	
	Women (1)	Men (2)	Women (3)	Men (4)
Stable employment	0.79 [0.41]	0.90 [0.30]	0.88 [0.33]	0.94 [0.23]
Med. age at first stable job	17 [3.90]	17 [3.91]	19 [3.71]	19 [3.47]
Temporary employment	0.08 [0.28]	0.15 [0.36]	0.25 [0.43]	0.25 [0.43]
Med. age at first temp. job	16 [5.78]	14 [4.36]	19 [4.74]	19 [4.80]
Cohabiting relationship	0.91 [0.29]	0.90 [0.31]	0.89 [0.31]	0.89 [0.32]
Med. age at first cohabiting relationship	22 [3.67]	24 [3.56]	22 [3.93]	24 [3.66]
Children	0.85 [0.35]	0.81 [0.39]	0.76 [0.43]	0.72 [0.45]
Med. age at first child	24 [3.87]	26 [3.79]	24 [4.26]	27 [4.00]
Med. school leaving age	16 [2.84]	16 [3.25]	18 [2.74]	18 [2.79]
Higher education	0.10 [0.30]	0.12 [0.33]	0.23 [0.42]	0.22 [0.41]
Secondary education	0.10 [0.30]	0.10 [0.31]	0.15 [0.36]	0.14 [0.34]
High SES	0.28 [0.45]	0.27 [0.44]	0.31 [0.46]	0.32 [0.47]
N	4401	3870	3360	2999

Note: The table refers to the same samples as Tables 2.A3 and 3.A18. Each row corresponds to a specific variable. Rows 1 and 3 show the employment rate over the life-course for stable jobs and temporary jobs respectively. Row 2 shows the median age at first stable job, and row 4 shows the same figure for temporary jobs. Row 5 corresponds to the proportion of individuals who lived in a cohabiting relationship before the survey, and row 8 to the proportion of individuals who had a child. Rows 6 and 7 show median age at first cohabiting relationship and at first child respectively. Row 9 gives the median school leaving age. Row 10 indicates the proportion of men and women who graduated from higher education and row 11 from high school. Row 12 shows the proportion of individuals whose father held a high SES position. For each variable and each sample, the table reports the mean or the median of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 3.A16 – Descriptive statistics – *HB* Survey – Timing of events analysis – Individuals with simultaneous transitions – Model (3.5)

	Older cohorts		Younger cohorts	
	Women (1)	Men (2)	Women (3)	Men (4)
Stable employment	1.00 [0.00]	1.00 [0.00]	1.00 [0.00]	1.00 [0.00]
Med. age at first stable job	21 [2.68]	23 [2.67]	21 [2.40]	23 [2.64]
Temporary employment	0.11 [0.32]	0.11 [0.32]	0.25 [0.43]	0.26 [0.44]
Med. age at first temp. job	18 [3.90]	16 [4.23]	20 [3.97]	19 [3.68]
Cohabiting relationship	1.00 [0.00]	1.00 [0.00]	1.00 [0.00]	1.00 [0.00]
Med. age at first cohabiting relationship	21 [2.68]	23 [2.67]	21 [2.40]	23 [2.64]
Children	0.96 [0.19]	0.98 [0.14]	0.85 [0.36]	0.76 [0.43]
Med. age at first child	24 [3.46]	25 [3.55]	24 [3.70]	26 [3.85]
Med. school leaving age	18 [3.79]	20 [4.57]	20 [2.57]	21 [3.14]
Higher education	0.36 [0.48]	0.45 [0.50]	0.39 [0.49]	0.50 [0.50]
Secondary education	0.14 [0.35]	0.09 [0.29]	0.25 [0.43]	0.14 [0.35]
High SES	0.37 [0.48]	0.43 [0.50]	0.41 [0.49]	0.46 [0.50]
N	159	202	379	243

Note: The table refers to similar samples as Tables 2.A3 and 3.A18 restricted to individuals who experienced the two events during the same year. Each row corresponds to a specific variable. Rows 1 and 3 show the employment rate over the life-course for stable jobs and temporary jobs respectively. Row 2 shows the median age at first stable job, and row 4 shows the same figure for temporary jobs. Row 5 corresponds to the proportion of individuals who lived in a cohabiting relationship before the survey, and row 8 to the proportion of individuals who had a child. Rows 6 and 7 show median age at first cohabiting relationship and at first child respectively. Row 9 gives the median school leaving age. Row 10 indicates the proportion of men and women who graduated from higher education and row 11 from high school. Row 12 shows the proportion of individuals whose father held a high SES position. For each variable and each sample, the table reports the mean or the median of the left hand side variable among the corresponding sample. Standard deviations are in brackets.

Table 3.A17 – Timing of events analysis – Model (3.5) – Younger cohorts from *HB* survey

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	1.069*** (0.326)	1.183*** (0.110)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	0.633** (0.311)	0.226* (0.131)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	3360	2999
Sample log-likelihood	-12769.6	-12268.6

Note: The table refers to the samples of men and women from the *HB* survey born in or after 1954 who finished initial schooling before the survey, did not start their first stable job and their first cohabiting relationship during the same year, and experienced each event either between schooling age and the minimum of 20 years later and 1999-2000 or never experienced it. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A18 – Timing of events analysis – Model (3.3) – Younger cohorts from *HB* survey

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	1.399*** (0.129)	1.147*** (0.081)
First temporary job	0.009 (0.074)	0.136 (0.112)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	0.288** (0.128)	-0.204* (0.118)
First temporary job	1.560*** (0.172)	1.440*** (0.247)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first temporary job</i>		
First cohabiting relationship	-0.090 (0.171)	-0.338* (0.187)
First stable job	-0.506*** (0.179)	0.037 (0.348)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	3278	2959
Sample log-likelihood	-14764.9	-14245.0

