Thèse

en vue de l'obtention du

Doctorat en Sciences Économiques

Délivré par l'école doctorale de l'École Polytechnique

Sous la direction de Christian Belzil

Présentée et soutenue publiquement le 2 septembre 2015 par :

Mathias André

Choix scolaires et trajectoires professionnelles

Essais sur le risque, la personnalité et les diplômes

Jury :

Christian Belzil	Professeur, École Polytechnique – *Directeur de Thèse*
Raïcho Bojilov	Professeur, École Polytechnique
Nicolas Jacquemet	Professeur, Université de Lorraine / PSE – *Rapporteur*
Thierry Kamionka	Directeur de recherche CNRS, Crest
Robert M. Sauer	Professeur, Université de Londres – *Rapporteur*

Unité de recherche : Laboratoire d'économétrie de l'École Polytechnique

Remerciements

Il est d'usage de faire commencer un manuscrit de thèse par «Les remerciements», sorte de passage obligé que beaucoup se contentent de lire d'ailleurs... Je ferai court pour laisser plus de temps au lecteur pressé pour s'intéresser au fond des travaux de cette thèse.

Je tiens en premier lieu à chaleureusement remercier mon directeur de thèse, Christian Belzil. Je voudrais lui exprimer ma gratitude pour avoir accepté de m'encadrer et avoir suivi mes travaux tout au long de cette thèse. Il a su m'apprendre à naviguer seul dans les méandres de l'économétrie structurelle. Si j'ai tant appris pendant ces années, c'est en grande partie grâce à lui, à sa disponibilité et à la confiance qu'il m'a accordée dans nos collaborations. Ses conseils ont aussi été précieux dans la découverte de ce milieu si particulier qu'est le monde universitaire. Il s'est révélé être un excellent encadrant pour mes travaux de recherches, et même le meilleur directeur de thèse que j'ai eu. Merci Christian.

Je tiens ensuite à remercier Nicolas Jacquemet et Robert Sauer d'avoir accepté d'être rapporteurs de ces travaux. Leur regard sur mes travaux après tant d'années à les faire avancer me sont précieux et me permettent d'envisager avec plaisir de les pousser plus loin. J'ai apprécié tout particulièrement l'exigence et la bienveillance dont ils ont su faire preuve pour les dernières étapes de mon doctorat.

Mes remerciements s'adressent aussi à Thierry Kamionka et Raïcho Bojilov qui ont accepté d'examiner cette thèse et de participer à ma soutenance. Leur appui n'a pas démarré avec la pré-soutenance car c'est aussi au cours des différentes présentations en séminaire au Crest et à Polytechnique, qu'ils ont su prendre le temps de discuter mes travaux et m'adresser de précieuses suggestions.

Je souhaiterais également pleinement remercier François Poinas et Maxime Tô, deux coauteurs avec qui j'ai échangé si souvent durant ces années de recherche et auprès de qui j'ai beaucoup appris. Je tiens à remercier les nombreux chercheurs à qui j'ai pu présenter mes travaux, au détour de séminaires et de conférences. Ils ont participé à les orienter et à les enrichir.

Durant ma thèse, j'ai pu profité de conditions de travail idéales et je tiens à

remercier les directeurs des laboratoires qui m'ont accueilli, au Crest comme à l'X : Francis Bloch, Xavier d'Haultfœuille, Thierry Kamionka et Pierre Picard. Je souhaite aussi exprimer ma gratitude à Alfred Galichon et Pierre Cahuc, qui n'ont pas seulement suivi l'avancée de mes travaux mais m'ont également permis d'être financé durant ma thèse et ce, dans les meilleures conditions.

Mes quatre années entre le Crest et l'École Polytechnique ont été rendues agréables par tous ceux que j'y ai côtoyés, et je les en remercie. Je pense plus particulièrement aux autres thésards que j'ai pu rencontrés et avec qui j'ai pu échangé sur nos travaux de recherche mais pas seulement. Et notamment celles et ceux avec qui j'ai partagé mes bureaux, au Crest, à l'X et à l'EUI. Je tiens aussi à remercier tous ceux que j'ai rencontrés à Florence, lors de ma visite au département d'économie de l'Institut européen universitaire. Cette magnifique expérience n'aurait pas été possible sans l'appui d'Alain Trognon pour ma formation complémentaire d'administrateur de l'Insee et l'accueil de Jérôme Adda, qu'ils en soient ici vivement remerciés.

Je tiens à exprimer ma reconnaissance à mes collègues de la Drees, qui ont su m'épauler lors des derniers mois pour finaliser ce manuscrit.

Enfin, je remercie particulièrement tous mes proches qui m'ont soutenu, entouré et encouragé au long de ces années. Aucun n'a douté de moi et je les en remercie avec affection. Et toi aussi lecteur pressé, je te remercie de ne pas t'arrêter là et d'entamer la découverte de mes travaux dans les pages d'introduction qui suivent.

Table des matières

Introduction générale

Cette thèse rassemble trois travaux en microéconométrie du travail et de l'éducation. Ces travaux ont en commun de s'intéresser aux déterminants de choix individuels. Il s'agit d'étudier les caractéristiques reliant les décisions passées aux positions futures. Bien que ces recherches soient relativement indépendantes, nous nous concentrons sur l'influence du risque, de la personnalité et des diplômes sur les trajectoires individuelles professionnelles. Nous attachons ainsi une importance à l'étude des choix scolaires et aux inégalités qui en découlent sur le marché du travail.

La contribution de cette thèse à la recherche en sciences économiques s'inscrit dans un cadre empirique dans la mesure où elle cherche à répondre à des problématiques originales en se basant sur des données microéconomiques adaptées et une modélisation cohérente des décisions individuelles. Une attention particulière est accordée à l'étude de l'endogénéité des choix.

Massification scolaire et augmentation du chômage

Une publication récente de l'Institut national des statistiques et des études économiques (Insee)[1] présente les évolutions des trajectoires scolaires et situations professionnelles sur les trente dernières années. Elle fournit le cadre statistique descriptif sur longue période des phénomènes que cette thèse aborde. On se concentre ici sur le cas français mais les tendances sont similaires pour les économies développées européennes ou nord-américaines.

[1] Trente ans de vie économique et sociale, Insee Références, édition 2014.

Selon ce rapport, depuis 1985, l'espérance de scolarisation à la naissance a fortement augmenté, passant de 19,7 ans à 21,5 ans. Plus de quatre jeunes sur dix sortent du système scolaire avec un diplôme de l'enseignement supérieur, contre moins de deux sur dix il y a trente ans. Le baccalauréat s'est généralisé dans tous les milieux sociaux mais les types de baccalauréat obtenu sont divers. Les enfants des milieux défavorisés obtiennent plus souvent des baccalauréats professionnels et ils poursuivent des études supérieures plus courtes (BTS-DUT). À l'autre bout de l'échelle sociale, un baccalauréat est la norme ainsi que des études supérieures dans des filières plus longues et plus sélectives.

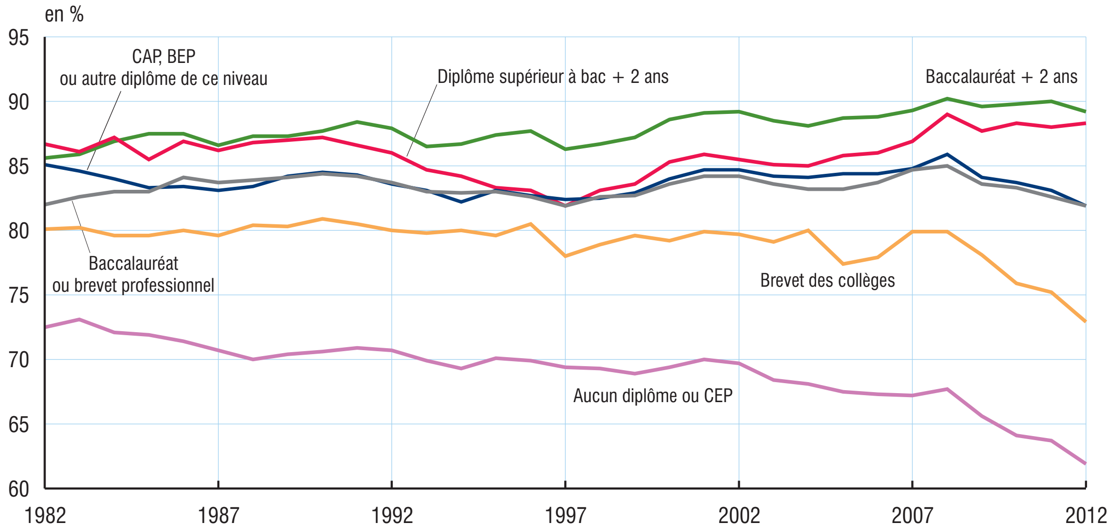
En trente ans, quitter le système scolaire sans diplôme est devenu plus rare, mais cela reste fréquent, surtout dans les milieux défavorisés : 21 % des enfants d'ouvriers ou d'employés ayant quitté l'école entre 2008 et 2010, sortent sans diplôme contre 7 % des enfants de cadres ou professions intermédiaires. De manière similaire, un quart des élèves qui étaient en sixième en 1995 n'ont pas terminé leur cursus secondaire avec un diplôme[2]. Le parcours scolaire est corrélé aux caractéristiques socio-économiques et détermine l'insertion professionnelle des individus. Ainsi, en 2010, la moitié des jeunes sans diplôme se retrouve au chômage en entrant sur le marché du travail.

Les choix scolaires impliquent donc des situations distinctes sur le marché du travail. Par exemple, en 2009, au plus fort de la crise, près d'un jeune actif sans diplôme sur deux connaissait une période de chômage d'une à quatre années après la fin des études, contre un sur dix pour les diplômés du supérieur. Depuis 1980, le taux de chômage des non-diplômés dépasse de vingt points celui des diplômés du supérieur. Cet écart a augmenté entre 2003 et 2011, montrant que la crise des dernières années a plus fortement touché les moins diplômés.

Parallèlement aux mutations scolaires, le monde du travail a profondément évolué dans les dernières décennies, principalement sous l'effet de la démographie et des choix d'activité. Selon l'Insee, les femmes se sont largement insérées sur le marché

[2] *Les décrocheurs du système éducatif : de qui parle-t-on ?*, France portrait social, Insee Références, édition 2013.

FIGURE 1 – Taux d'emploi des 25 à 49 ans selon le diplôme le plus élevé



Champ : France métropolitaine, population des ménages, personnes de 25 à 49 ans.
Note : données corrigées pour les ruptures de série.
Source : Insee, enquêtes Emploi.

du travail, les jeunes et les seniors en ont été en partie exclus. La structure des emplois a beaucoup changé, notamment par l'augmentation nette du chômage et la multiplication du temps partiel. En 2012, les chômeurs représentent 10 % de la population active contre 3 % en 1975 et les personnes travaillant à temps partiel, 16 % contre 7 %. C'est entre 1975 et 1985 que le taux de chômage a nettement augmenté pour ensuite fluctuer avec le cycle économique. Mais les caractéristiques individuelles telles que l'âge et la qualification ont engendré des trajectoires diverses. Le graphique 1 indique l'évolution du taux d'emploi sur trente ans par catégorie de diplôme, montrant qu'une scolarité élevée protège du chômage.

Il ressort de l'analyse descriptive précédente que les trajectoires professionnelles sont influencées par les qualifications obtenues dans le système scolaire, en plus des caractéristiques socio-économiques usuelles et des évolutions macroéconomiques. La question des déterminants des choix scolaires est donc cruciale. D'un point de vue théorique, ce thème est traité par l'économie de l'éducation, cadre dans lequel souhaite s'inscrire cette thèse.

Des théories complémentaires

En sciences sociales, un angle récurrent d'analyse des sociétés actuelles consiste à s'intéresser à la mobilité sociale. Il s'agit d'étudier la transmission intergénérationnelle des positions sociales, prises au sens large, comme par exemple, la place dans l'échelle des revenus ou bien l'appartenance à différents types de catégories socioprofessionnelles.

Les recherches récentes ont confirmé le constat ancien que l'éducation est la variable la plus importante pour assurer la mobilité sociale. Il a ainsi été démontré que les diplômes sont un des principaux déterminants de la position dans la distribution des revenus et des inégalités sociales, notamment via leur impact sur la qualité des emplois, la probabilité d'être au chômage et la progression dans les carrières. Dans la théorie économique, il existe principalement trois vecteurs par lesquels l'éducation d'un individu peut influencer sa trajectoire socioprofessionnelle.

La théorie dite du «capital humain» (Mincer (1958), Becker (1962)) assimile l'éducation à un investissement. Dans un premier temps, elle s'apparente pour l'individu à une dépense, constituée notamment de coûts directs tels que les droits de scolarité ou les coûts de mobilité géographique. Elle implique aussi un coût d'opportunité correspondant au salaire non perçu en renonçant à travailler pendant la période de scolarité. La justification d'un tel investissement s'appuie alors sur la recherche d'une augmentation de sa productivité et ses connaissances pour ensuite les valoriser sur le marché du travail. Il pourra ainsi diminuer sa probabilité d'être au chômage ou augmenter son salaire. Dans ce cadre, l'éducation constitue une composante du capital humain dont le rendement permet d'accéder à des revenus supérieurs et de meilleures opportunités de carrière.

Mincer (1974) a ainsi proposé une spécification linéaire simple de l'accumulation du capital humain développée par Becker, liant le niveau d'éducation et l'expérience professionnelle au salaire. Les rendements de l'éducation correspondent aux gains de revenus salariaux imputables à l'allongement de la durée des études. Cette spécification a donné lieu à de multiples travaux empiriques et à une vaste littérature. De nombreuses spécifications de l'équation de Mincer ont été ainsi testées en contrôlant par un ensemble de caractéristiques individuelles et collectives : genre, secteur d'activité, statut matrimonial, etc. Heckman, Lochner, and Todd (2006a) proposent une revue de littérature qui réconcilie les résultats empiriques avec un socle théorique, en s'appuyant notamment sur les fonctions valeurs associées. Cette étude permet de comprendre l'interprétation économique de l'équation de Mincer en termes de rendements *ex post* sur le marché du travail.

L'équation de Mincer étant fondamentalement statique, elle s'appuie sur l'exogénéité des parcours post scolarité. Les besoins d'un cadre dynamique ont poussé Ben-Porath (1967) à considérer un modèle dynamique de choix discrets d'investissement en capital humain (Mincer (1994)). D'une manière générale, il s'agit de postuler l'existence d'une fonction de production déterminant les aptitudes des individus à travailler en fonction des compétences et des savoirs acquis. L'expérience accumulée sur le marché du travail et l'éducation en sont les principaux facteurs. L'arbitrage intertemporel entre ces deux facteurs substituables détermine les choix d'éducation, supposés optimaux pour les agents.

Une théorie alternative dite «du signal» Spence (1973) stipule que l'éducation constitue un simple présage et non une mesure intrinsèque de la productivité. Le niveau d'éducation obtenu n'a, dans ce cadre, qu'une valeur indicative sur les capacités individuelles et ne fournit pas une mesure directe et précise des connaissances et compétences acquises. Ceci s'applique principalement à l'enseignement supérieur, une fois que les savoirs de base sont assimilés par tous. En France, le cas des grandes écoles en offre un exemple frappant : les étudiants font valoir leur diplôme en tant que tel, plus que des compétences ou savoirs acquis durant leur cursus. Selon cette théorie du signal, les employeurs utilisent le diplôme comme un indicateur de la productivité pour distinguer les individus et les classer mais sans être capable de déterminer *a priori* leurs capacités réelles.

Un autre type de modèles, non contradictoires avec ceux qui précèdent, se concentrent sur les caractéristiques intrinsèques du marché du travail (*job search*, voir notamment Mortensen and Pissarides (1994)). Ces travaux attribuent moins d'importance aux différences individuelles comme déterminants des choix observés. Ce sont les conditions et les caractéristiques du marché du travail qui orientent les trajectoires, principalement par le biais principal des emplois disponibles, c'est-à-dire les positions vacantes proposées par les entreprises. L'éducation permet alors de classer les individus pour un poste vacant donné. Mais le véritable mécanisme à la source de la mobilité est la disponibilité des offres d'emploi. Ce cadre accorde une importance première aux déterminants et aux conséquences des rigidités sur le marché du travail. Les études s'intégrant dans ce schéma de pensée cherchent à modéliser et reproduire les «imperfections» du marché du travail.

La discussion précédente est usuelle en économie de l'éducation. Elle a été complétée par de nombreux travaux plus récents qui se sont concentrés sur l'élargissement des modèles originaux à des mécanismes de transition entre système éducatif et marché du travail plus précis. En effet, au-delà de la situation conjoncturelle, il existe une grande diversité des situations individuelles pour un diplôme donné. C'est pourquoi une analyse au niveau individuel nous semble nécessaire pour mettre en avant les caractéristiques qui influencent la valorisation des niveaux d'éducation au début de la carrière professionnelle.