Note: The table refers to the samples of men and women from the *HB* survey born in or after 1954 who finished initial schooling before the survey, did not start their first stable job, their first temporary job or their first cohabiting relationship during the same year, and experienced each event either between schooling age and the minimum of 20 years later and 1999-2000 or never experienced it. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable and temporary job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable and temporary job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A19 – Timing of events analysis – Model (3.5) – Older cohorts – Detailed results

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	0.564*** (0.049)	0.976*** (0.071)
Higher education	0.442*** (0.086)	1.340*** (0.095)
Secondary education	0.601*** (0.083)	0.835*** (0.081)
High SES	0.193*** (0.051)	0.147*** (0.054)
Unemployment rate	0.137*** (0.016)	0.152*** (0.018)
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	-1.198*** (0.119)	0.748*** (0.141)
Higher education	1.165*** (0.290)	0.033 (0.066)
Secondary education	0.818*** (0.151)	0.246*** (0.063)
High SES	0.037 (0.065)	0.037 (0.044)
Unemployment rate	0.213*** (0.025)	0.159*** (0.018)
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	4401	3870
Sample log-likelihood	-17952.6	-16830.2

Note: The table refers to the same samples as Table 2.A3. Each column corresponds to a specific regression where the impact of a first stable job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.

* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A20 – Unobserved heterogeneity distribution – Model (3.5) – Older cohorts

	Women	Men
p_{11}	0.112	0.134
p_{12}	0.034	0.012
p_{21}	0.592	0.757
p_{22}	0.261	0.097
$v_{i,r}^1$	-6.567*** (0.152)	-7.921*** (0.159)
$v_{i,r}^2$	-3.547*** (0.054)	-4.591*** (0.074)
$v_{i,sj}^1$	0.250*** (0.084)	-0.029 (0.028)
$v_{i,sj}^2$	-1.900*** (0.261)	-4.988*** (0.305)
$corr(N_r, N_{sj})$	-0.055	-0.034

Note: The table refers to the same samples as Table 2.A3. The two columns show the estimated parameters for the distribution of unobserved characteristics impacting transitions into stable employment and cohabiting relationships. Standard errors are in parentheses.
* significant at 10%. ** significant at 5%. *** significant at 1%.

Table 3.A21 – Timing of events analysis – Model (3.5) – Older cohorts – Robustness checks

	Women	Men
<i>Hazard rate of a first cohabiting relationship</i>		
First stable job	0.490*** (0.049)	0.980*** (0.075)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
<i>Hazard rate of a first stable job</i>		
First cohabiting relationship	-1.374*** (0.126)	-0.238 (0.182)
Control variables	✓	✓
Piecewise constant duration	✓	✓
Unobserved characteristics	✓	✓
Number of mass points	2	2
N	4193	3638
Sample log-likelihood	-17103.0	-15647.1

Note: The table refers to the same samples as Table 2.A3 restricted to individuals who did not experience the different events during two consecutive years. The first column refers to the subsample of women and the second column to the subsample of men. Each column corresponds to a specific regression where the impact of a first stable job on the hazard rate of a first cohabiting relationship and the impact of a first cohabiting relationship on the hazard rate of a first stable job are estimated jointly. All regressions include controls for individuals' observed and unobserved characteristics, yearly unemployment rates and duration variables. Standard errors are in parentheses.
* significant at 10%. ** significant at 5%. *** significant at 1%.