L'analyse microéconométrique comme cadre méthodologique

Un attention méthodologique particulière est nécessaire pour estimer les réels déterminants des choix scolaires. En effet, les comportements s'expriment au sein du modèle dans lequel ils sont étudiés et sont eux-mêmes générés par les conditions et les hypothèses du modèle. Les choix sont contingents au modèle au sens où celui-ci précise explicitement les coûts et les bénéfices que peuvent retirer les individus de ces investissements éducatifs. En outre, la difficulté de l'analyse empirique réside principalement dans le fait que l'ensemble des déterminants n'est en général pas observé par l'économètre. Et l'omission d'un déterminant du choix qui a lui-même un impact direct sur le bénéfice de ce choix implique un biais d'estimation.

Pour illustrer ce propos, prenons l'exemple d'un étudiant motivé qui souhaite suivre une filière précise à l'université, dans laquelle le recrutement se fait uniquement sur la base des résultats à une seule option donnée. En outre, supposons pour simplifier que seule la motivation des élèves influence la réussite dans cette option-là mais que l'économètre n'observe pas la motivation des individus. Comme la motivation influence également les résultats dans les autres matières scolaires, les plus motivés ont de meilleurs résultats dans toutes les options. Sans prendre en compte la motivation comme déterminant explicite du choix et du résultat, une analyse empirique simple montrera que les étudiants de cette filière sont ceux qui obtiennent des meilleurs résultats scolaires dans toutes les options. Cette conclusion est fallacieuse car à motivation donnée, le rôle des autres options est nul pour le recrutement dans cette filière par construction. Le raisonnement est similaire si la motivation est mesurée imparfaitement ou s'il y a plusieurs vecteurs de sélection endogène.

Plus généralement, Roy (1951) a montré que le mécanisme de «sélection» impose aux économistes de prendre en compte les corrélations entre les déterminants des choix et les choix étudiés. Dans le cadre d'une décision binaire, les individus se sélectionnent selon la position qu'ils ont le plus intérêt à adopter. Les deux populations, une fois le choix effectué, se distinguent également par des caractéristiques *ex ante* sans que cela ne puisse s'interprèter comme un effet direct de ces caractéristiques sur la variable de choix. L'économètre doit donc en tenir compte s'il souhaite comparer deux populations basées sur un choix individuel. Les nombreux travaux d'Heckman (voir par exemple Heckman (1990) et Heckman (2008)) ont apporté des réponses empiriques pour résoudre cela. Il s'agit d'utiliser des sources de variabilité jouant sur les mécanismes de sélection sans avoir d'effet sur les bénéfices attendus après le choix. Ces variables sont dite exogènes et permettent l'identification de l'effet recherché.

La microéconométrie fournit un large panorama de stratégies empiriques pour estimer les effets recherchés. Le choix de la méthode est guidé par la question posée et les données disponibles pour y répondre. Les variables disponibles ou bien le nombre d'individus présents dans une enquête limitent ainsi les méthodes qui peuvent être appliquées. L'économétrie structurelle par exemple est une branche de la microéconométrie qui s'intéresse à l'estimation des primitives des mécanismes économiques, paramètres invariants aux politiques et déterminant le comportement des agents. Il s'agit de modéliser les décisions des agents sous certaines hypothèses, de résoudre le modèle pour une valeur donnée des paramètres du modèles et d'estimer ces paramètres via une fonction objectif, généralement la fonction de vraisemblance. Les méthodes d'estimation sont en général exigeantes en termes de programmation mais un des principaux avantages est de permettre la simulation de contrefactuels, une fois les estimations obtenues (tels que dans Belzil and Hansen (2002) ou Cameron and Heckman (2001) par exemple).

En outre, un avantage d'une telle approche est d'être explicite sur les hypothèses sous-jacentes aux estimations, basées sur l'écriture de modèle économique correspondant. Un des champs porteur de cette littérature est constitué des modèles dits de *job search* : il s'agit de modéliser les suites de choix discrets des individus sur le marché du travail soumis à la possibilité de chômage. Un autre domaine de l'économétrie structurelle s'intéresse au problème d'arrêt optimal, par exemple pour le choix de la dernière année d'études.

Cette branche de l'économétrie permet ainsi de modéliser les suites de choix rationnels des agents en s'appuyant sur les outils de d'optimisation dynamique aux méthodes récursives, en partie développés par Bellman. Le caractère dynamique du modèle correspond à la résolution intertemporelle du choix des agents. Selon les modèles et les travaux, la dimension intertemporelle peut être résolue formellement (explicite mathématiquement), estimée numériquement ou approchée par différentes méthodes. Tenir compte de la dynamique, c'est-à-dire de l'influence actualisée des choix futurs et des conséquences des trajectoires, ajoute une dimension supplémentaire riche à une modélisation de base. Enfin, l'estimation des paramètres repose sur des théorèmes d'identification relativement puissants des modèles de choix discrets dynamiques (Magnac and Thesmar (2002a)).

Belzil (2007) présente les travaux les plus influents de la littérature structurelle sur les rendements de l'éducation, en étudiant notamment les différences avec l'approche par instrumentation. Il commente entre autres les papiers de Willis and Rosen (1979) et Keane and Wolpin (1997), contributions les plus importantes de cette littérature. Willis and Rosen (1979) s'appuient sur le modèle d'autosélection de Roy pour modéliser la décision individuelle de poursuite des études à l'université. Le modèle semi-structurel repose sur les avantages comparatifs de productivité entre les individus qui maximisent leur revenu intertemporel selon le modèle de Becker et Mincer.

Keane and Wolpin (1997) développent un modèle de choix entre éducation mais aussi occupation sur le marché du travail dont la modélisation et l'estimation structurelles permettent de tenir compte d'une grande hétérogénéité individuelle. Cependant, il est clair que les décisions des agents s'effectuent en présence de mécanismes non prévisibles. Par exemple, la valeur attendue des rendements peut s'avérer être différente des gains salariaux effectifs, observée une fois sur le marché du travail. Les chocs futurs sur les salaires ne peuvent pas non plus être prévus parfaitement. En outre, la réussite aux examens n'est pas assurée si l'individu choisit de prolonger sa formation et la qualité du diplôme n'est pas parfaitement anticipée par les étudiants au moment de leur choix. C'est pour cela que nous nous intéressons à l'influence du risque et au rôle des traits psychologiques.

Cadre et contributions de la thèse

Une diversité de mécanismes liant scolarité et parcours professionnels

Depuis l'après-guerre, le système scolaire a évolué dans son ensemble en raison notamment de la massification de l'enseignement. L'augmentation conjointe de la productivité dans la même période semble contredire la théorie du signal et plutôt valider le rôle du capital humain. Mais dans un contexte actuel avec un taux de chômage élevé, notamment pour les jeunes et même pour les plus diplômés, la grande disparité des diplômes d'un même niveau peut empêcher les employeurs de traduire le niveau d'études en compétences. Ainsi, la tertiarisation de l'économie alliée à une hausse rapide du progrès technique, ont mené nos économies vers ce qui est appelée la polarisation des emplois[3] stipulant que les créations et destructions d'emploi diffèrent selon les secteurs et leur intensité technologique.

Les travaux empiriques ont rejeté un lien direct entre le progrès technique et le chômage (voir notamment Bean and Pissarides (1993)). Mais l'influence du progrès technique existe de façon dynamique : les salariés les plus qualifiés sont d'autant plus productifs que la technologie progresse car ils s'adaptent mieux et plus vite. Ce cadre permet de fournir une explication à l'augmentation des inégalités salariales entre les diplômés du supérieur et du secondaire observée depuis les années 1990 (Acemoglu and Autor (2011)). Et cet effet se retrouve aussi sur la marge extensive, à savoir la participation au marché du travail : l'emploi très qualifié progresse alors que l'emploi peu qualifié stagne voire régresse. La technologie permet d'automatiser certaines tâches et accentue ainsi ce mécanisme de polarisation des emplois. Aussi, le rôle des diplômes est d'autant plus central que les emplois moins qualifiés sont plus exposés au risque du chômage. En outre, d'après Autor (2011) les échanges

[3]cf. Pierre Cahuc, *La polarisation des emplois*, Chaire «sécurisation des parcours professionnels».

commerciaux mondiaux accélèrent la disparition des métiers manufacturiers dans les pays développés et accentuent d'autant plus la polarisation des emplois, en en sélectionnant une partie. Les effets sont différents selon le secteur d'activité où les interactions sociales sont cruciales et donc hors de portée des systèmes informatisés. C'est par exemple le cas des services ou des métiers nécessitant abstraction ou adaptation. Autor (2014) propose des pistes de réflexion au sujet des complémentarités ou substitutions entre capital humain et technologiques informatiques (soulignant que «paradoxe de Polanyi», *we can know more than we can tell*, pourrait être dépassé par des machines qui apprennent des exemples humains).

Les services aux entreprises offrent des carrières plus stables et mieux rémunérées que les secteurs manufacturiers ou agricoles. Ainsi, les disparités dans les secteurs ont permis d'accroître la mobilité intergénérationelle en termes absolus. Mais il semble que ces mécanismes soient maintenant stabilisés alors que les inégalités de revenus et d'accès à l'emploi augmentent. Si l'éducation est vue comme un moyen de se positionner et de se classer dans une distribution au sein d'une population, alors chacun est incité à chercher plus et mieux, à niveau de diplôme donné. Des disparités existent ainsi au sein d'un même niveau de diplôme. C'est pourquoi la valeur des diplômes est maintenant interprétée différemment selon le lieu où il été délivré et non plus en tant que telle.

Les «nouvelles classes moyennes»[4] perçoivent ainsi la compétition scolaire comme une stratégie de «résistance au déclassement scolaire». En se rapprochant donc de la théorie du signal, en tout cas dans la perception et l'anticipation qu'ont les familles des classes moyennes, Goux et Maurin présentent la réussite scolaire comme le seul objectif pour se classer par rapport aux autres, peu importe la matière ou la filière. C'est dans ce cadre que le «sésame des grandes écoles» engendre attentes et frustrations selon la position des ménages dans cette «compétition scolaire». Ainsi le choix des trajectoires scolaires relèvent joue un rôle central dans la dynamique

[4]D. Goux et É. Maurin,*Les nouvelles classes moyennes*, éd. Le Seuil, 2012.

sociale, tout du moins dans la vision qu'en ont les individus.

L'accès généralisé à l'éducation, vue sous différents angles est donc un vecteur de mobilité sociale et influence les inégalités au premier ordre. Néanmoins, elle ne garantit pas à elle seule l'égalité. Ainsi, l'origine sociale a été depuis longtemps documentée comme une influence persistante, dans l'accès à l'éducation en premier lieu. Mais même après avoir contrôlé par les diplômes obtenus, les carrières sont aussi influencées par des mécanismes diffus tels que les effets de réseau ou de transmission des richesses permettant par exemple l'accès au logement et notamment l'adaptation à la carte scolaire en France.

Cependant, dans les économies tertiarisées telles que la France, d'autres facteurs sont mis en évidence pour lier capital humain et trajectoires professionnelles. Ainsi les *soft skills*, les compétences dites non techniques, comme la personnalité, la confiance en soi ou dans les autres, le *locus of control* (capacité d'un individu à penser qu'il influence, ou non, son environnement direct et que ses décisions ou ses actions déterminent le cours de sa vie) sont introduites comme variables individuelles influençant les choix éducatifs et la réussite scolaire.

Comme ce sont principalement des caractéristiques qui sont transmises par la famille et l'environnement non éducatif, elles peuvent être un des vecteurs explicatifs des inégalités scolaires. C'est par exemple le cas pour la socialisation primaire, notamment en ce qui concerne les femmes. Les traits psychologiques sont issus de mécanismes encore difficilement expliqués par les sciences sociales ou cognitives. Sans oublier les déterminants systémiques telles que les habitudes sociales du corps enseignant ou pédagogique, les traits psychologiques ont une influence directe sur les choix individuels. Le «plafond de verre» peut ainsi s'interpréter dans une certaine mesure par un mécanisme d'auto-sélection, certaines femmes refusant de jouer le jeu de la compétition sociale ou s'estimant, à tort, moins légitimes ou compétentes que certains hommes. Et ceci engendre donc des populations différentes selon le genre, que ce soit dans certaines filières du système scolaire ou plus tard, dans certains types de métiers, plus rémunérateurs, compétitifs ou risqués.

Risque et capital humain

Un des angles novateurs que propose cette thèse est d'approfondir l'influence de l'aversion au risque dans les décisions scolaires et les trajectoires professionnelles. Comme nous l'avons vu, les déterminants des décisions d'éducation sont multiples et s'inscrivent au sein d'un processus d'arbitrage complexe. Nous proposons de compléter les analyses usuelles en étudiant le rôle des traits psychologiques sur les choix scolaires. Considérée comme un investissement dans la théorie du capital humain, l'éducation possède un rendement et donc, il est nécessaire d'étudier les risques afférents à cet investissement. Les travaux de recherche de cette thèse se proposent de contribuer en quantifiant la manière dont la notion de risque intervient dans les décisions d'orientation des agents. Ils reposent sur le constat que l'investissement éducatif n'a pas un rendement certain et que sa valeur anticipée dans la population est hétérogène. Dans ce cadre sous incertitude, l'aversion au risque influence les choix éducatifs.

Certains économistes ont analysé l'effet de l'incertitude sur les décisions d'éducation dans des modèles dynamiques de choix discrets mais la nouveauté de cet apport est de tenir compte des préférences intrinsèques des agents. En effet, une littérature récente s'intéresse à l'importance des préférences individuelles, et notamment à l'impact de l'hétérogénéité des agents vis-à-vis du risque dans les décisions d'orientation scolaire. La notion de risque, absente de l'analyse de Mincer, doit donc être prise en compte dans ce contexte d'investissement sous incertitudes. Un numéro spécial de *Labour Economics* a ainsi été consacré en 2007 à la notion de risque dans les décisions éducatives. Cette publication se concentre sur des aspects assez descriptifs mais montre que cette problématique est porteuse au sein de la recherche actuelle et que de nombreux aspects nécessitent encore d'être approfondis.

L'éducation peut dans un premier temps être considérée comme un investissement risqué. Les coûts (direct, psychologique ou d'opportunité) associés à la poursuite des études implique des décisions dont les rendements ne sont pas certains étant données les incertitudes sur le marché du travail. Le risque associé à la poursuite des études n'est donc pas négligeable. Toutes choses égales par ailleurs, les agents ayant peur du risque peuvent préférer s'orienter vers des cursus courts ou avec des rendements moins incertains.

Cependant, même s'il apparaît a priori que l'accès à un niveau d'éducation plus élevé peut représenter une plus haute exposition au risque, le cas inverse est aussi envisageable d'un point de vue théorique. Les incertitudes du côté de la demande de travail sont importantes. L'éducation peut en effet modifier les trajectoires sur le marché du travail (par exemple, en diminuant la volatilité des revenus ou en augmentant la probabilité d'offres de travail, etc.). C'est-à-dire que pour beaucoup d'individus, l'obtention d'un niveau d'éducation plus élevé peut représenter une forme d'assurance, en permettant aux individus de s'adapter plus facilement aux changements technologiques dans un secteur donné. Elle implique alors une réduction dans le niveau de risque *ex ante*. L'aversion au risque a alors un effet contraire à la situation précédente car les agents peu amènes à prendre des risques ont tendance à choisir des cursus plus longs.

Par exemple, dans une étude sur données espagnoles de Hartog and Díaz-Serrano (2007), les auteurs trouvent un rôle primordial du risque (mesuré comme la variance des revenus) dans la demande d'éducation supérieure avec ces deux effets distincts : l'aversion au risque limite l'entrée à l'université (l'investissement est perçue comme risqué) sauf pour les ménages très peu averses au risque qui adoptent un comportement inverse (l'université est alors vue comme une assurance). Ainsi, le risque peut être différencié selon le niveau d'étude et plus généralement, selon le type de diplôme ou les filières scolaires.

Essais sur le risque, la personnalité et les diplômes

Les chapitres de cette thèse se veulent apporter une contribution aux débats présentés ci-dessus. Le capital humain au moment d'entrer sur le marché du travail correspond aux connaissances et savoirs accumulés dans le système scolaire. Mais sa composition peut être plus riche que la théorie l'avait initialement stipulée. Le premier article de cette thèse introduit une seconde dimension dans le capital humain accumulé à l'école. Nous appelons «spécialisation» la nature du diplôme, une fois le nombre donné d'années d'étude. Il s'agit de documenter l'hétérogénéité des trajectoires professionnelles pour un même niveau d'études. La cadre adopté cherche à évaluer l'importance de la professionnalisation au cours des études sur l'entrée sur le marché du travail.