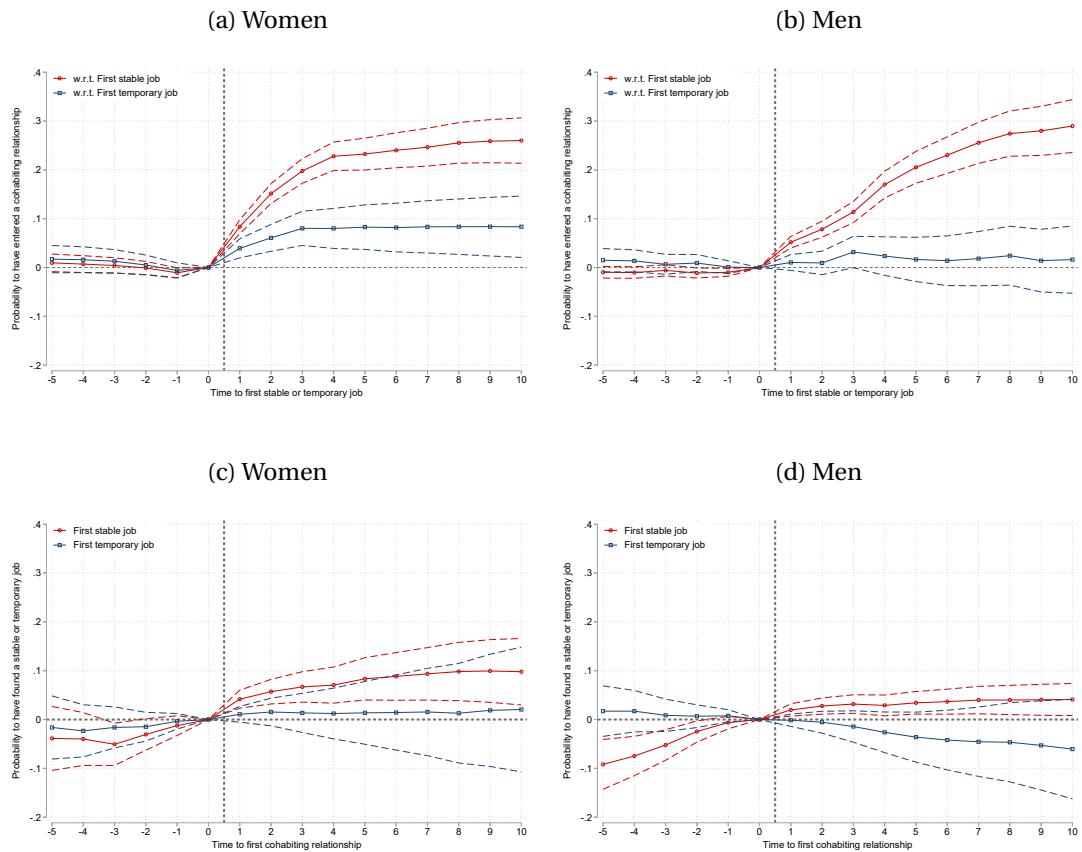


Figure 3.A1 – First stable job, first temporary job and first cohabiting relationship

Note: The figures refer to the samples of men and women from the *FE* survey who completed initial schooling before the survey. Additionally, Figures 3.A1a and 3.A1b focus on men and women who started their first stable or temporary job between age 18 and 1993 or 1994. Figures 3.A1c and 3.A1d focus on men and women who started their first cohabiting relationship between age 23 and 1993 or 1994. Figures 3.A1a and 3.A1b show the event time coefficients estimated for Equations (3.1), and Figures 3.A1c and 3.A1d show the event time coefficients estimated for Equations (3.2). Dotted lines represent 95% confidence intervals computed with robust standard errors clustered at the individual level.

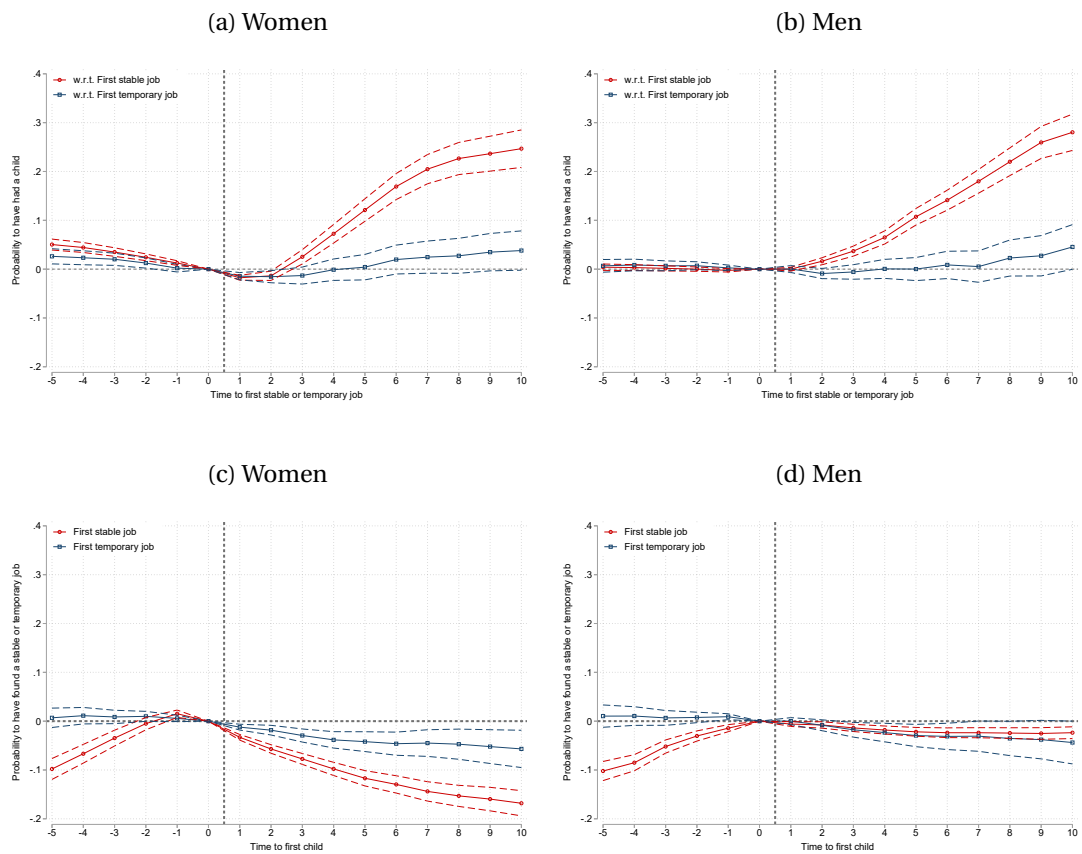


Figure 3.A2 – First stable job, first temporary job and first child

Note: The figures refer to the samples of men and women from the *FE* survey who completed initial schooling before survey and experienced the event used for the horizontal axis between age 18 and 2003 or 2004. Figures 3.A2a and 3.A2b show the event time coefficients estimated for Equations (3.1), and Figures 3.A2c and 3.A2d show the event time coefficients estimated for Equations (3.2) where the event "first cohabiting relationship" is replaced by "first child". Dotted lines represent 95% confidence intervals computed with robust standard errors clustered at the individual level.