L'acquisition d'un savoir professionnel au cours des études est une question récurrente dans le débat français et international : une spécialisation précoce au cours des études est par exemple souvent associée à des difficultés d'adaptation sur le marché du travail en cas de perte d'emploi. L'allongement général des études et une spécialisation plus tardive n'est cependant pas sans conséquence pour les jeunes qui se dirigent préférablement vers des études courtes. On peut ainsi confronter la logique des systèmes allemand, où les cursus professionnalisants sont mis en place très jeune, au système anglais et nord-américain, où les étudiants se voient proposer un large panel de matières jusque tard dans les études supérieures. En France, l'Insee indique que 21% des décrocheurs du système scolaire sont principalement passés par des classes spécialisées. La compréhension des dynamiques d'éducation en termes de prolongement des études et de spécialisation constitue donc un aspect important pour alimenter les instruments de politique publique.

Dans le cadre de la thématique de choix en univers incertain, la spécialisation peut offrir une assurance vis-à-vis des chocs sur le marché du travail. En effet, à niveau scolaire donné, un diplôme spécialisé implique une exposition au risque différencié si un choc technologique dans le secteur associé se produit. À l'image d'un arbitrage entre *adaptabilité* entre les secteurs et *productivité* dans un secteur donné, un diplôme général permet d'avoir une trajectoire sur le marché du travail moins sensible aux évolutions technologiques. Dans une certaine mesure, un parallèle peut être fait avec la théorie des postes vacants des modèles de *job search*, réconciliant la deuxième dimension du capital humain avec un critère de tri par grappes au sein de chaque secteur. Ainsi, la dimension horizontale des choix scolaires est primordiale, une fois que l'on a contrôlé par le niveau d'éducation, exprimé usuellement en nombre d'années d'études (dimension verticale).

Le premier chapitre de cette thèse se concentre sur les trajectoires en début de carrière professionnelle et s'inspire des travaux effectués sur la modélisation des choix discret dynamique développés par Keane and Wolpin (1997). Il modélise de manière explicite les choix successifs d'éducation à chaque étape de la scolarité. Dans ce modèle, les capacités d'accueil local des institutions scolaires (lycées, écoles, etc.) sont utilisées comme déterminant des choix d'éducation : elles contraignent les choix de scolarisation et de spécialisation sans pour autant être liées aux revenus futurs des individus. Les résultats des estimations obtenus à partir de l'enquête Génération 1998 du Céreq montrent que la spécialisation a un impact prégnant sur les futures trajectoires professionnelles : choisir un diplôme qui oriente vers les métiers de l'industrie donne lieu à des revenus supérieurs, notamment si la spécialisation a lieu après le baccalauréat.

Les autres travaux de cette thèse se concentrent sur le rôle de la personnalité et notamment de l'aversion au risque sur le marché du travail et dans les parcours scolaires. Le deuxième chapitre est un *companion paper* d'un projet plus ambitieux, non présenté dans cette thèse.

Partant du constat que la décision d'effectuer des études supérieures est multidimensionnelle, il s'agit d'étudier comment elle peut s'appuyer à la fois sur les capacités cognitives (par l'intermédiaire des résultats scolaires), les préférences individuelles (telles que l'aversion au risque) mais aussi les traits de personnalité (la confiance en soi, la motivation ou autres aspects du « Big Five » développé par les psychologues) et le contexte socio-économique (situation des parents ou variables macroéconomiques). En général, il est difficile d'avoir de telles mesures dans un cadre unifié à la fois au niveau empirique que théorique, notamment en raison de la corrélation avec les caractéristiques des parents ou d'effets individuels spécifiques. À partir des données allemandes du GSOEP, nous appliquons les modèles à facteurs développés par J. Heckman et ses coauteurs à un modèle de choix d'éducation en utilisant un questionnaire de personnalité : notre modèle s'inspire ainsi à la fois de la psychométrie et de la microéconomie.

Le projet original se propose de comprendre comment l'aversion au risque influence la décision de faire des études supérieures. Ainsi, il est crucial de modéliser précisément les mécanismes incertains à l'oeuvre au moment des choix. Une attention particulière est accordée à la modélisation et l'estimation de l'aversion au risque et l'influence de ce facteur dans les décisions individuelles.

Nous écrivons les choix individuels comme une forme réduite d'un modèle dynamique de choix discret et décomposons les déterminants de l'éducation en différents éléments distincts : aversion au risque, capacités cognitives, motivation, disposition à être consciencieux, confiance, réciprocité et revenus parentaux. Tous les facteurs sont mesurés à 17 ans, avant le choix d'éducation que nous étudions, la probabilité d'entrer à l'université, et dépendent des caractéristiques des parents.

Le chapitre inclus dans cette thèse étudie plus en détail la distribution de l'aversion au risque et des traits psychologiques, mesurés à 17 et 18 ans. Le modèle développé fait correspondre les caractéristiques familiales et des facteurs latents inobservés, en tenant compte de l'erreur de mesure. Notre analyse se concentre sur la distinction entre les deux voies scolaires, académique ou professionnelle.

Les résultats montrent que la différence de position scolaire à 17 ans, académique ou professionnelle, implique de fortes différences dans les facteurs latents, tels que l'intelligence et la motivation. En revanche, l'aversion au risque est faiblement corrélée à la trajectoire scolaire. L'hypothèse usuelle d'orthogonalité entre les facteurs ne résiste pas à ces estimations : en particulier, les facteurs cognitifs et non cognitifs sont très corrélés alors que l'aversion au risque l'est moins.

Ainsi, la faible corrélation entre l'aversion au risque et la position scolaire à 17 ans, est compatible avec l'existence d'une relation causale entre l'aversion au risque et les choix scolaires futures, notamment le choix de faire des études supérieures. Les travaux parallèles devront donc chercher à expliquer si le fait d'avoir effectué un parcours académique avant les études supérieures est un déterminant essentiel des choix individuels dans la mesure où le risque est accru pour les trajectoires éducatives supérieures.

Le troisième et dernier chapitre utilise les données américaines du PSID afin d'étudier le lien entre le salaire et l'aversion au risque. Théoriquement, les individus qui ont plus tendance à prendre des risques rencontrent une croissance de revenus plus importante. Il s'agit ici de combler un manque de la littérature en enrichissant les composants d'hétérogénéité individuelle influençant la croissance des revenus.

Une grande partie des débats sur les séries individuelles de revenus a porté sur l'importance des chocs persistants sur l'évolution des revenus et de l'influence que peut avoir l'hétérogénéité individuelle sur ces trajectoires alors que les fondations théoriques de tels mécanismes sont rarement examinées. La littérature est souvent restée statistique par essence, avec l'estimation de séries temporelles. L'accroissement des inégalités salariales rencontré par l'économie américaine entre les années 70 et 90 a ainsi été expliqué par Gottschalk and Moffitt (1994a) par une augmentation de la variance du composant transitoire des revenus. Mais elle reste silencieuse sur l'importance probable de rendements des capacités individuelles ou l'influence de l'exposition au risque. L'approche macroéconomique a tendance à considérer les processus de revenus comme exogènes alors qu'au niveau structurel, les agents peuvent adopter des comportements dynamiques d'épargne et de consommation (ou de choix de travail et de loisir) ; ce qui limite le niveau d'hétérogénéité considéré dans les modèles usuels.

Les mesures individuelles de l'aversion au risque nous permettent d'estimer un modèle en forme réduite des profils de revenus que nous décomposons en deux sources d'hétérogénéité persistante : des différences dans les attitudes face au risque et des différences dans une hétérogénéité résiduelle, assimilable à une erreur de mesure. Nous estimons la distribution de l'aversion au risque conditionnellement au trajectoires d'accumulation du capital humain et des conditions initiales sur le marché du travail. Dans cette contribution, nous accordons donc une attention particulière à l'hétérogénéité individuelle, notamment vis-à-vis du risque *ex ante*, avec pour objectif principal d'estimer l'importance de ces mécanismes dans les inégalités de revenus.

Les résultats montrent que la tolérance au risque joue un rôle déterminant dans la croissance des revenus : plus les individus aime le risque, plus le salaire croit vite, d'autant plus en début de carrière alors qu'au bout de plusieurs dizaines d'année, l'influence s'estompe. Selon différentes estimations, la tolérance vis-à-vis du risque offre un rendement de l'ordre de 2% sur les dix à vingt premières années de profil de revenus et que c'est surtout en milieu de carrière que les comportements vis-à-vis du risque expliquent une grande partie de la variance des salaires.

Chapitre 1

Specialization during Education : Impact of Specialization for New Comers in the Labor Market[1]

This paper studies and quantifies the returns to specialization defined as the schooling choice of a specific field. We consider two different dimensions of education : a quantitative one, the number of accumulated years of schooling, and a qualitative one, the field of study. In addition to the estimation of returns to specialization, the timing of specialization during education is also taken into account.
To understand the impact of early and late specialization, we build a dynamic discrete choice model allowing for heterogeneity in returns to human capital : individuals are supposed to be forward looking agents, making sequential schooling decisions and then facing a labor market specific to her field. Returns to schooling, job destruction probability and offer arrival rates are measured through the labor market structure and differ by specialization.
We use French panel data *Génération 98* , with more than 6500 men over 7 years. All these individuals exit school at the same date and then face the same macroeconomic context.
Our results underline that the choice of the specialization have a large impact on individual labor market trajectories with a particular role of utility of schooling. Our specifications allow us to capture heterogeneity in both dimensions of schooling choices, especially distinct returns of schooling.

[1]This chapter was written with Maxime Tô (Crest and Sciences Po)

Introduction

Education is usually defined as the number of years of schooling. Returns to education are then measured as the increase in earnings obtained after spending an additional year at school instead of entering the labor market.[2] This simple way of measuring education may be too restrictive. The large diversity of schools and curricula may bring different skills that are likely to be rewarded differently by the labor market. As a consequence, returns to education will differ : two individuals with the same number of years of schooling may actually have distinct labor income.

Furthermore, a recurrent debate about vocational education raises the question of specialization during education. Some degrees provide skills that are specific to particular occupations on the labor market. And then, specialized individuals potentially face different job opportunities. Although specific skills may respond to the needs of some industries or firms, policy makers might be interested in knowing the impact of early or late schooling specialization on future trajectories.

The aim of this paper is to describe the differences in labor market trajectories of people acquiring different skills during their school path. Thus, we argue that education has at least two dimensions : a quantitative one - the number of years of schooling - and a qualitative one that characterizes the professional knowledge or skills acquired at school that is called *specialization*.

In the early years of the career path, specialization is often characterized by vocational education. Focusing on this aspect of education Adda, Dustmann, Meghir, and Robin (2013) show its importance on future trajectories.[3] Skilled and non skilled workers in Germany have different wage profiles and job opportunities : although skilled workers experience a fast wage increase at the beginning of their career, they face lower destruction rates and receive less job opportunities which may prevent

[2] Card (1999), Card (2001) and Belzil (2007) survey different contributions of the literature on the estimation of returns to education.

[3] One can also be concerned about the returns of the choice of different curricula during high school. However, as detailed by Altonji, Blom, and Meghir (2012), the returns to curricula at early ages are week compared to the impact of vocational education.

them from finding new jobs after a negative employment shock. In post secondary education, specialization mainly corresponds to the choice of a major. For higher level of schooling, Arcidiacono (2004) and Arcidiacono, Hotz, and Kang (2012) underline that major choice may have a large impact on future individual earnings.

Finally, Hanushek, Woessmann, and Zhang (2011) question the timing of the specialization. Considering the trade-off between immediate productivity and future adaptability, they compare different countries where the age of specialization differs. They show that later specialization is associated to worse employment outcomes but that this initial penalty lowers with time. Focusing on this same question, Malamud (2010) compares England and Scotland who have different educational systems in terms of specialization. He shows that the timing in accumulation of field specific skills has important implications for career paths : switching to occupations unrelated to the field of studies lowers wages and the cost of switching is much higher for English workers who specialize earlier than Scottish ones.

Altogether, these results confirm the intuition that individuals face different labor market conditions according to the skills acquired during education and the timing of the acquisition. In particular, returns to education, offer arrival rates and probability of job destruction may differ according to the educational path of individuals.

In this paper, we propose a model of individual labor market trajectories and schooling choices with a specialization dimension. The key issue in order to measure the differences in job opportunities and returns to specialization is the endogeneity of schooling decisions. We build a dynamic discrete choice model of schooling and career choices *à la* Keane and Wolpin (1997) : individuals first make sequential schooling decisions choosing a level of schooling and a specialization ; then they leave school once and for all and enter the labor market. At this second stage, people face job offers, decide to accept or reject it and may lose their job while working.

This dynamic discrete choice structure allows us to take into account the actualized values of alternatives in individual decisions and is relevant to identify the importance of the timing for optimal individual decisions. Moreover, unobserved heterogeneity is added to the model in order to solve for endogeneity issues linked to the fact that individuals base their choice on future expected earnings.[4] Following Belzil and Hansen (2002), each discrete type of unobserved heterogeneity has a schooling ability and a labor market ability and individuals belonging to this type base his decisions on these two elements. This allows us to identify returns to specialization by accounting for potential correlation between both dimensions of unobserved heterogeneity.

To secure identification, we use local capacity constraints of schools as exclusion variables that only influence schooling decisions but are supposed to have no direct impact on future earnings. These variables are considered as exogenous from the individual point of view : in the presence of moving costs, individual decisions are guided by local schooling opportunities, measured by the number of available positions for each specialization and degree.

The model is estimated on the *Generation 98* survey, a French panel data of young people leaving school in 1998. We measure specialization from the observed field of studies or the domain of the highest diploma. We find large heterogeneity in labor market outputs according to the degree of individuals. In particular, we find that the choice of the specialization have a large impact on individual labor market trajectories. This heterogeneity comes from direct returns of education but also from differences in subsequent returns to labor market experience and costs of schooling faced by individuals.

The next section details the model and identification strategy. Section 1.2 presents the data. Then, section 1.3 analyzes the empirical results. The last section concludes.

[4] Arcidiacono (2004), Arcidiacono, Hotz, and Kang (2012), Beffy, Fougère, and Maurel (2012) in the case of major choice, or by Carneiro, Heckman, and Vytlacil (2011) for the decision to attend college.

1.1 Model

Our goal is to measure differences in labor market trajectories taking into account the endogeneity of education decisions. Therefore, we build a two-stage model. The first part explains education decisions and the second one describes labor market decisions. The two stages are sequential : once they entered the labor market, individuals cannot change their education level or specialization by going back to school. This feature of the model significantly simplifies the resolution of the model and is empirically supported by the fact that in our data less than 1% of individuals actually go back to school after entering the labor market.

During their schooling path, individuals are supposed to make sequential discrete decisions. At each level of education, they choose to enter the labor market or to continue to the next level of schooling. When continuing in the schooling system, individuals have to choose to specialize toward a particular set of occupations or not. We consider three types of schooling opportunities : no specialization, specialization toward production occupations and specialization toward service occupations.

Thus, during the schooling stage, an individual state is characterized by a position in the path in terms of level and specialization and he faces four alternatives : three specialization alternatives at the next level of the schooling system and one labor market alternative.

During the second stage, individuals face a labor market that is specific to the final level and chosen specialization. At each period, they receive job offers and decide to accept or reject it. When employed, positions can also be destroyed at a rate that is specialization specific.

Individuals are supposed to be forward looking in the sense that they take into account the future impact of their present decisions. This allows us to link the two stages of the model : the value of an educational degree will depend on its labor market value.

The two stages are also related through individual heterogeneity linking unobserved cost of schooling to unobserved ability on the labor market. Letting these two dimensions of unobserved heterogeneity to be correlated allows us to explicitly model the correlation between education and labor market and then to solve for the endogeneity problem.

In order to formalize this model, let Ω_t be the state space at time t, including state variables and random draws, and d_t^j an indicator variable for choice j at time t. The Bellman equations are :

$$\begin{aligned} V_t(\Omega_t) &= \max_{1 \le j \le K} V_t^j(\Omega_t) \\ V_t^j(\Omega_t) &= u_t^j + \beta \mathrm{E}[V_{t+1}(\Omega_{t+1}) | d_t^j = 1, \Omega_t] \end{aligned}$$

where β is the discount factor and u_t^j the direct utility of choice j. To avoid notation burden, we make the conditioning in the $\mathrm{E}\max$ term implicit and we omit individual subscripts. Given initial characteristics Ω_0, the set Ω_t can be written in three parts : state variables at time t, ω_t, random draws at time t, $\underline{\omega_t}$, and past state space : $\Omega_t = \Omega_{t-1} \cup \omega_t \cup \underline{\omega_t}$.

We distinguish the value functions for schooling decision $VS_t^j(\Omega_t)$ from value functions for the second stage of the model $VL_t^j(\Omega_t)$.[5]

1.1.1 Schooling decisions

In the schooling stage, individuals choose a level of schooling and a specialization maximizing their utility. At the end of her schooling path, an individual is characterized by a level of schooling d and a specialization k. We distinguish six different levels of schooling :

- $d = 1$: Junior high school : most of children have age 16 which corresponds to

[5]A summary of notations and details on the derivation of value functions and the likelihood are given in Appendix 1.A.

the maximum of compulsory age of schooling in France. These individuals are high school drop outs.

- $d = 2$: First year of High school or Short professional track that ended with a qualification (*BEP/CAP*).
- $d = 3$: High school diploma (*Baccalauréat*)
- $d = 4$: Early higher education : Community college (BTS, DUT) or University first year
- $d = 5$: Bachelor degree
- $d = 6$: Higher degrees (Master/PhD degree and French *Grandes Écoles*)

And as stated before, three specializations are available at each level of schooling : no specialization, specialization toward production occupation and specialization toward service occupation.

At each state (d, k), individuals have the possibility to leave school and face the labor market corresponding to the pair (d, k). Alternatively, they have the possibility continue at school to the next schooling level and choose a new specialization. We denote by $\mathcal{C}_{d,k}$, the set of possible choices at state (d,k). $\mathcal{C}_{d,k} = \{(d+1,1);...;(d+1,K);LM\}$ where LM is the labor market alternative with level d and specialization k. At the last level of schooling, $d = 6$ individuals necessarily enter the labor market.

Transition cost from level d to level d' is denoted by $c^d_{d',k'}(W_{d',k'},\eta^d_{d'})$ where $W_{d',k'}$ is a cost shifter of schooling and $\eta^d_{d'}$ is a normally distributed random shock. It is important to acknowledge that French public schools are free and that there are no fee at all level of schooling. Thus, we do not include additional covariates is the cost specification which we assume to be linear. These costs are written :

$$c^d_{d',k'} = \beta_{d',k'} + \varphi_{d'} \cdot W_{d',k'} + \eta^d_{d',k'}$$

Prior the last period of schooling, the value function of choice (d', k') conditional on being at state d :

$$VS^{d}_{d',k'} = c^{d}_{d',k'}(W_{d',k'}, \eta^{d}_{d'}) + \beta \mathrm{E}[\max_{j \in \mathcal{C}_{d',k'}} V^{j}(\Omega_{t+1})] \qquad \forall d : 1 \leq d < D$$

where V^j can be value functions for schooling VS^j or labor market values VL^j, defined below ; the expectation is taken on future realizations of random draws. The flows $c^{d}_{d',k'}$ can be interpreted as utility of schooling, usually reported as *costs of schooling* or *psychic costs* in the literature.

1.1.2 Labor Market Alternatives

At each state of the schooling path, individual can choose to enter the labor market. The value of working depends on both his level of schooling and his specialization.

While in the schooling system, the labor market value is given by :

$$VL_t(d,k) = \lambda_0(d,k) \times VE_t(d,k) + (1 - \lambda_0(d,k)) \times VU_t(d,k)$$

wher $VE_t(d,k)$ is the intertemporal value of being employed, $VU_t(d,k)$ the intertemporal value of being unemployed at time t and $\lambda_0(d,k)$ the probability to receive an offer while not working. Thus individuals do not receive automatically a job offer and then can be unemployed.

1.1.2.1 Reward functions

Intertemporal values of being employed and unemployed depend on the potential flows of utility characterized by the rewards associated to the decision to accept or not an offer.

The reward functions of employment for an individual with schooling level d and specialization k is given by :

$$R_t^E(d,k,X_t) \quad = \quad g(d,k,X_t) \times e^{\varepsilon_t(k)}$$

where $\varepsilon_t(k)$ is the error term which distribution depends of the specialization and X_t denote the working experience. The g function has the following shape :

$$g(d,k,X_t) = \exp\left(\alpha_0^k + \alpha_1^k X_t + \alpha_2^k X_t^2 + \sum_{s=1}^{D} \alpha_{3d}^k 1(s = d)\right)$$

X_t is thus a state variable. Its evolution is deterministic given the decision to work $(d_t^E = 1)$: $X_{t+1} = X_t + d_t^E$.

To allow for more flexible on-the-job wage evolution we model error terms as autoregressive processes (see Adda, Dustmann, Meghir, and Robin (2013)). When entering a new job, unobserved heterogeneity of labor income is supposed to follow a normal distribution with variance parameter specific to the chosen specialization : $\varepsilon_t(k) \sim \mathcal{N}(0, \sigma_k)$. Then, while staying in the same job, individuals face shocks that are added to the previous term. Thus we have :

$$\varepsilon_t^m(k) = \varepsilon_{t-1}^m(k) + e_t(k)$$

The innovation term $e_t(k)$ follow a normal distribution : $e_t(k) \sim \mathcal{N}(0, \sigma_v^k)$. ε are drawn when the individual starts a new job, this is a match specific draw and the e is interpretable as a productivity shock at each period.

On the other hand, the unemployment reward function is specified as :

$$R_t^U(d,k) = b_0 + .6 \cdot d_{t-1}^E \cdot R_{t-1}^E(d,k) + u_t$$

where b_0, can be interpreted as the basic level of unemployment benefit or home production, and u_t is normally distributed with variance σ_u. In accordance with administrative rules, unemployment benefits corresponds to 60% of labor income of the previous period. Individuals being eligible to unemployment benefit if they worked at least 6 months and for a limited time period, we add this term in the utility function only if the individual was working at the previous period (each period lasts six months).

1.1.2.2 Value functions

At each period, for individual characterized by (d, k), the timing is the following :

1. *Random draws.* Individual receive offers :
   (a) Match are destroyed with probability $\delta(d, k)$
   (b) If they are working and job is not destroyed :
      i. Wage evolves through experience accumulation and productivity schocks.
      ii. With probability $\lambda_1(d, k)$, individual receive external offers with associated wage's error term $\varepsilon*$.
   (c) If they are unemployed, individuals receive offers with probability $\lambda_0(d, k)$ and associated wage's error term $\varepsilon*$.
2. Individual compare values of all available positions and choose whether or not to work maximizing their actualized lifetime utility.

From this timing, we derive the value functions of working and being unemployed depending on the instantaneous reward functions R_t^E and R_t^U given above. Making the dependence on d an k implicit, we have :

$$
\begin{aligned}
VE_t &= R_t^E \\
&+\beta\{(1-\delta)\cdot\lambda_1\cdot \mathrm{E}[\max(VE_{t+1}, VE^*_{t+1}, VU_{t+1})] \\
&+(1-\delta)\cdot(1-\lambda_1)\cdot \mathrm{E}[\max(VE_{t+1}, VU_{t+1})] \\
&+\delta\cdot VU_{t+1}\}
\end{aligned}
$$

$$
VU_t = R_t^U + \beta\{\lambda_0\cdot \mathrm{E}[\max(VE^*_{t+1}, VU_{t+1})] + (1-\lambda_0)\cdot VU_{t+1}\}
$$

More precisely, we know from the data if the end of a job is due to resignation, end of contract with or without new offer, accepted or not. The different possibilities are detailed in appendix with all possible transitions of the likelihood.

Individuals are supposed to stay on the labor market until retirement. However, we do not solve for the whole dynamic program from this final period T but we set a value for a pseudo terminal period T^* which is the last period observed in our data. The values of all possible states (d, k, X_t) at these pseudo terminal period are estimated from the labor force survey as the actualized aggregate sum of labor income. An alternative method is given by Keane and Wolpin (2001) who specify the pseudo terminal values at period T^* as a flexible function of the state variable and estimate the parameters of this function joint with the model.

1.1.3 Identification Issues

We estimate this model by maximum likelihood methods. Individual contributions are written for each transition in both stages of our model. Overall, usual parametric specification of utility and error terms leads to an explicit expression of the likelihood. However, given the discrete choice structure, several normalization are necessary.

Following the identification results of Magnac and Thesmar (2002b) about discrete choice models, we fix the discount factor β to 0.95.

Schooling decisions rely on two main elements : on the one hand, individuals take into account the expected present values of potential future payoffs ; on the other hand, they consider the opportunity cost of being in school rather than entering the labor market. Endogeneity issues of schooling decision clearly appears if there is any correlation between unobserved heterogeneity of schooling acquisition costs and unobserved ability in the labor market.

In order to solve for this issue, we define discrete types of heterogeneity following Heckman and Singer (1984). As in Belzil and Hansen (2002), this allows us to link unobserved heterogeneity in schooling decisions and labor market outputs. For each type m of the M types of unobserved heterogeneity, we define p_m the probability to belong to this type. Unobserved heterogeneity is supposed to be two dimensional : the first dimension is schooling ability (f_φ^m) and impacts the transition from one level of schooling to the next one. The second dimension is ability on the labor market (f_α^m).

In order to allow the unobserve heterogeneity to affect the different transitions in different way, our specification uses factor loadings for intercepts of each equation. On the one hand, schooling ability is fixed for a given type but influences all schooling transitions in a different way. Thus, the costs of schooling intercept is $\beta_{d,k,m} = f_\beta^m \cdot \beta_{d,k}$. On the other hand, for each specialization k, the wages profile intercept of type m is $\alpha_0^{k,m} = f_\alpha^m \cdot \alpha_0^k$. . The use of such additional structure also implies several normalizations. In particular, we fix the vector (f_β^m, f_α^m) of one Heckman-Singer type to one.[6]

As pointed out by Kasahara and Shimotsu (2009) identification of unobserved heterogeneity dispersion using discrete types also depends on the support and variability of covariates in the model. In our case, in order to secure identification, we consider local capacity constraints of schools as cost shifters. These variables are

[6]This is equivalent to normalize one factor loading of one for each latent

continuous and are supposed to only affect schooling decisions. According to their local context, individuals face different opportunities in terms of schooling which is supposed to be exogenous.

Finally, the specification of discrete types of unobserved heterogeneity and the use of exclusion variables allows us to account for endogeneity of schooling decision and identify the timing of optimal choices and returns to specialization.

1.2 Data

1.2.1 The *Génération 98* Panel Data

The *Génération 98* survey is a representative survey of 16 000 young people leaving the French schooling system for the first time in 1998. This large scale survey is conducted by the Céreq.[7] After leaving school, individuals are followed during 7 years reporting the different steps of their working career during three retrospective interviews in 2001, 2003 and 2005.[8] *Génération 98* has the particular advantage to document many aspects of early labor market transitions and contains individuals facing the same labor market conditions. We restrict the sample to men who do not work in the public sector. Our final sample is composed by 5.350 individual trajectories.

1.2.1.1 Education decisions

During the first interview, individuals detailed their education trajectory which allow us to rebuild the sequence of their schooling decisions. As previously described, levels of schooling correspond to keep moments in the schooling path : end of junior high

[7]French Center for Research on Education, Training and Employment.

[8]About 742 000 individuals left the schooling system in France in 1998. In the first wave in 2001, 54 000 young people were interviewed from whom 33 000 were selected to enter the panel data. In 2003, 22 000 out of the 33 000 were asked and at the end, about 16 000 individuals are in the selected subsample. If we look at the probability to be surveyed at the three dates, they are a little less unemployed and more educated.

school (d = 1), first year of high school and short professional tracks (d = 2), high school degree (d = 3), early higher education (d = 4), bachelor degree (d = 5) and master degrees (d = 6).

Specialization is observed in the data as the field of education or major. It indicates whether individuals acquire knowledge that is more likely to be used for some occupations. We distinguish three categories : no specialization (k = 1), specialization toward production occupations (k = 2) and specialization toward services sector (k = 3).

More concretely, individual not specialized (k = 1) corresponds to a large range of disciplines, like sciences or humanities but education in these fields is not applied to specific occupations. In this respect, we define these individuals as not specialized. Individuals specialized in production occupation (k = 2), acquire techniques that are supposed to be applied in the production sector (Processing jobs, Mechanics, engineering, etc.). Finally, specialization in services represent degrees that give practical knowledge in services (Management, Trade, Communication, etc.). Given that high school drop-outs have no degree at all and may face a very particular labor market, we consider them as a particular case (k = 0).

Except for high-school drop-outs, all specializations are available at all levels. However, individual cannot enter the labor market with all specialization and all transitions between two levels of schooling from one specialization to another are not observed. The scheme in figure 1.1 presents the observed transition in the data. We exclude unobserved transitions from the model.

Table 1.1 describes the observed distribution of education. 16.2% of individuals are high school drop outs and about 40% have less than a high school degree. About 24% of the population leaves school with a high school degree. Among the people who go to college, most of them enter the labor market with a community college degree[9] (17% of the total population).

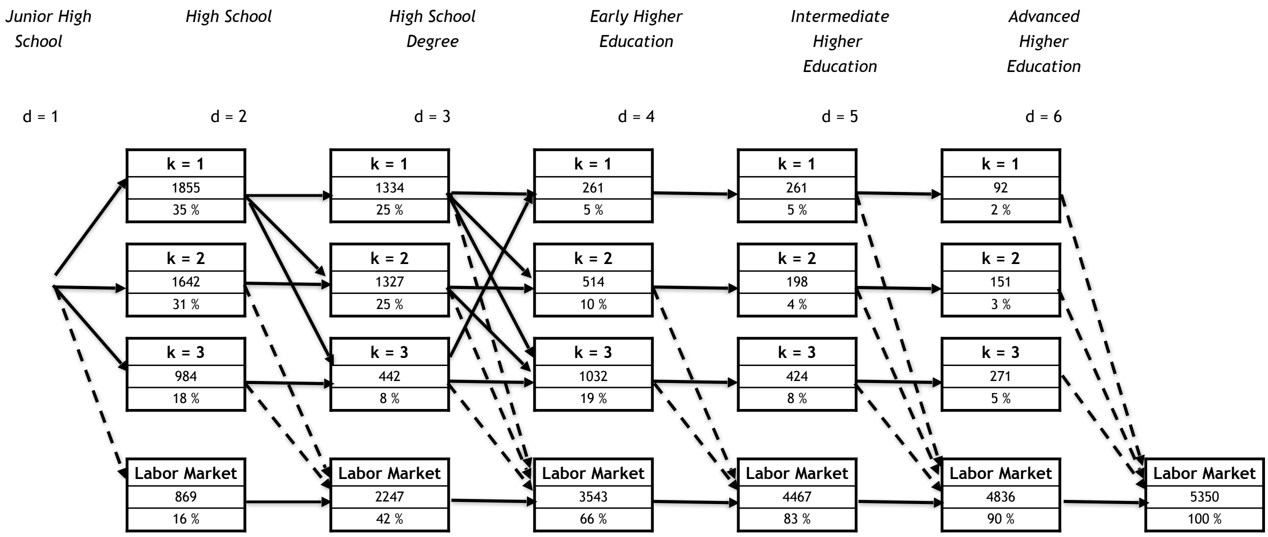
[9]BTS/DUT in the French schooling system.

TABLE 1.1 – Distribution of Final Level of Education and Specialization

	k = 0	k = 1	k = 2	k = 3	Total
d = 1	**869** 16.24 100 100				**869** 16.24 100
d = 2			**1145** 21.4 83.09 39.94	**233** 4.36 16.91 20	**1378** 25.76 100
d = 3		**188** 3.51 14.51 41.87	**865** 16.17 66.74 30.17	**243** 4.54 18.75 20.86	**1296** 24.22 100
d = 4			**548** 10.24 59.31 19.11	**376** 7.03 40.69 32.27	**924** 17.27 100
d = 5		**169** 3.16 45.8 37.64	**67** 1.25 18.16 2.34	**133** 2.49 36.04 11.42	**369** 6.9 100
d = 6		**92** 1.72 17.9 20.49	**242** 4.52 47.08 8.44	**180** 3.36 35.02 15.45	**514** 9.6 100
Total	**869** 16.24 100	**449** 8.39 100	**2867** 53.59 100	**1165** 21.78 100	**5350** 100 NA

Note : Source : *Enquête Génération 98*. The columns (k) correspond to the specialization, the rows corresponds to the level of education. Each cell contains the frequency, the cell percentage, the row percentage and the column percentage

FIGURE 1.1 – Aggregate Choices Tree (Levels d and Specializations k)



The distribution in terms of final specialization is very unbalanced : 8% of the total population have no particular specialization, 53% have a specialization toward production occupations and 21 % a specialization in services. Moreover, the balance differs according to the level of schooling considered. In particular, one can observe that individuals with bachelor degree are much less likely have a specialization toward production occupations.