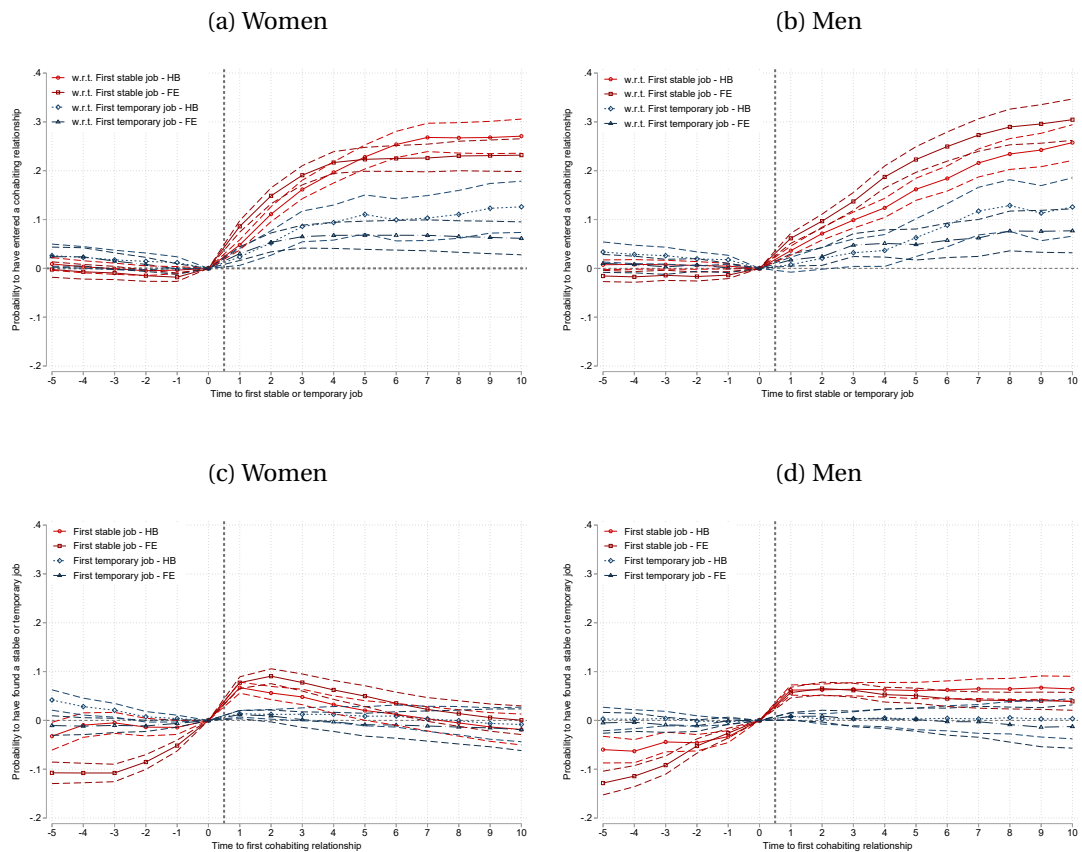


Figure 3.A3 – First stable job, first temporary job and first cohabiting relationship – *FE* vs. *HB* surveys

Note: The figures refer either to the samples of men and women from the *FE* survey who completed initial schooling before the survey and experienced the event used for the horizontal axis between age 18 and 2003 or 2004, or to the samples of men and women from the *HB* survey born in or after 1954 who completed initial schooling before the survey and experienced the event used for the horizontal axis between school leaving age and 1999 or 2000. Figures 3.A3a and 3.A3b show the event time coefficients estimated for Equations (3.1), and Figures 3.A3c and 3.A3d show the event time coefficients estimated for Equations (3.2). Dotted lines represent 95% confidence intervals computed with robust standard errors clustered at the individual level.

Conclusion générale

Les travaux présentés dans cette thèse ont eu pour objectif d'éclairer l'importance du contexte institutionnel et social pour rendre compte des trajectoires scolaires, professionnelles et familiales. Pour conclure, nous nous proposons de revenir sur les résultats de chaque chapitre, pour en souligner les implications mais aussi les limites et voies de recherches futures.

Le premier chapitre de cette thèse s'est intéressé au rôle de l'environnement scolaire dans les choix d'orientation des lycéens de l'académie de Paris. En comparant les trajectoires scolaires des derniers collégiens admis dans les lycées parisiens les plus sélectifs avec celles des premiers recalés, nous avons montré que l'accès à un établissement scolaire sélectif n'avait aucun impact sur les performances au baccalauréat, mais un impact très négatif sur la propension des jeunes lycéennes à s'orienter dans la filière scientifique.

Nos résultats soulignent ainsi qu'un environnement scolaire plus ou moins sélectif et compétitif peut jouer sur la sous-représentation des femmes en sciences. Les processus d'affectation des élèves pourraient donc être envisagés comme des leviers pour lutter contre le manque de femmes dans ce domaine. Néanmoins, de nouvelles recherches restent à développer pour comprendre toutes les conséquences induites par différents processus d'affectation.

Dans cette perspective, de nouvelles recherches devront également être conduites pour mieux comprendre les raisons pour lesquelles un environnement plus sélectif affecte les choix des filles et non ceux des garçons. Dans le contexte parisien, l'accès à un lycée sélectif s'accompagne surtout d'une amélioration du niveau académique des camarades de classe, notamment en sciences. Nos résultats concordent donc avec les travaux d'économie et psychologie expérimentale suggérant que les filles sont plus réticentes que les garçons à choisir des voies compétitives. Cependant, les données administratives dont nous disposons ne permettent pas de fournir des preuves directes pour asseoir cette explication.

Enfin, les données utilisées ne nous ont pas non plus permis d'étudier les conséquences de nos résultats pour les choix d'orientation dans le supérieur et les trajectoires professionnelles. Nous avons souligné dans ce chapitre le caractère clé des choix d'orientation au lycée pour les trajectoires ultérieures, les filières scientifiques de l'enseignement supérieur étant largement inaccessibles aux bacheliers des autres filières. Là encore, de nouvelles recherches sont nécessaires pour éclairer précisément les implications à long termes de notre étude.