1.2.1.2 Labor Market transitions

Labor market situations of individuals are reported through a monthly retrospective calendar in the survey. For each employment spell, wages are reported at the beginning and at the end of the spell. Additional information on wages is recorded at the moment of the three interviews. For each individual, we reduce the data to biannual observations and use interpolation to predict wages at the corresponding date if the information is not observed at this precise moment.

Table 1.2 provides information on labor market transitions observed from our biannual dataset. We observed 79 585 transitions during the seven first years of labor market histories. The large majority of transition are people who stay in the job (65%) and people who remain unemployed between two periods (9%). We still

TABLE 1.2 – Observed Labor Market Transitions

	Freq.	Perc.	$d \leq 3$ (%)	$d \geq 3$ (%)
Stays Unemployed	9208	11.57	12.06	10.61
From Unempl. to Empl.	7360	9.25	9.46	8.84
Unempl. after resignation	693	0.87	0.97	0.67
Unempl. after destruction	1734	2.18	2.44	1.67
Stays in the same job	54848	68.92	67.19	72.32
Changes job	5742	7.21	7.88	5.89
Total	79585	100	100	100

Note : Source : *Enquête Génération 98.*

TABLE 1.3 – Observed Labor Market Transitions by Specialization

	$k = 1$ (%)	$k = 2$ (%)	$k = 3$ (%)
Stays Unemployed	17.23	8.81	10.63
From Unempl. to Empl.	9.09	9.17	9.03
Unempl. after resignation	0.84	0.81	0.9
Unempl. after destruction	2.16	1.96	1.86
Stays in the same job	65.04	71.78	70.72
Changes job	5.64	7.47	6.87
Total	100	100	100

Note : Source : *Enquête Génération 98.*

observe a large number of mobility : more than 2 000 transitions from employment to unemployment, 7 360 from unemployment to employment and 5 742 changes of job.

1.2.2 Computing Capacity Constraints of Schools

Part of the identification of the model relies in the use of exclusion variables that influence the schooling decisions. We use proxies of local capacity constraints of schools in terms of both levels and specializations. To approximate these capacities, we use exhaustive French administrative data giving the number of individuals registered in each level of schooling and specialization. This information being available from 1994 to 2002, we approximate the local capacity constraints of education by averaging the figures on this period.

The average is taken at several geographical levels according to the level of schooling we consider. For levels lower than high school ($d \leq 3$), we average the data at the employment zone level.[10] For levels higher than high school, we aggregate the data at the Department level.[11] By doing so, we consider that individuals first take into account the very local possibilities of schooling for low levels of schooling and then a broader area at subsequent levels.

Given that we only observe the location of people at 6^{th} grade and at the end of schooling, we attribute to each individual the information related to its 6^{th} grade location which we consider as predetermined with respect to his schooling decisions.

For each level of schooling, we proxy capacity constraints by the variables defined as the ratio between the number of individuals locally registered in the previous schooling level and the number of individual locally registered in the considered specialization at the next schooling level.

Denoting by N_d the number of individuals enrolled in level d in a given area and $N_{d,k}$ the number individuals enrolled in specialization k in this same level, we build the exclusion variables as follow. For transition from level d, we use $W_{d,k}$ the ratio between students in level $d1$ and student in the previous level $d-1$:

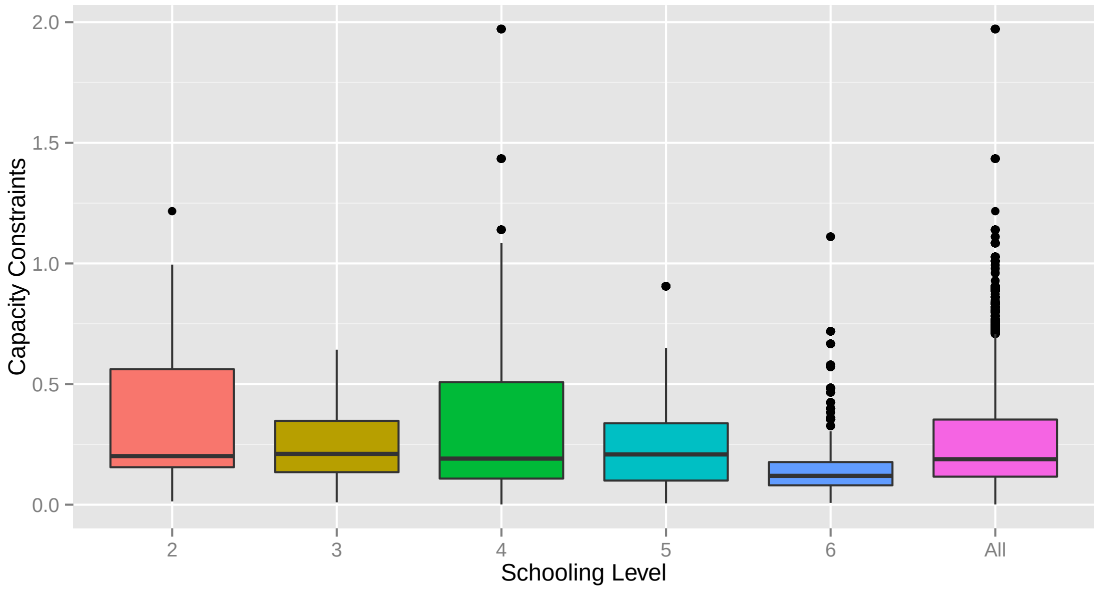
$$W_{d,k} = \frac{N_{d,k}}{N_{d-1}}$$

These variables are supposed to be exogenous in the model. This is equivalent to assume that there is no general equilibrium effect, that seems a fair assumption considering the short-term rigidity of local schooling capacity. These variables play a role of cost shifters in the utility of schooling in order to add individual heterogeneous exclusion restrictions.

[10]The "Employment Zone" is defined by Insee as "a geographical area within which most of the labor force lives and works".

[11]France is divided into 105 "Department" that are the suitable geographical areas to the post-secondary education. The median area of Departments is $5880km^2$.

FIGURE 1.2 – Distribution of Capacity Constraint by Level



Numerically the value of these variable are between 0 and 2. The overall average value is 0.25 and median is 0.189. Figures 1.2 and 1.3 describe the distribution of these variables by level and specialization. The dispersion appears to be bigger for early High School ($d = 2$) and early post-secondary education ($s = 4$). Capacity in general education ($k = 1$) is also more dispersed with a standard deviation of 0.24. In comparison, the standard deviation for production ($k = 2$) is 0.068 and 0.15 for services ($k = 3$).

In order to test for the relevance of the exclusion variables, we estimate a reduced form version of the schooling decision at each level of schooling. On the population of individuals who actually reached level $d - 1$ of schooling, we estimate a multinomial conditional logit model where the alternatives are the specialization at the next level of schooling. The utility of choosing specialization k at level d is given by :

$$U_{dk} = a_{dk} + b_d W_{d,k} + \eta_{dk}$$

One of the specialization correspond to the choice of labor market and is normalized to 0. The coefficients associated to the exclusion variables are thus constant by level

FIGURE 1.3 – Density of Capacity Constraints by Specialization

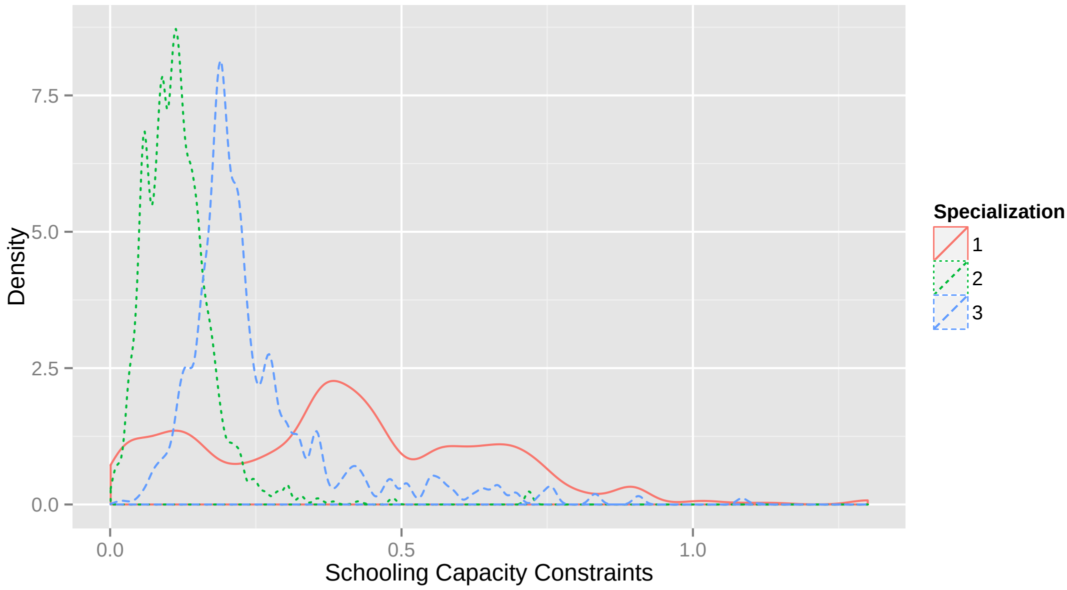


TABLE 1.4 – Reduced form estimation of schooling decisions

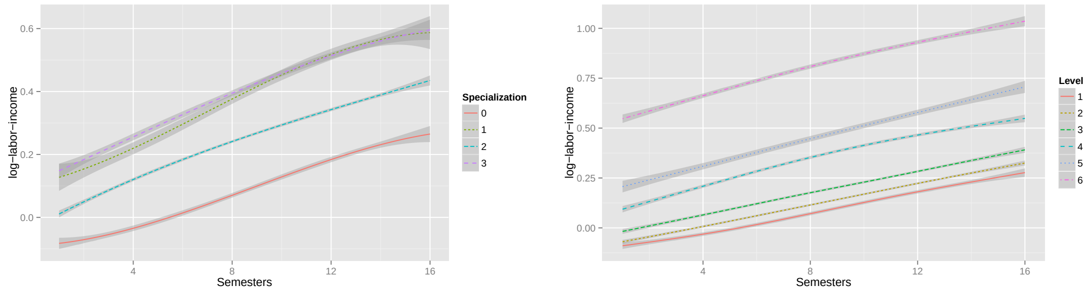
Level	Coef.	Std. Err.	Wald statistic	p-value
2	1.022558	.1861766	5.49	0.000
3	1.649873	.3466311	4.76	0.000
4	.3358804	.0996002	3.37	0.001
5	.7736837	.218844	3.54	0.000
6	.597257	.3740111	1.60	0.110

of schooling. Table 1.4 provides the estimates of these coefficients b_d for each of the five schooling decision.

From this table, we observe that the coefficients are highly significant at all levels of schooling except the last one. For this level of schooling, the coefficient is only significant at a 11% level. Although these estimations do not control for the expected labor market values of each alternative, it shows that our exclusion variable significantly influences individual decisions. It must also be noted that the sign of the coefficient is consistent with the fact that the more capacity there is in the specialization, the more likely is the individual to choose it.

FIGURE 1.4 – Average Log income and schooling levels

(a) Specialization (b) Level



Note : Source : *Enquête Génération 98*. Average trajectories were fitted by polynomial approximation

1.2.3 Descriptive Analysis of labor market trajectories

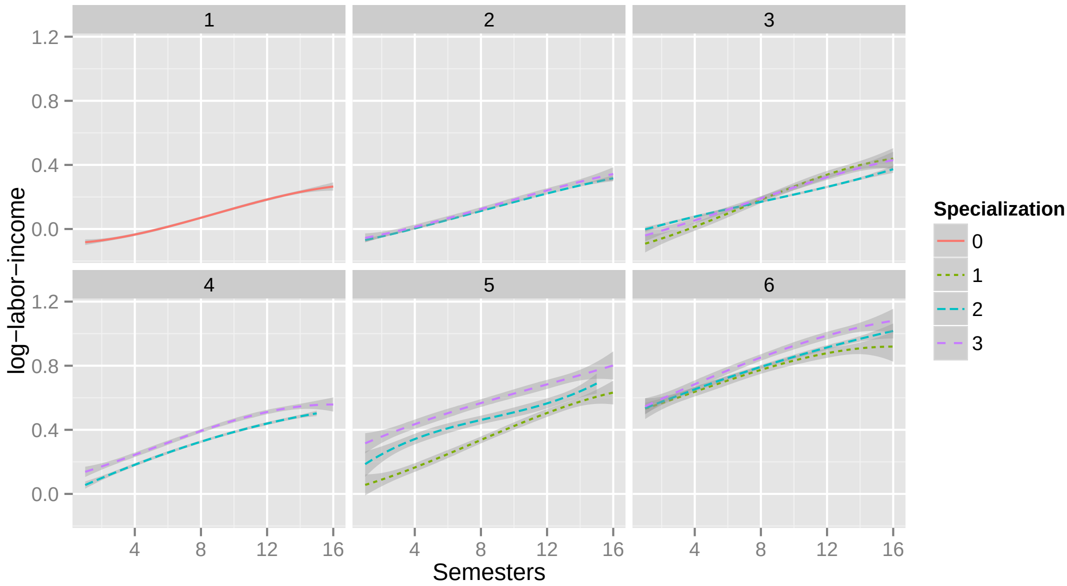
Focusing on the wage and employment trajectories, figures 1.4 to 1.7 provide details about the difference observed by level and specialization.

In terms of wages, we can observe on figures 1.4b that wage paths are almost parallel across levels of schooling : returns to experience are almost similar for the different levels of education. The parallelism is less visible but still exists when comparing the average paths for different specializations (Fig. 1.4a). Figure 1.5 takes into account the fact that specializations differs in terms of level composition. We observe in this figure that, for a given level of schooling, returns to experience clearly differ between specializations.

In particular, for individuals with high school degree (level $d = 3$), individuals specialized in the production occupations have higher initial income but their income increase much more slowly than individuals who have no specialization or specialized in services.

Simple panel regression analysis confirm these facts. Table 1.5 details fixed effect estimation of wages on quadratic experience and shows that there are important differences of returns to experience when considering separately the different populations of specialized people. Using a random effect estimate allows us to consider

Figure 1.5 – Average Log income by schooling levels and specialization



Note : Source : *Enquête Génération 98.* Average trajectories were fitted by polynomial approximation

the differences in returns to education by specialization (Table 1.5). Although the interpretation of the coefficient of these regression are not easy to interpret given that all specialization are not available at all levels of schooling, we observe from the differences between the intercepts that returns to high school degree (which is the reference level of schooling) differs between specialization as well as the returns to master degrees (and higher degrees). It appears that specialization is not beneficial for high school degree but actually bring a premium for higher degrees.

In terms of employment, transitions also differ a lot with respect to the level of schooling and the specialization. Specialized individuals have a much faster transition to employment than non-specialized although the unemployment rate of non-specialized converges to the rate of specialized ones (See figure 1.6a). This fact is also observed when considering separately high-school graduates and bachelor graduates. For higher degrees, the difference is less clear.

This analysis is descriptive and does not control for the endogeneity of education

Table 1.5 – Fixed effect estimation

	All	k = 0	k = 1	k = 2	k = 3
$X/10$	0.408 (0.00648)	0.334 (0.01844)	0.526 (0.02529)	0.374 (0.00827)	0.485 (0.01394)
$(X/10)^2$	−0.06 (0.00423)	−0.035 (0.01205)	−0.078 (0.01692)	−0.051 (0.00537)	−0.084 (0.00913)
R^2	0.411	0.309	0.477	0.408	0.408
N	55178	7844	4396	30814	12124

Note : Source : *Enquête Génération 98*. Estimates of the fixed effect estimation by specialization k.