Le second chapitre de cette thèse a montré que les politiques de redoublement en classes

préparatoires scientifiques ont des effets directs très largement positifs pour les redoublants, mais des effets négatifs pour les nouveaux arrivants. Plus une classe compte de redoublants de haut niveau académique plus les nouveaux arrivants sont en difficulté au moment des concours, alors que leurs performances sont largement insensibles au nombre de redoublants de niveau moyen ou faible. Si les redoublements ont une influence sur les nouveaux arrivants, cela ne semble pas parce qu'ils contribuent à surcharger les classes, mais plutôt parce qu'ils en modifient le profil.

Les résultats présentés dans ce chapitre vont à l'encontre de l'idée généralement admise selon laquelle les bons élèves généreraient des effets d'entraînement vertueux bénéfiques au moins bons élèves. Un mécanisme susceptible d'expliquer nos résultats serait la sensibilité des pratiques enseignantes au profil des élèves : un surcroît de bons élèves peut induire les enseignants à produire des cours d'un niveau trop ambitieux pour la moyenne des nouveaux arrivants. Cette explication permettrait de rendre compte à la fois des effets très positifs du redoublement pour les redoublants, et des effets négatifs pour les nouveaux arrivants. Il serait cependant nécessaire de conduire des recherches plus approfondies pour valider cette hypothèse.

En termes de politiques publiques, les résultats présentés dans ce chapitre dressent un bilan mitigé des politiques de redoublement en classes préparatoires scientifiques. Les effets positifs du redoublement pour les redoublants semble in fine contrebalancés par les effets négatifs pour les nouveaux arrivants.

Plus généralement, le paradoxe mis en lumière concernant les potentiels effets vertueux des bons élèves pose la question des classes de niveaux. Comme l'ont entre autres souligné Duflo, Dupas et Kremer (2011) et Booi, Leuven et Oosterbeek (2017) dans le contexte kényan et hollandais, les classes de niveaux auraient des effets positifs pour l'ensemble des élèves, quand bien même elles priveraient les moins bons élèves de leurs meilleurs camarades. Il pourrait s'avérer intéressant d'étudier cette question dans le contexte des classes préparatoires scientifiques où cette pratique est déjà répandue, et de tâcher de mieux comprendre les mécanismes à l'œuvre (pratiques enseignantes, interactions, ...).

Le dernier chapitre de cette thèse s'est intéressé au rôle de l'institutionnalisation des emplois précaires et des difficultés rencontrées par les jeunes pour s'insérer durablement dans l'emploi pour expliquer les retards observés dans les trajectoires familiales des jeunes adultes.

Depuis le milieu des années soixante-dix, l'âge à la première cohabitation conjugale et au premier enfant ont nettement reculé en France. Ces évolutions sont traditionnellement mises en regard avec les changements de normes survenus à cette même période et avec l'allongement des études. Dans ce chapitre, nous montrons que les évolutions opérées sur le

marché du travail jouent également un rôle important pour rendre compte des trajectoires familiales. Nos résultats révèlent que l'accès à l'emploi – et surtout à l'emploi stable – constitue une étape importante pour la mise en couple et l'arrivée d'un premier enfant. Entre les générations nées au milieu des années cinquante et les générations nées au début des années soixante-dix, la montée du chômage des jeunes et de la part des emplois précaires expliquerait environ 25% des retards observés dans l'âge à la première mise en couple, et 40% des retards dans l'âge au premier enfant.

En termes de politiques publiques, une question importante reste de déterminer si ces retards dans les trajectoires familiales ont entraîné ou non une baisse totale de la fertilité. Nos données ne nous ont pas permis d'éclairer cette question, et de futures recherches pourraient être conduites sur ce sujet. S'il s'avérait que les reculs dans l'âge au premier enfant n'avaient pas de conséquences sur la fertilité totale, une autre perspective serait d'étudier les conséquences de ces reculs pour les enfants et leurs parents. En particulier, les retards dans les trajectoires familiales sont susceptibles de diminuer l'espacement des naissances et d'affecter ainsi les trajectoires professionnelles des mères et les trajectoires scolaires des enfants.

Enfin, de futures recherches devront être conduites pour éclairer la question des mécanismes derrière les liens entre accès à l'emploi stable et formation de la famille. En particulier, le marché de l'immobilier semble un candidat potentiel pour expliquer nos résultats, mais les données et la méthodologie utilisées ne nous ont pas permis de fournir des éléments de réponse à ce sujet.

References

- Aassve, Arnstein, Simon Burgess, Carol Propper and Matt Dickson. 2006. "Employment, Family Union and Childbearing Decisions in Great Britain." *Journal of the Royal Statistical Society Series A*, 169(4): 781–804.
- Abbring, Jaap H. and Gerard J. van den Berg. 2003. "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica*, 71(5): 1491–1517.
- Abbring, Jaap H. and Gerard J. van den Berg. 2004. "Analyzing the Effect of Dynamically Assigned Treatments Using Duration Models, Binary Treatment Models, and Panel Data Models." *Empirical Economics*, 29(1): 5–20.
- Abdulkadiroğlu, Atila, Joshua Angrist, Yusuke Narita and Parag Pathak. 2017. "Research Design Meets Market Design: Using Centralized Assignment for Impact Evaluation." *Econometrica*, 85(5): 1373–1432.
- Abdulkadiroğlu, Atila, Joshua Angrist and Parag Pathak. 2014. "The Elite Illusion: Achievement Effects at Boston and New York Exam Schools." *Econometrica*, 82(1): 137–196.
- Adsera, Alicia. 2005. "Vanishing Children: From High Unemployment to Low Fertility in Developed Countries." *American Economic Review*, 95(2): 189–193.
- Angrist, Joshua. 2014. "The Perils of Peer Effects." *Labour Economics*, 30(C): 98–108.
- Arcidiacono, Peter, Esteban M. Aucejo and V. Joseph Hotz. 2016. "University Differences in the Graduation of Minorities in STEM Fields: Evidence from California." *American Economic Review*, 106(3): 525–562.
- Becker, Gary S. 1964. *Human Capital, A Theoretical and Empirical Analysis, with Special Reference to Education*. New York: Columbia University Press for the National Bureau of Economic Research.
- Becker, Gary S. 1981. *A Treatise on the Family*. Cambridge, MA: Harvard University Press.
- Bettinger, Eric P. and Bridget T. Long. 2005. "Do Faculty Serve as Role Models? The Impact of Instructor Gender on Female Students." *American Economic Review*, 95(2): 152–157.
- Black, Dan A., Terra G. McKinnish and Seth G. Sanders. 2003. "Does the Availability of High-Wage Jobs for Low-Skilled Men Affect Welfare Expenditures? Evidence from Shocks to the Steel and Coal Industries." *Journal of Public Economics*, 87(9–10): 1921–1942.

- Booij, Adam S., Edwin Leuven and Hessel Oosterbeek. 2017. "Ability Peer Effects in University: Evidence from a Randomized Experiment." *Review of Economic Studies*, 84(2): 547–578.
- Buckles, Kasey S. and Elizabeth L. Munnich. 2012. "Birth Spacing and Sibling Outcomes." *Journal of Human Resources*, 47(3): 613–642.
- Bursztyn, Leonardo, Thomas Fujiwara and Amanda Pallais. 2017. "‘Acting Wife’: Marriage Market Incentives and Labor Market Investments." *American Economic Review*, 107(11): 3288–3319.
- Buser, Thomas, Muriel Niederle and Hessel Oosterbeek. 2014. "Gender, Competitiveness, and Career Choices." *The Quarterly Journal of Economics*, 129(3): 1409–1447.
- Buser, Thomas, Noemi Peter and Stefan C. Wolter. 2017. "Gender, Competitiveness, and Study Choices in High School: Evidence from Switzerland." *American Economic Review*, 107(5): 125–130.
- Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica*, 82(6): 2295–2326.
- Card, David, Alexandre Mas and Jesse Rothstein. 2008. "Tipping and the Dynamics of Segregation." *The Quarterly Journal of Economics*, 123(1): 177–218.
- Carrell, Scott E., Marianne E. Page and James E. West. 2010. "Sex and Science: How Professor Gender Perpetuates the Gender Gap." *The Quarterly Journal of Economics*, 125(3): 1101–1144.
- Chen, Xianglei and Thomas Weko. 2009. "Students Who Study Science, Technology, Engineering, and Mathematics (STEM) in Postsecondary Education." Stats in Brief, U.S. Department of Education.
- Clark, Damon and Emilia Del Bono. 2016. "The Long-Run Effects of Attending an Elite School: Evidence from the United Kingdom." *American Economic Journal: Applied Economics*, 8(1): 150–76.
- Croson, Rachel and Uri Gneezy. 2009. "Gender Differences in Preferences." *Journal of Economic Literature*, 47(2): 448–474.
- Cullen, Julie B., Brian A. Jacob and Steven Levitt. 2006. "The Effect of School Choice on Participants: Evidence from Randomized Lotteries." *Econometrica*, 74(5): 1191–1230.

- d'Albis, Hippolyte, Angela Greulich and Grégory Ponthière. 2017. "Education, Labour, and the Demographic Consequences of Birth Postponement in Europe." *Demographic Research*, 36(23): 691–728.
- Dee, Thomas S. 2007. "Teachers and the Gender Gaps in Student Achievement." *Journal of Human Resources*, 42(3): 528–554.
- Direction de l'évaluation, de la prospective et de la performance. 2012. "L'état de l'école." Technical Report, Ministère de l'éducation nationale.
- Direction de l'évaluation, de la prospective et de la performance. 2014. "Filles et garçons sur le chemin de l'égalité de l'école à l'enseignement supérieur." Technical Report, Ministère de l'éducation nationale.
- Dobbie, Will and Roland G. Fryer Jr. 2014. "The Impact of Attending a School with High-Achieving Peers: Evidence from the New York City Exam Schools." *American Economic Journal: Applied Economics*, 6(3): 58–75.
- Dougherty, Shaun M., Joshua S. Goodman, Darryl V. Hill, Erica G. Litke and Lindsay C. Page. 2017. "Objective Course Placement and College Readiness: Evidence from Targeted Middle School Math Acceleration." *Economics of Education Review*, 58(C): 141–161.
- Duflo, Esther, Pascaline Dupas and Michael Kremer. 2011. "Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya." *American Economic Review*, 101(5): 1739–1774.
- Eisenkopf, Gerald, Zohal Hessami, Urs Fischbacher and Heinrich W. Ursprung. 2015. "Academic Performance and Single-Sex Schooling: Evidence from a Natural Experiment in Switzerland." *Journal of Economic Behavior & Organization*, 115: 123–143.
- Ekert-Jaffé, Olivia and Anne Solaz. 2001. "Unemployment, Marriage, and Cohabitation in France." *Journal of Behavioral and Experimental Economics (formerly The Journal of Socio-Economics)*, 30(1): 75–98.
- Ekert-Jaffé, Olivia and Anne Solaz. 2002. "Couple Formation in France: The Changing Importance of Labor Market Early Career Path." *Journal of Bioeconomics*, 4(3): 223–239.
- Estrada, Ricardo and Jérémie Gignoux. 2017. "Benefits to Elite Schools and the Formation of Expected Returns to Education: Evidence from Mexico City." *European Economic Review*, 95(C): 168–194.