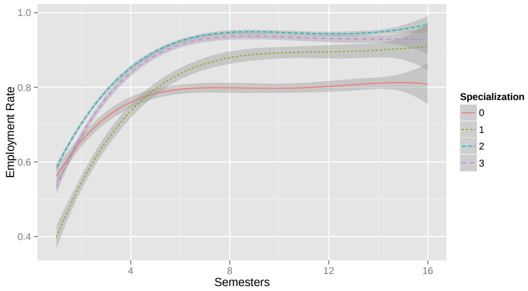
Table 1.6 – Random effect estimation

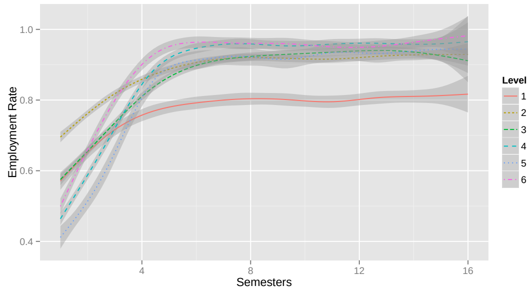
	All	k = 0	k = 1	k = 2	k = 3
Intercept	−0.055 (0.0065)	−0.121 (0.00898)	−0.101 (0.0231)	−0.048 (0.00701)	−0.069 (0.01735)
$X/10$	0.407 (0.00648)	0.331 (0.01841)	0.526 (0.02525)	0.374 (0.00826)	0.484 (0.01395)
$(X/10)^2$	−0.06 (0.00423)	−0.034 (0.01204)	−0.077 (0.0169)	−0.051 (0.00536)	−0.083 (0.00914)
d = 1	−0.1 (0.00992)				
d = 2	−0.076 (0.00864)			−0.069 (0.00866)	−0.081 (0.02415)
d = 3	*ref.*		*ref.*	*ref.*	*ref.*
d = 4	0.159 (0.0094)			0.149 (0.01022)	0.166 (0.02135)
d = 5	0.261 (0.01281)		0.145 (0.03098)	0.292 (0.0238)	0.348 (0.02757)
d = 6	0.607 (0.01128)		0.549 (0.03766)	0.608 (0.01348)	0.609 (0.025)
R^2	0.42	0.291	0.469	0.421	0.421
N	59841	8560	4791	33342	13145

Note : Source : *Enquête Génération 98*. Estimates of the random effect estimation by specialization k.

Figure 1.6 – Employment Fit

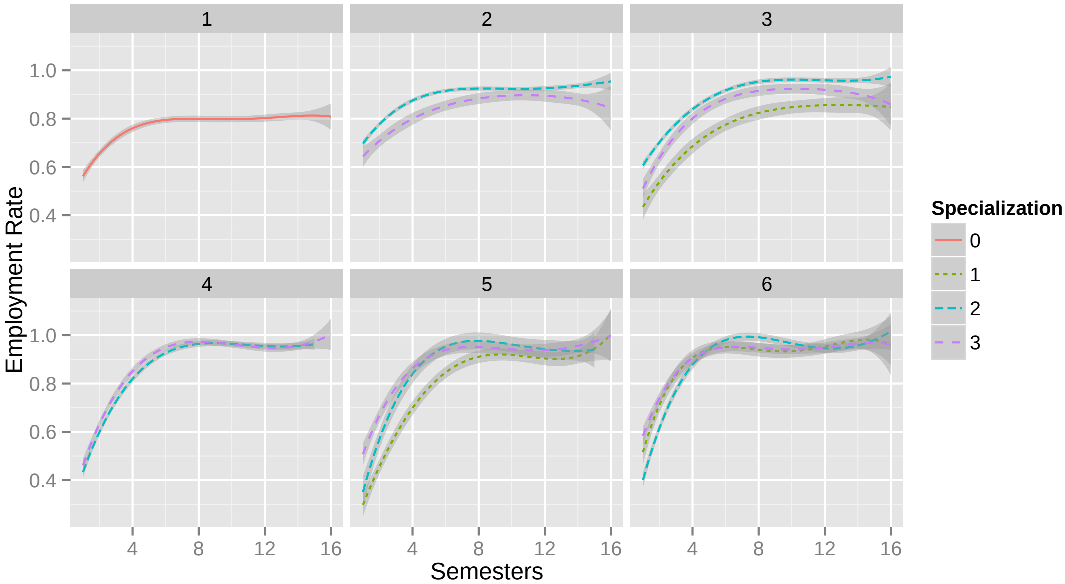
(a) Specialization (b) Level





Note : Source : *Enquête Génération 98*. Average trajectories were fitted by polynomial approximation

Figure 1.7 – Employment



Note : Source : *Enquête Génération 98*. Average trajectories were fitted by polynomial approximation decision.

1.3 Results

We estimate two distinct models. The benchmark model only considers the labor market, educational decisions are given and not endogeneous. The labor market values of education are then potentially biased. We compare these estimates with the results from the estimation of the full model previously described. In this last estimation, we use three types of unobserved heterogeneity[12]. The differences between the estimates of the two models allows us to assess the degree of endogeneity of educational choices.

1.3.1 Distinct Returns of Specializations

In order to summarize the differences between the two models, we present in tables 1.7 and 1.8 the labor market values of each possible combination of level/specialization.[13] Labor market values differ significantly between the two. In a general way, compared to the labor market drop out, values of all levels of education and specialization are much higher when we account for selection.

The hierarchy between degrees is not changed except in two cases. First, at the highest level of schooling, acquiring specific skills about service occupation appears to have bigger relatively returns than general education at this level : from 42% and 48 % respectivelly, it is estimated as 70% and 64 % taking into account selection. Second, for community college degrees ($d = 4$) the relative advantage for specialization toward production occupation ($k = 2$) and specialization toward services is changed when accounting for selection. Acquiring knowledge about production occupation becomes more advantageous and close to services specialization (resp. 0.20 and 0.21).

[12]We test estimation with 2 and 4 types of unobserved heterogeneity. Estimates with 3 and 4 types being close, we only display estimations with 3 types.

[13]These are actualized value functions at the beginning of labor market trajectory.

Table 1.7 – Labor Market Value of Education (% difference with respect to the labor value for high school drop out) : model with no types of unobserved heterogeneity

	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
$k = 1$		0.25		0.05	0.48
$k = 2$	0.02	0.12	0.20	0.16	0.42
$k = 3$	0.20	0.06	0.21	0.05	0.37

Note :*Source : Enquête Génération 1998.* Each value of this table corresponds to the percentage difference of the labor market value of a given level (d)/specialization (k) with respect to the value of high school drop outs. For instance, individual with high school degree (d = 3) and specialization in production occupations (k = 2) have on average labor market values that are 12% higher than high school drop out.

Table 1.8 – Labor Market Value of Education (% difference with respect to the labor value for high school drop out) : model with 3 types of unobserved heterogeneity

	$d = 2$	$d = 3$	$d = 4$	$d = 5$	$d = 6$
$k = 1$		0.32		0.18	0.64
$k = 2$	0.08	0.35	0.44	0.42	0.70
$k = 3$	0.38	0.19	0.38	0.18	0.65

Note :*Source : Enquête Génération 1998.* Each value of this table corresponds to the percentage difference of the labor market value of a given level (d)/specialization (k) with respect to the value of high school drop outs. For instance, individual with high school degree (d = 3) and specialization in production occupations (k = 2) have on average labor market values that are 35% higher than high school drop out.

Altogether, although this gross comparison of estimates does not constitute a proper test for endogeneity, accounting for the dynamic and intertemporality of decisions substantially modifies the estimates and appears to be necessary in the estimation of returns to education and specialization.

The table 1.16 in appendix and table 1.11 above give mincer coefficients estimates (respectively for benchmark model and for the full model). Because taking into account endogeneity leads to higher value functions, estimated returns of the benchmark model are higher : bias is downward and bigger for lower level of schooling and specialization $k = 3$. Estimated variances of idiosyncratic shocks are also higher in the full model. Services specialization $k = 3$ is optimal for highest level of schooling when general specialization is not available[14] and specialization $k = 2$ (production) has the highest returns of the two first levels.

[14]There is no general specialization for levels $d = 2$ and $d = 4$.

1.3.2 Costs of schooling

TABLE 1.9 – Costs parameters : impact of instruments (φ_d)

	Slope (ϕ_1)
d = 2	0.05291 (0.01696)
d = 3	0.13809 (0.03565)
d = 4	0.01245 (0.00863)
d = 5	−0.00977 (0.01052)
d = 6	0.05454 (0.02764)

Table 1.9 indicate that the coefficients associated to the capacity constraints variables are significant and positive for almost all levels.[15] The only parameter that is not significantly different from 0 is associated to the instrument in the cost of reaching level 5. Other parameters confirm the reduced form results presented above. Instruments have a positive impact on the utility of schooling meaning that the more capacity there is in a level/specialization, the more likely people are to continue schooling in this level/specialization.

Table 1.10 gives the average of utility of schooling for all individuals. Each figure of this table is computed as the mean over individuals who differ with respect to the capacity constraint they face. These numbers are monetary interpretable given at each level, schooling utilities are compared to the labor market values of the current state.

TABLE 1.10 – Average Individual Utility of Schooling

	d = 2	d = 3	d = 4	d = 5	d = 6
k = 1	-1.55	0.75	-0.64	-0.71	-0.65
k = 2	-0.48	-0.19	-0.28	-1.02	0.61
k = 3	-0.75	-0.62	0.11	-0.57	0.08

Utility of schooling is negative for almost all levels and specializations[16] : this is why we call them *costs*. This imply that the consumption values of schooling are

[15] Table 1.14 in appendix contains the estimates of intercepts for each specializations and types.

[16] Except the utility of reaching the high school degree with no specialization and the utility of obtaining a specialization at the last level of schooling.

negative, in addition to fees and other sources of psychic costs. The positive utility of finishing high school with no specialization may be explained by the fact that this degree, corresponding to the *Baccalauréat*, is one of the most attended in the French school system and is socially rewarded so that its consumption value is higher. The fact that costs of specialization are negative at the last level of education $d = 6$ can be explained by the fact that individuals often make a simultaneous choice of going to a master degree at the same time they start their bachelor degree so that monetary utility of schooling is positive (and close to 0 for services).

1.3.3 Labor market trajectories

From table 1.8 it appears that labor market values differ a lot according to the level and specialization.[17]

Compared to high school drop-outs (which is the reference group), returns to education are heterogeneous. The smallest value is for level $d = 2$ specialized in production occupations. Except for this level of education, individual specialized in production occupation have much higher labor market values conditional on the level of schooling. Specialization toward services leads to returns to education similar to the one of individuals who have no specialization.

The differences between labor market values can be analyzed from the detail of the parameters given in table 1.11. Direct returns to education at high level of education are more important for individual with no specialization and individual specialized in services. However, as summarized in figures 1.9 and 1.8, marginal returns to experience are much more important for individual specialized in production occupations.

Although returns are globally increasing with the level of schooling it is important to notice that labor market values are not monotonous and that specialization

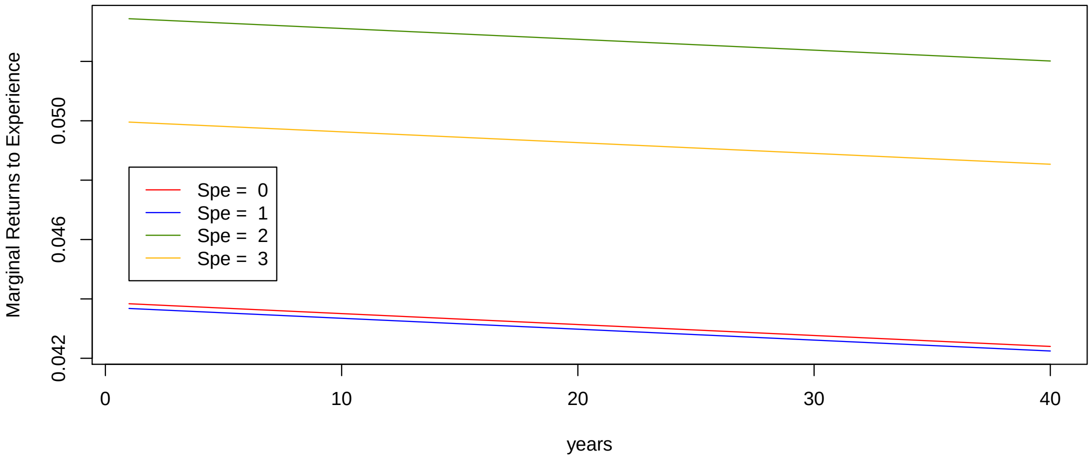
[17]Estimates for the hazard rate of job offers and job destructions are presented in appendix in tables 1.17, 1.18 and 1.19. Parameters of the wage equations are presented in table 1.11

TABLE 1.11 – Labor market parameters estimates

	k = 0	k = 1	k = 2	k = 3
α_0^1	−0.19621 (3e−04)	−0.19621 (3e−04)	−0.19621 (3e−04)	−0.19621 (3e−04)
α_0^2	0.41379 (0.01511)	0.41379 (0.01511)	0.41379 (0.01511)	0.41379 (0.01511)
α_0^3	1.1294 (0.09922)	1.1294 (0.09922)	1.1294 (0.09922)	1.1294 (0.09922)
α_x	0.08775 (0.00442)	0.08743 (0.00148)	0.10695 (0.00259)	0.09998 (0.00442)
α_{x^2}	−0.07375 (0.00385)	−0.07348 (0.00262)	−0.07306 (0.0023)	−0.07279 (0.00143)
$\alpha_{d=2}$	. (.)	. (.)	0.12746 (0.00988)	0.04249 (0.00894)
$\alpha_{d=3}$	. (.)	0.14148 (0.0227)	0.15315 (0.01062)	0.14104 (0.00881)
$\alpha_{d=4}$	. (.)	. (.)	0.16913 (0.01043)	0.19018 (0.01017)
$\alpha_{d=5}$	. (.)	0.161 (0.02418)	0.13054 (0.04424)	0.1154 (0.02488)
$\alpha_{d=6}$	. (.)	0.21987 (0.03149)	0.19344 (0.01969)	0.22251 (0.01469)
σ_e	1.23254 (0.18579)	0.65741 (0.01318)	0.82286 (0.12705)	0.78861 (0.04466)
σ_v	0.71018 (0.04313)	0.08775 (0.00787)	0.8784 (0.06162)	0.76443 (0.03898)

plays different roles at different levels. In particular it appears from the comparison between levels 2 to 5 that a well chosen specialization may increase more individual labor market values than spending more time at school.

FIGURE 1.8 – Marginal Return to experience



1.3.4 Ex-ante returns to schooling

Although estimates suggest a hierarchy of degrees in terms of labor market values, degrees with high values are not the most popular degrees. Costs often appear to be too important for individuals to continue schooling and to choose specializations. In order to account for costs in the computation of returns, we compute two versions of the internal return rates (IRR) following Heckman, Lochner, and Todd (2006b). Given that IRR are define to compare two available alternatives, we choose to focus on nodes on the schooling path where labor market alternative are available. At each of these nodes, we compare the labor market value of the individual to each of the available schooling paths. So, for an individual who reached level d with specialization k, we compare labor market values ($VL(d,k)$) to schooling values ($VS^{d}_{d',k'}$).

We compare two alternative measures of the IRR. The first one does not account for the direct utility of schooling and the second one does. In the first case, the only cost of schooling is thus the opportunity cost of not entering the labor market immediately.

The first measure is :

$$IRR_1(d,k,d',k') = \frac{\beta VL(d',k') - VL(d,k)}{VL(d,k)}$$

and the second :

$$IRR_2(d,k,d',k') = \frac{\beta VL(d',k') - (VL(d,k) - c^d_{d',k'})}{VL(d,k) - c^d_{d',k'}} = \frac{VS^d_{d',k'} - VL(d,k)}{VL(d,k) - c^d_{d',k'}}$$

Tables 1.12 and 1.13 give the average values of these internal rates of returns.

Table 1.12 – Internal return rate of education comparing labor market values for the degree (d,k) to the value of education (d',k') (IRR_2)

	$k'=1$	$k'=2$	$k'=3$
$d=1,\ k=1$	-0.01	-0.02	-0.23
$d=2,\ k=2$		0.20	
$d=2,\ k=3$			-0.32
$d=3,\ k=1$	-0.75	-0.69	-0.57
$d=3,\ k=2$		-0.69	-0.58
$d=3,\ k=3$	-0.73		-0.52
$d=4,\ k=2$		-0.13	
$d=4,\ k=3$			-0.13
$d=5,\ k=1$	-0.28		
$d=5,\ k=2$		1.38	
$d=5,\ k=3$			0.36

Table 1.13 – Internal return rate of education comparing labor market values for the degree (d,k) to the value of education (d',k') setting cost to 0 (IRR_1)

	$k'=1$	$k'=2$	$k'=3$
$d=1,\ k=1$	1.73	0.51	0.42
$d=2,\ k=2$		0.43	
$d=2,\ k=3$			0.03
$d=3,\ k=1$	-0.61	-0.61	-0.61
$d=3,\ k=2$		-0.62	-0.62
$d=3,\ k=3$	-0.57		-0.57
$d=4,\ k=2$		0.57	
$d=4,\ k=3$			0.28
$d=5,\ k=1$	0.12		
$d=5,\ k=2$		0.14	
$d=5,\ k=3$			0.25

In these tables, each line correspond to a level/specialization of schooling where the individual can choose to enter the labor market. Each value of the line corresponds to the IRR of the next level with the specialization indicated by the column. Given that all specialization are not available at all levels/specializations, many cells of these table are empty.

Generally, the IRR taking into account for costs are negative, which may be related to the fact that for most decisions individuals only a minority of individual choose any of the alternatives. However, one striking result from these table is the fact that the average returns from going from level 3 (high school degrees) to level 4 are negative even without taking direct costs into account. In addition to the fact that the labor market values of these degrees are high, the negativity of the internal return rates show the importance of both the opportunity cost of schooling and the direct cost. As a consequence, these results show that in order to increase the share of individual who undertake post secondary education, an important effort must be made in order to lower the cost of schooling at this level. Moreover, given that measurements of the IRR are closer to zero at subsequent levels of schooling, it appears that reducing the cost of these levels would increase the general of education by much more than one level.

Concluding Remarks

In this paper we propose and estimate a dynamic discrete choice model with unobserved heterogeneity that allow us to estimate the returns to specialization during education.

In order to identify the model, we use capacity constraints as exclusion variables that influence schooling decision but have no impact on future labor market earnings. The estimations show that these variables have a significant impact on decisions.

From the estimation of the value of labor market trajectory, we observe that acquiring skills to the production sector is more rewarded than other specialization, in particular if the specialization takes place after the high-school level. Specialization toward services has a larger impact prior high school graduation and an impact similar to the one of having no specialization from the high school level.

The difference between the labor market values of the trajectories comes from differences of direct rewards of the level of education across specialization and in a more important way, from differences in returns to experience in the labor market that are much larger for individual specialized toward production occupations.

From the analysis of internal rates of returns, we observe that both direct costs and opportunity costs play important roles that explain the fact that most valuable degrees may not be the most popular ones. In particular, there exists an important barrier to enter post-secondary education with any of the specializations considered.

1.A Summary of Notations

Variables	**Description**
$W_{d',k'}$	Local Supply of Education by level (d) and specialization (k)
X	Experience (in month)

Function	**Variables**	**Parameters**	**Error term**
Schooling Costs $c^d_{d',k'}(.)$ from d to (d',k')	$(1, W_{d',k'})$	$(\beta_{d',k'}, \phi_{d'})$	$\eta^d_{d',k'}$
Rewards $R^E_t(k)$	$(1, X, X^2, d)$	$(\alpha^k_0, \ldots, \alpha^k_{3D})$	$\varepsilon_t(k)$

Other Notation	Description
Θ	Full set of parameters
Ω_t	Information set at time t
β	Discount factor
k	Specialization
d or d'	Level of schooling (d when known and d' for future possibilities)
VU and VE	Value functions (resp. for employment and unemployment)
VS and VL	Value functions (resp. for schooling and labor market)
b_0, b_s	Unemployment benefits parameters
λ_0, λ_1 and δ	Offer arrival rate when unemployed and employed and destruction rate

1.B Model Likelihood

This section give more precisions about the estimation of our model and the explicit form of likelihood contributions. We first write the extended expression of the likelihood for each part of the model (labor market and schooling choices) and present in depth the resolution of the dynamic programming structure, i.e. how to compute the Value Functions and Emax terms.

The individual likelihood can be expressed as a product of conditional densities for each transition $k - k'$. We described schooling choices, trajectories to the labor market, series of wages if individuals are working and transitions to unemployment.

We use Heckman-Singer semi-parametric techniques with J points of support for unobserved heterogeneity. The full likelihood is :

$$\mathcal{L}(\Theta) = \prod_i \prod_{k-k'} \mathcal{L}_{k-k'}(\Theta) = \prod_i \int_j \mathcal{L}_S^j(\Theta_S) \cdot \mathcal{L}_L^j(\Theta_L) \mathrm{dF}_j$$

where j is the heterogeneity distribution, i individuals, S for the schooling model and L for the labor market. We stipulate a logit form for discrete heterogeneity types such that :

$$\mathcal{L}(\Theta) = \prod_i \sum_{j=1}^{J} \frac{e^{\theta_j}}{\sum_l e^{\theta_l}} \cdot \mathcal{L}_S^j(\Theta_S) \cdot \mathcal{L}_L^j(\Theta_L)$$

The error terms of our model are :

- η^k : speciality choice
- $\kappa_d^{d'}$: costs of schooling
- ε_t : wages shocks
- u_t : unemployment benefits

At a given period t and for a specific job m, $\varepsilon_t = \varepsilon_t^m + \varepsilon_t^e$ is the stochastic part of wage evolution. Following Adda et al., this wage error term is thus decomposed as a match-specific term ε_t^m and an idiosyncratic shock at each period ε_t^e :

- The value of the position-specific term ε_t^m takes the previous value of the error term if individual stays in the same job $\varepsilon_t^m = \varepsilon_{t-1}$ or is drawn in a random law $\varepsilon^* \sim \mathcal{N}(0, \sigma_m)$ if this is a new job (distribution of outside offers).
- Idiosyncratic shocks are drawn at each period and normally distributed $\varepsilon_t^e \sim \mathcal{N}(0, \sigma_e)$

We use e characters for realizations of random variables ε. The exact timing of realizations are explained in the text. In order to have readable contributions, we simplify the notation of value functions[18] :

- $VS_t^k = VS_t^k(\kappa) = S_t^k + \kappa$ value of schooling level k
- $VL_t^k = VL_t^k(\eta, e) = L_t^k(e) + \eta$ value of entering the labor market with speciality k
- $VU_t = VU_t(u_t)$ value of unemployment

[18] We only write in parenthesis error terms evolving in the current period. Generally speaking, value functions depend on the whole state space : $V_t(\Omega_t)$.

- $VE_t = VE_t(e_t)$ value of working (same job)
- $\widetilde{VE}_t = VE_t(e_t^*)$ value of working (new job)

For each transition, the individual contributions of the likelihood for period t are written below. We first derive all contributions in the labor market and then for schooling. We present the general formula in term of probabilities and give the formal expression based on our functional assumptions.

1.B.1 Labor market contributions

1.B.1.1 Timing and trajectories

Because ε_t is a continuous random variable, it is needed to discretize it in order to compute individual contributions of working trajectories. We take K points of support for this distribution so that the state space in the labor market is indexed by $k \in \{0, 1, \dots, K\}$: 0 stands for unemployment and working choices for $1 \leq k \leq K$. We denote VE_t^k and $\widetilde{VE_t^{k*}}$ the values of VE_t and $\widetilde{VE}_t$. All possible cases are written below (probabilities of transition are between parentheses) :

- If individual is unemployed (state 0) :
  - if no offer (with probability $1 - \lambda_0$), next state is 0.
  - if one offer (with probability λ_0) of a wage w_k :
    - Accept choice k if $VU_t < VE_t^k$, next state is k.
    - Reject choice k if $VU_t > VE_t^k$, next state is 0.
- If individual is employed (state k) :
  - if job position is destroyed (with probability δ), next state is 0.
  - if no destruction (with probability $(1 - \delta)$), on the job wage evolution :
    - If no offer (with probability $1 - \lambda_1$) :

- if $VU_t < VE_t^k$, same next state k.
- if $VU_t > VE_t^k$, next state is 0.

$*$ if new offer w_k^* (with probability λ_1) :

- if $\max(\widetilde{VE_t^{k^*}}, VU_t) < VE_t^k)$, next state is k
- if $\max(VU_t, VE_t^k) < \widetilde{VE_t^{k^*}})$, next state is k^*
- if $\max(\widetilde{VE_t^{k^*}}, VE_t^k) < VU_t)$, next state is 0

1.B.1.2 Emax Computations

Before expliciting the likelihood contributions, it is necessary to solve the dynamic programming problem and write the value functions. We denote by R_t and U_t the deterministic part of the rewards in the labor market (resp. while working and unemployed). To the end of this section, we simplify value functions notations to avoid notation burden :

$$\begin{aligned} VU_t &= V_t(0) \\ \forall k > 0, \quad VE_t^k &= V_t(k) \\ \forall k^* > 0, \quad \widetilde{VE_t^{k^*}} &= V_t(k^*) \end{aligned}$$

As said before, we use K points of support for match specific error term. The discretization is done with intervals of the random variable ε. We equally split its distribution so that $\varepsilon_k = E(\varepsilon | \varepsilon \in [q_k, q_{k+1}])$ with q_k and q_{k+1} two quantiles of the distribution of ε (if employed, realizations are written e_k of the same job or e_k^* for a new job). We mainly have two types of value functions :

$$\begin{aligned} V_t(0) &= U_t + u_t + \beta \cdot \mathrm{E}(\max_k V_{t+1}(k)) \\ V_t(l) &= R_t + e_t + \beta \cdot \mathrm{E}(\max_k V_{t+1}(k)) \end{aligned}$$

We detail below the Emax term in each of the two cases.

- **Unemployment :**

$$V_t(0) = U_t + u_t + \beta \cdot \mathrm{E}(\max_k V_{t+1}(k))$$

$$\begin{aligned} \text{with : } \mathrm{E}[\max V_{t+1}(k)] &= (1-\lambda_0)\cdot \mathrm{E}[V_t(0)] \\ &\quad +\lambda_0 \cdot \sum_{k>0} \mathrm{P}(\varepsilon_{t+1} = e_k)\cdot \mathrm{E}[\max\{V_{t+1}(0), V_{t+1}(k)\}|\varepsilon_{t+1} = e_k] \end{aligned}$$

In this expression, $\mathrm{E}[V_t(0)]$ is calculated with the next period value functions due to backward induction solution and $\mathrm{E}[u_t] = 0$. The other term $\mathrm{E}[\max\{V_{t+1}(0), V_{t+1}(k)\}|\varepsilon_{t+1} = e_k]$ equals (we omit $_{t+1}$ for the Vs) :

$$\begin{aligned} &= \mathrm{E}[V(0)|V(0) > V(k), \varepsilon_{t+1} = e_k]\cdot \mathrm{P}(V(0) > V(k)|\varepsilon_{t+1} = e_k) \\ &\quad +\mathrm{E}[V(k)|V(0) < V(k), \varepsilon_{t+1} = e_k]\cdot \mathrm{P}(V(0) < V(k)|\varepsilon_{t+1} = e_k) \\ &= (U + \mathrm{E}[u|u > R + e_k - U])\cdot \mathrm{P}(u > R + e_k - U) \\ &\quad +(R + e_k)\cdot \mathrm{P}(u < R + e_k - U) \end{aligned}$$

Noting $r_{t+1} = \frac{R_{t+1}+e_k-U_{t+1}}{\sigma_u}$, we have under normality assumption :

$$\begin{aligned} \mathrm{E}[u|u > R + e_k - U] &= \sigma_u \frac{\varphi(r)}{1-\Phi(r)} \\ \mathrm{P}(u > R + e_k - U) &= 1 - \Phi(r) \end{aligned}$$

Thus we have :

$$V_t(0) = U_t + u_t + \beta \times \{U_{t+1} + (R_{t+1} - U_{t+1} + e_k) \cdot \Phi(r) + \sigma_u \phi(r)\}$$

- **Employment :**

$$V_t(l) = R_t + e_l + \beta \cdot \mathrm{E}(\max_k V_{t+1}(k))$$

And without writing $_{t+1}$ subscripts for the Vs, we have :

$$\begin{aligned}
\mathrm{E}(\max_k V(k)) \quad = \quad & \delta \cdot \mathrm{E}[V(0)] \\
& +(1-\delta) \cdot \sum_{k>0} \mathrm{P}(\varepsilon_{t+1} = e_k | \varepsilon_t = e_l) \\
& \times\{(1-\lambda_1) \cdot \mathrm{E}[\max(V(0), V(k)) | \varepsilon_{t+1} = e_k] \\
& +\lambda_1 \cdot \sum_{k^*>0} \mathrm{P}(\varepsilon^*_{t+1} = e^*_{k^*}) \cdot \mathrm{E}[\max(V(k^*), V(k), V(0)) | \varepsilon_{t+1} = e_k, \varepsilon^*_{t+1} = e^*_{k^*}]\}
\end{aligned}$$

The first $\mathrm{E}\max$ is detailed before, two remaining elements need to be detailed. The evolution in the same job is (recall that $\varepsilon^m_{t+1} = \varepsilon_t$ in this case) :

$$\mathrm{P}(\varepsilon_{t+1} = e_k | \varepsilon_t = e_l) = \mathrm{P}(\varepsilon_{t+1} \in [q_k, q_{k+1}] | \varepsilon_t \in [q_l, q_{l+1}])$$

The new offer is just a draw of the previous wage in a normal with mean e_l. Then, the transition probability is :

$$\mathrm{P}(\varepsilon_{t+1} = e_k | \varepsilon_t = e_l) = \Phi\left(\frac{q_{k+1} - e_l}{\sigma_v}\right) - \Phi\left(\frac{q_k - e_l}{\sigma_v}\right)$$

Indeed, given that we only know that ε is in a interval, we have to integrate the previous probability over the values between q_l and q_{l+1} :

$$\mathrm{P}(\varepsilon_{t+1} \in [q_k, q_{k+1}] | \varepsilon_t \in [q_l, q_{l+1}]) = \int_{q_l}^{q_{l+1}} \mathrm{P}(\varepsilon_{t+1} \in [q_k, q_{k+1}] | \varepsilon_t = e_l) \mathrm{d}e_l$$

which is calculated by simulation (see subsection below). Note that simulation must be done only once, at the beginning of likelihood. The E max can then be written (the only remaining stochastic part is the error term of unemployment) :

$$\begin{aligned}\mathrm{E}[\max(V(k^*),V(k),V(0))] &= \mathrm{P}(\max(V(k^*),V(k))>V(0))\cdot\max(V(k^*),V(k)) \\ &\quad +\mathrm{P}(\max(V(k^*),V(k))<V(0))\cdot\mathrm{E}(V(0))\end{aligned}$$

- **Simulation of the transition probability :**

We calculate the transition probability by simulations :

$$\mathrm{P}(\varepsilon_{t+1}\in[q_k,q_{k+1}]|\varepsilon_t\in[q_l,q_{l+1}]) = \int_{q_l}^{q_{l+1}}\mathrm{P}(\varepsilon_{t+1}\in[q_k,q_{k+1}]|\varepsilon_t=e_l)\mathrm{d}e_l$$

We only need to draw values of ε_t that are draws in a normal distribution truncated between q_l and q_{l+1}. We can easily obtain these values by inverting the CDF of the truncated normal distribution :

$$F_\varepsilon(x) = \frac{\Phi\left(\frac{x}{\sigma_\varepsilon}\right)-\Phi\left(\frac{q_l}{\sigma_\varepsilon}\right)}{\Phi\left(\frac{q_{l+1}}{\sigma_\varepsilon}\right)-\Phi\left(\frac{q_l}{\sigma_\varepsilon}\right)}$$

Then F_ε^{-1} is given by :

$$F_\varepsilon^{-1}(u) = \sigma_\varepsilon\Phi^{-1}\left(u\left[\Phi\left(\frac{q_{l+1}}{\sigma_\varepsilon}\right)-\Phi\left(\frac{q_l}{\sigma_\varepsilon}\right)\right]+\Phi\left(\frac{q_l}{\sigma_\varepsilon}\right)\right)$$

So from a set of uniform simulations u, one can easily obtain a set of truncated normal simulations.

1.B.1.3 General Form of Likelihood Contributions

Value functions are solved by backward induction according to previous section. This paragraph gives likelihood contribution for each transition. Integrals are computed by simulations.

- **Transition from unemployment to work :**

$$\mathcal{L}_{U-E} = \lambda_0 \times \mathrm{P}(VE_t > VU_t) \times \mathrm{pdf}_\varepsilon(e_t)$$

$$L(0 \to k) = \lambda_0 \cdot \mathrm{P}(V(0) < V(k) | \varepsilon = e_k) \cdot \mathrm{P}(\varepsilon = e_k)$$

with $\mathrm{P}(\varepsilon = e_k) = \mathrm{P}(\varepsilon \in [q_k, q_{k+1}]) = \frac{1}{\sigma_k}\varphi\left(\frac{e_k}{\sigma_k}\right)$.