- Fack, Gabrielle, Julien Grenet and Yinghua He. 2015. "Beyond Truth-Telling: Preference Estimation with Centralized School Choice." CEPR Discussion Papers, Center for Economic Policy Research.
- Fruehwirth, Jane C., Salvador Navarro and Yuya Takahashi. 2016. "How the Timing of Grade Retention Affects Outcomes: Identification and Estimation of Time-Varying Treatment Effects." *Journal of Labor Economics*, 34(4): 979–1021.
- Gary-Bobo, Robert J., Marion Goussé and Jean-Marc Robin. 2016. "Grade Retention and Unobserved Heterogeneity." *Quantitative Economics*, 7(3): 781–820.
- Gaure, Simen, Knut Røed and Tao Zhang. 2007. "Time and Causality: A Monte Carlo Assessment of the Timing-of-Events Approach." *Journal of Econometrics*, 141(2): 1159–1195.
- Goldin, Claudia. 2015. "Gender and the Undergraduate Economics Major: Notes on the Undergraduate Economics Major at a Highly Selective Liberal Arts College." Report on the status of undergraduate women in economics at Adams College.
- Goldstein, Joshua, Michaela Kreyenfeld, Aiva Jasilioniene and Deniz D. Karaman Örsal. 2013. "Fertility Reactions to the "Great Recession" in Europe: Recent Evidence from Order-Specific Data." *Demographic Research*, 29 85–104.
- Hansen, Bruce E. 2000. "Sample Splitting and Threshold Estimation." *Econometrica*, 68(3): 575–603.
- He, Ziwei. 2016. "The Impact of Academic Tracking Assignment on Students' Subject Specializations." Unpublished manuscript.
- Hill, Andrew J. 2014. "The Costs of Failure: Negative Externalities in High School Course Repetition." *Economics of Education Review*, 43(C): 91–105.
- Hiller, Victor and Olivier Tercieux. 2014. "Choix d'écoles en France. Une évaluation de la procédure Affelnet." *Revue économique*, 65(3): 619–656.
- Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *The Review of Economics and Statistics*, 91(4): 717–724.
- Huttunen, Kristiina and Jenni Kellokumpu. 2016. "The Effect of Job Displacement on Couples' Fertility Decisions." *Journal of Labor Economics*, 34(2): 403–442.
- Imbens, Guido and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies*, 79(3): 933–959.

- Jackson, C. Kirabo. 2010. "Do Students Benefit from Attending Better Schools? Evidence from Rule-Based Student Assignments in Trinidad and Tobago." *Economic Journal*, 120(549): 1399–1429.
- Jackson, C. Kirabo. 2012. "Single-Sex Schools, Student Achievement, and Course Selection: Evidence from Rule-Based Student Assignments in Trinidad and Tobago." *Journal of Public Economics*, 96(1): 173–187.
- Jackson, C. Kirabo. 2013. "Can Higher-Achieving Peers Explain the Benefits to Attending Selective Schools? Evidence from Trinidad and Tobago." *Journal of Public Economics*, 108: 63–77.
- Jacob, Brian A. and Lars Lefgren. 2004. "Remedial Education and Student Achievement: A Regression-Discontinuity Analysis." *The Review of Economics and Statistics*, 86(1): 226–244.
- Jacob, Brian A. and Lars Lefgren. 2009. "The Effect of Grade Retention on High School Completion." *American Economic Journal: Applied Economics*, 1(3): 33–58.
- Joensen, Juanna S. and Helena S. Nielsen. 2009. "Is there a Causal Effect of High School Math on Labor Market Outcomes?." *Journal of Human Resources*, 44(1): 171–198.
- Joensen, Juanna S. and Helena S. Nielsen. 2016. "Mathematics and Gender: Heterogeneity in Causes and Consequences." *Economic Journal*, 126(593): 1129–1163.
- Kalmijn, Matthijs. 2011. "The Influence of Men's Income and Employment on Marriage and Cohabitation: Testing Oppenheimer's Theory in Europe." *European Journal of Population/Revue européenne de Démographie*, 27(3): 269–293.
- Kleven, Henrik, Camille Landais and Jakob E. Sogaard. 2018. "Children and Gender Inequality: Evidence from Denmark." NBER Working Papers, National Bureau of Economic Research, Inc.
- Lavy, Victor, M. Daniele Paserman and Analia Schlosser. 2012. "Inside the Black Box of Ability Peer Effects: Evidence from Variation in the Proportion of Low Achievers in the Classroom." *Economic Journal*, 122(559): 208–237.
- Lavy, Victor and Edith Sand. 2015. "On The Origins of Gender Human Capital Gaps: Short and Long Term Consequences of Teachers' Stereotypical Biases." NBER Working Papers, National Bureau of Economic Research.
- Lee, David S. and Thomas Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of economic Literature*, 48(2): 281–355.

- Lillard, Lee A. 1993. "Simultaneous Equations for Hazards: Marriage Duration and Fertility Timing." *Journal of Econometrics*, 56(1-2): 189–217.
- Lillard, Lee and Linda Waite. 2000. "Marriage, Divorce and the Work and Earning Careers of Spouses." Working Papers, University of Michigan, Michigan Retirement Research Center.
- Lim, Jaegeum and Jonathan Meer. 2017. "The Impact of Teacher-Student Gender Matches: Random Assignment Evidence from South Korea." *Journal of Human Resources*, 52(4): 979–997.
- Mamun, Arif A. 2008. "Effects of Employment on Marriage: Evidence from a Randomized Study of the Job Corps Program." Technical Report, Washington, DC: Mathematica Policy Research.
- Manacorda, Marco. 2012. "The Cost of Grade Retention." *The Review of Economics and Statistics*, 94(2): 596–606.
- McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics*, 142(2): 698–714.
- MENESR. 2013. *Parcours et réussite en STS, IUT et CPGE*. Vol. n°6 of L'état de l'enseignement supérieur et de la Recherche en France.
- MESRI. 2012. "Les bacheliers S : des poursuites d'études de plus en plus dispersées." Note d'information Enseignement Supérieur & Recherche, MESR-SIES.
- MESRI. 2013. "Réussite et échec en premier cycle." Note d'information Enseignement Supérieur & Recherche, MESR-SIES.
- MESRI. 2015. "Réussite et échec en premier cycle." Note d'information Enseignement Supérieur & Recherche, MESR-SIES.
- Modena, Francesca, Concetta Rondinelli and Fabio Sabatini. 2014. "Economic Insecurity and Fertility Intentions: The Case of Italy." *Review of Income and Wealth*, 60(S1): 233–255.
- Murphy, Richard and Felix Weinhardt. 2014. "Top of the Class: The Importance of Ordinal Rank." CESifo Working Paper Series, CESifo Group Munich.
- Niedergesäss, Markus. 2013. "Employment, Partnership and Childbearing Decisions of German Women and Men: A Simultaneous Hazards Approach." University of Tuebingen Working Papers in Economics and Finance, University of Tuebingen, Faculty of Economics and Social Sciences.

- Niederle, Muriel. 2016. *Gender*. second edition edition 481–553, Handbook of Experimental Economics: Princeton University Press.
- Niederle, Muriel and Lise Vesterlund. 2007. “Do Women Shy Away From Competition? Do Men Compete Too Much?.” *The Quarterly Journal of Economics*, 122(3): 1067–1101.
- Niederle, Muriel and Lise Vesterlund. 2010. “Explaining the Gender Gap in Math Test Scores: The Role of Competition.” *Journal of Economic Perspectives*, 24(2): 129–144.
- Niederle, Muriel and Lise Vesterlund. 2011. “Gender and Competition.” *Annual Review of Economics*, 3: 601–630.
- Office for Standards in Education (OFSTED). 2015. “A-Level Subject Take-Up.” Technical Report.
- Olsen, Randall J. and George Farkas. 1990. “The Effect of Economic Opportunity and Family Background on Adolescent Cohabitation on Adolescent Cohabitation and Childbearing among Low-Income Blacks.” *Journal of Labor Economics*, 8(3): 341–362.
- Pailhé, Ariane and Anne Solaz. 2012. “The Influence of Employment Uncertainty on Childbearing in France: A Tempo or Quantum Effect?.” *Demographic Research*, 26(1): 1–40.
- Pison, Gilles. 2018. “1968-2018: Four Demographic Surprises in France over the Last 50 Years.” *Population and Societies*, 553: 1–6.
- Pomerantz, Eva M., Ellen R. Altermatt and Jill L. Saxon. 2002. “Making the Grade but Feeling Distressed: Gender Differences in Academic Performance and Internal Distress.” *Journal of Educational Psychology*, 94(2): 396–404.
- Pop-Eleches, Cristian and Miguel Urquiola. 2013. “Going to a Better School: Effects and Behavioral Responses.” *American Economic Review*, 103(4): 1289–1324.
- Porter, Jack and Ping Yu. 2015. “Regression Discontinuity Designs with Unknown Discontinuity Points: Testing and Estimation.” *Journal of Econometrics*, 189(1): 132–147.
- Prifti, Ervin and Daniela Vuri. 2013. “Employment Protection and Fertility: Evidence from the 1990 Italian Reform.” *Labour Economics*, 23(C): 77–88.
- Prioux, France. 2003. “Age at First Union in France: A Two-Stage Process of Change.” *Population (English edition)*, 58(4): 559–578.
- Rask, Kevin and Jill Tiefenthaler. 2008. “The Role of Grade Sensitivity in Explaining the Gender Imbalance in Undergraduate Economics.” *Economics of Education Review*, 27(6): 676–687.

-
- Roth, Alvin E. 1982. "The Economics of Matching: Stability and Incentives." *Mathematics of Operations Research*, 7(4): 617–628.
- Rothstein, Donna S. 1995. "Do Female Faculty Influence Female Students' Educational and Labor Market Attainments?" *Industrial and Labor Relations Review*, 48(3): 515–530.
- Schwerdt, Guido, Martin R. West and Marcus A. Winters. 2017. "The Effects of Test-Based Retention on Student Outcomes Over Time: Regression Discontinuity Evidence from Florida." *Journal of Public Economics*, 152(C): 154–169.
- Stinebrickner, Ralph and Todd R. Stinebrickner. 2014. "A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout." *Review of Economic Studies*, 81(1): 426–472.
- Tafreschi, Darjusch and Petra Thiemann. 2016. "Doing it Twice, Getting it Right? The Effects of Grade Retention and Course Repetition in Higher Education." *Economics of Education Review*, 55(C): 198–219.
- Vergauwen, Jorik, Karel Neels and Jonas Wood. 2016. "Educational Differentials in the Impact of Micro- and Macro-Level Economic Conditions on Union Formation in France (1993-2008)." *Population (english edition)*, 0(4): 593–617.
- Wiswall, Matthew and Basit Zafar. 2015. "Determinants of College Major Choice: Identification using an Information Experiment." *Review of Economic Studies*, 82(2): 791–824.
- Zafar, Basit. 2013. "College Major Choice and the Gender Gap." *Journal of Human Resources*, 48(3): 545–595.