- **Transition from unemployment to unemployment :**

$$\mathcal{L}_{U-U} = 1 - \lambda_0 + \lambda_0 \cdot \mathrm{P}(VU_t > VE_t)$$

$$L(0 \to 0) \quad = \quad (1-\lambda_0) + \lambda_0 \cdot \int_\varepsilon \mathrm{P}(V(0) > V(k) | \varepsilon = e_k) de_k$$

- **Transition from work to unemployment :**

$$\mathcal{L}_{E-U} \quad = \quad \delta + (1-\delta) \times [\ (1-\lambda_1) \cdot \mathrm{P}(VU_t > VE_t) + \lambda_1 \cdot \mathrm{P}(VU_t > VE_t\,,\, VU_t > \widetilde{VE}_t)]$$

$$\begin{aligned} L(k \to 0) \quad &= \quad \delta + (1-\delta) \times \{ \int_\varepsilon (1-\lambda_1) \cdot \mathrm{P}(V(0) > V(l) | \varepsilon = e_l) de_l \\ &\quad + \lambda_1 \int_{(\varepsilon^*, \varepsilon)} \mathrm{P}(V(0) > \max(V(k^*), V(l)) | \varepsilon^* = e^*_{k^*}, \varepsilon = e_l) de_l de^*_{k^*} \} \end{aligned}$$

- **Transition from work to work (same job) :**

$$\begin{aligned}\mathcal{L}_{E-E} &= (1-\delta)\times[\,(1-\lambda_1)\cdot \mathrm{P}(VE_t > VU_t) \\ &+ \lambda_1\cdot \mathrm{P}(VE_t > VU_t\,,\, VE_t > \widetilde{VE}_t)]\times \mathrm{pdf}_{\varepsilon}(e_t)\end{aligned}$$

$$\begin{aligned}L(k\to k') &= (1-\delta)\cdot \mathrm{P}(\varepsilon_{t+1}=e_{k'}|\varepsilon_t=e_k)\times\{(1-\lambda_1)\cdot \mathrm{P}(V(k')>V(0)|\varepsilon_{t+1}=e_{k'}) \\ &+\lambda_1\cdot\int_{\varepsilon^*}\mathrm{P}(V(k')>\max(V(k^*),V(0))|\varepsilon_{t+1}=e_{k'},\varepsilon^*_{t+1}=e^*_{k^*})de^*_{k^*}\}\end{aligned}$$

- **Transition from work to work (new job) :**

$$\mathcal{L}_{E-\widetilde{E}} = (1-\delta)\lambda_1\times \mathrm{P}(\widetilde{VE}_t > VU_t\,,\, \widetilde{VE}_t > VE_t)\times \mathrm{pdf}_{\varepsilon^*}(e^*_t)$$

$$L(k\to l) = (1-\delta)\lambda_1\cdot \mathrm{P}(\varepsilon^*_{t+1}=e_l)\int_{\varepsilon_{t+1}}\mathrm{P}(V(l)>\max(V(k'),V(0))|\varepsilon_{t+1}=e_{k'},\varepsilon^*_{t+1}=e^*_{k^*})de_{k'}$$

1.B.2 Schooling contributions

1.B.2.1 Truncated Normal distribution

General notations

We focus on the joint distribution of X_1 and X_2 :

$$\begin{pmatrix}X_1\\X_2\end{pmatrix}\sim\mathcal{N}\left(\begin{pmatrix}0\\0\end{pmatrix},\begin{pmatrix}1&\rho\\\rho&1\end{pmatrix}\right)$$

The joint normal distribution φ_2 is given by :

$$\varphi_2(x,y,\rho)=\frac{1}{2\pi\sqrt{1-\rho^2}}\exp\left(\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right)$$

We define the following function :

$$\begin{aligned} f(a,b,\rho) &= \mathrm{E}[X_1.1_{\{}(X_1 > a, X_2 > b)\}] \\ &= \varphi(a)\Phi\left(\frac{\rho a - b}{\sqrt{1-\rho^2}}\right) + \rho\varphi(b)\Phi\left(\frac{\rho b - a}{\sqrt{1-\rho^2}}\right) \end{aligned}$$

Note the particularity of the cumulative distribution of the bivariate normal :

$$\begin{aligned} P(X_1 < a; X_2 < b) &= \Phi_2(a,b,\rho) \\ P(X_1 > a; X_2 > b) &= P(-X_1 < -a; -X_2 < -b) = \Phi_2(-a,-b,\rho) \\ P(X_1 > a; X_2 < b) &= P(-X_1 < -a; X_2 < b) = \Phi_2(-a,b,-\rho) \end{aligned}$$

Emax problem

We are interested in computing the Emax function defined as :

$$g(U_1,U_2,\sigma_1,\sigma_2,\rho,V) = \mathrm{E}[\max(U_1+\varepsilon_1, V, U_2+\varepsilon_2)]$$

where the random terms are ε_1 and ε_2 :

$$\begin{aligned} \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}\right) \\ \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 - \varepsilon_1 \end{pmatrix} &\sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 - \sigma_1^2 \\ \underbrace{\rho\sigma_1\sigma_2 - \sigma_1^2}_{\tilde{\rho}\sigma_1\tilde{\sigma}} & \underbrace{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}_{\tilde{\sigma}^2} \end{pmatrix}\right) \end{aligned}$$

Thus we define :

$$\begin{aligned} \tilde{\rho} &= \frac{\rho\sigma_2 - \sigma_1}{\tilde{\sigma}} \\ \tilde{\sigma} &= \sqrt{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2} \end{aligned}$$

The function g can be decomposed into three parts :

$$
\begin{aligned}
g &= \underbrace{\mathrm{E}[(U_1+\varepsilon_1).1_{\{U_1+\varepsilon_1>V, U_1+\varepsilon_1>U_2+\varepsilon_2\}}]}_{①} \\
&\quad + V.\underbrace{\mathrm{E}[1_{\{V>U_1+\varepsilon_1, V>U_2+\varepsilon_2\}}]}_{②} \\
&\quad + \underbrace{\mathrm{E}[(U_2+\varepsilon_2).1_{\{U_2+\varepsilon_2>V, U_2+\varepsilon_2>U_1+\varepsilon_1\}}]}_{③}
\end{aligned}
$$

We can compute an analytic form of each element of g :

$$
\begin{aligned}
① &= U_1 \mathrm{P}\left(\frac{U_1+\varepsilon_1}{\sigma_1} > \frac{V}{\sigma_1}\right) + \sigma_1 \mathrm{E}\left[\frac{\varepsilon_1}{\sigma_1}.1_{\left\{\frac{U_1+\varepsilon_1}{\sigma_1}>\frac{V}{\sigma_1}, U_1+\frac{U_1-U_2}{\tilde{\sigma}}>\frac{\varepsilon_2-\varepsilon_1}{\tilde{\sigma}}\right\}}\right] \\
&= U_1 \Phi_2\left(\frac{U_1-V}{\sigma_1}, \frac{U_1-U_2}{\tilde{\sigma}}, -\tilde{\rho}\right) + \sigma_1 f\left(\frac{U_1-V}{\sigma_1}, \frac{U_1-U_2}{\tilde{\sigma}}, -\tilde{\rho}\right)
\end{aligned}
$$

$$
② = \mathrm{P}\left(\frac{V}{\sigma_1} > \frac{U_1+\varepsilon_1}{\sigma_1}, \frac{V}{\sigma_2} > \frac{U_2+\varepsilon_2}{\sigma_2}\right) = \Phi_2\left(\frac{V-U_1}{\sigma_1}, \frac{V-U_2}{\sigma_2}, \rho\right)
$$

$$
③ = U_2 \Phi_2\left(\frac{U_2-U_1}{\tilde{\sigma}}, \frac{U_2-V}{\sigma_2}, \frac{\rho\sigma_1-\sigma_2}{\tilde{\sigma}}\right) + \sigma_2 f\left(\frac{U_2-U_1}{\tilde{\sigma}}, \frac{U_2-V}{\sigma_2}, \frac{\rho\sigma_1-\sigma_2}{\tilde{\sigma}}\right)
$$

Adaptation to our problem

In our application, the only Emax function that takes this form is the Emax function for schooling 4. In this case, we need to compute :

$$
\mathrm{E}\left[\max(VS_t^4(3,\kappa_3^4), VS_t^5(3,\kappa_3^5), VW_t(3)\right]
$$

where $VW_t(3)$ is the value of working with level 3 of schooling. Then we can easily identify the determinist and random part of the previous calculus :

$$\begin{aligned} U_1 &= c_3^4(W^4, Z, 0) + \beta^{l_4}\mathrm{E}[\max_{l\in\mathcal{C}_4} V_{t+1}^l(\Omega_{t+1})] \\ U_2 &= c_3^5(W^5, Z, 0) + \beta^{l_5}\mathrm{E}[\max_{l\in\mathcal{C}_5} V_{t+1}^l(\Omega_{t+1})] \\ \varepsilon_1 &= \kappa_3^4 \\ \varepsilon_2 &= \kappa_3^5 \\ V &= VW_t(3) \end{aligned}$$

1.B.2.2 Likelihood contributions

• Transition between two levels of schooling :

$$\begin{aligned} \mathcal{L}_{d-k} &= \mathrm{P}(VS_t^k = \max_{l\in\mathcal{C}_d} V_t^l) \\ &= \mathrm{P}\left(\forall l \in \mathcal{S}_d, \kappa^l - \kappa^k < S_t^k - S_t^l\,,\, \forall l \in \mathcal{V}_d, \eta^l - \kappa^k < S_t^k - L_t^l\right) \end{aligned}$$

Because $\mathcal{S}_d$ has at most 2 elements, the first part is not empty only if $\mathcal{S}_d \backslash \{k\}$ is not empty, i.e $\mathcal{S}_d = \{k_1, k_2\}$. In this case, supposing $k = k_1$:

$$\mathcal{L}_{d-k} = \mathrm{P}\left(\kappa^{k_2} - \kappa^k < S_t^k - S_t^{k_2}\,,\, \forall l \in \mathcal{V}_d, \eta^l - \kappa^k < S_t^k - L_t^l\right)$$

Otherwise :

$$\mathcal{L}_{d-k} = \mathrm{P}\left(\forall l \in \mathcal{V}_d, \eta^l - \kappa^k < S_t^k - L_t^l\right)$$

And then these probabilities are multinomial probit.

- **Transition from schooling level d to labor market with speciality k :**

$$\begin{aligned}\mathcal{L}_{S-L^k} &= \mathrm{P}(VL_t^k = \max_{l \in \mathcal{C}_d} V_t^l) \\ &= \mathrm{P}\left(\forall l \in \mathcal{S}_d, \kappa^l - \eta^k < L_t^k - S_t^l\,,\, \forall l \in \mathcal{V}_d \backslash \{k\}, \eta^l - \eta^k < L_t^k - L_t^l\right)\end{aligned}$$

1.C Estimates

1.C.1 Schooling costs Intercepts

TABLE 1.14 – Costs parameters : Intercept

$k = 1$	Cste Spe = 1 type = 1	Cste Spe = 1 type = 2	Cste Spe = 1 type = 3
Level 1	−1.88453 (0.11634)	−1.91607 (0.13201)	−1.19686 (0.657)
Level 2	−0.55737 (0.05038)	−0.5667 (0.05347)	−0.35398 (0.19571)
Level 3	−0.84906 (0.05217)	−0.86327 (0.05858)	−0.53923 (0.29567)
Level 4	0.20908 (0.15074)	0.21258 (0.15354)	0.13278 (0.11974)
Level 5	. (.)	. (.)	. (.)
$k = 2$	Cste Spe = 2 type = 1	Cste Spe = 2 type = 2	Cste Spe = 2 type = 3
Level 1	−0.34224 (0.06074)	−0.34797 (0.06265)	−0.21735 (0.12399)
Level 2	−0.91271 (0.09006)	−0.92798 (0.09568)	−0.57966 (0.31994)
Level 3	−0.29354 (0.05113)	−0.29845 (0.05286)	−0.18643 (0.10659)
Level 4	0.06086 (0.04802)	0.06187 (0.04885)	0.03865 (0.03684)
Level 5	−0.69041 (0.04883)	−0.70196 (0.05402)	−0.43847 (0.24064)
$k = 3$	Cste Spe = 3 type = 1	Cste Spe = 3 type = 2	Cste Spe = 3 type = 3
Level 1	−1.01318 (0.16729)	−1.03014 (0.17248)	−0.64347 (0.3657)
Level 2	−0.54158 (0.0877)	−0.55064 (0.09106)	−0.34395 (0.19378)
Level 3	−0.69852 (0.13274)	−0.7102 (0.13666)	−0.44362 (0.25527)
Level 4	0.56272 (0.18915)	0.57213 (0.19302)	0.35738 (0.22874)
Level 5	−0.00755 (0.10912)	−0.00768 (0.11094)	−0.0048 (0.06938)

1.C.2 Offers arrival rate

1.C.2.1 Only Labor Market

Table 1.15 – Labor market estimates for Benchmark Model

	k = 0	k = 1	k = 2	k = 3
α_0	-0.20374 (7e–05)	-0.20374 (7e–05)	-0.20374 (7e–05)	-0.20374 (7e–05)
α_x	0.09959 (0.0074)	0.08805 (0.01159)	0.07198 (0.0062)	0.09157 (0.00461)
α_{x^2}	-0.0101 (0.00359)	-0.08501 (0.0055)	$-3e-05$ (0.00275)	$-5e-04$ (0.00216)
$\alpha_{d=2}$	. (.)	. (.)	0.14465 (0.00705)	0.15683 (0.00612)
$\alpha_{d=3}$	. (.)	0.17066 (0.00011)	0.16356 (0.00789)	0.14932 (0.00626)
$\alpha_{d=4}$	. (.)	. (.)	0.168 (0.00933)	0.19219 (0.00725)
$\alpha_{d=5}$	. (.)	0.19065 (0.01388)	0.13135 (0.02388)	0.11161 (0.0127)
$\alpha_{d=6}$	. (.)	0.21823 (0.00015)	0.212 (0.01301)	0.23356 (0.00991)
σ_e	0.381 (0.00301)	0.34887 (0.00454)	0.32077 (0.00221)	0.33443 (0.00167)
σ_v	0.07355 (0.00013)	0.07541 (0.00033)	0.07549 (0.00012)	0.07476 (1e–04)

	k = 0	k = 1	k = 2	k = 3
$\lambda_{0,d=1}$	0.48519 (0.00623)	. (.)	. (.)	. (.)
$\lambda_{0,d=2}$	. (.)	. (.)	0.47799 (0.00668)	0.5231 (0.00607)
$\lambda_{0,d=3}$	. (.)	0.50948 (0.01415)	0.48079 (0.00883)	0.47412 (0.00666)
$\lambda_{0,d=4}$	. (.)	. (.)	0.47971 (0.0102)	0.48676 (0.00746)
$\lambda_{0,d=5}$	. (.)	0.43629 (0.01278)	0.42679 (0.03746)	0.44091 (0.01356)
$\lambda_{0,d=6}$	. (.)	0.46765 (0.01999)	0.44901 (0.01693)	0.41492 (0.01212)

	k = 0	k = 1	k = 2	k = 3
$\lambda_{1,d=1}$	0.38738 (0.00841)	. (.)	. (.)	. (.)
$\lambda_{1,d=2}$	. (.)	. (.)	0.35811 (0.01005)	0.29993 (0.00762)
$\lambda_{1,d=3}$	. (.)	0.29811 (0.01571)	0.30944 (0.01064)	0.3254 (0.00875)
$\lambda_{1,d=4}$	. (.)	. (.)	0.2865 (0.01241)	0.29156 (0.00835)
$\lambda_{1,d=5}$	. (.)	0.30505 (0.01885)	0.28968 (0.0284)	0.307 (0.01906)
$\lambda_{1,d=6}$	. (.)	0.21463 (0.01827)	0.25484 (0.01779)	0.24349 (0.01179)

Table 1.16 – Parameters estimates of the Benchmarck Model

	k = 0	k = 1	k = 2	k = 3
$\delta_{d=1}$	0.41074 (−0.08652)	. (.)	. (.)	. (.)
$\delta_{d=2}$	. (.)	. (.)	0.41759 (0.0015)	0.4133 (0.00144)
$\delta_{d=3}$	. (.)	0.41453 (0.00297)	0.40962 (0.00184)	0.41294 (0.00147)
$\delta_{d=4}$	. (.)	. (.)	0.41108 (0.00249)	0.40832 (0.00161)
$\delta_{d=5}$	. (.)	0.41339 (0.00299)	0.40038 (0.00702)	0.41051 (0.00332)
$\delta_{d=6}$	. (.)	0.41529 (0.0041)	0.40922 (0.00351)	0.40903 (0.00295)

1.C.2.2 Full Model

Table 1.17 – Probability of receiving an offer while not working

	k = 0	k = 1	k= 2	k = 3
$\lambda_{0,d=1}$	0.45407 (0.00647)	. (.)	. (.)	. (.)
$\lambda_{0,d=2}$	. (.)	. (.)	0.45932 (0.00664)	0.53554 (0.00589)
$\lambda_{0,d=3}$	. (.)	0.49698 (0.0141)	0.50577 (0.00791)	0.48035 (0.00628)
$\lambda_{0,d=4}$	. (.)	. (.)	0.49846 (0.00996)	0.49301 (0.00737)
$\lambda_{0,d=5}$	. (.)	0.44861 (0.01225)	0.45141 (0.0367)	0.44091 (0.01349)
$\lambda_{0,d=6}$	. (.)	0.46143 (0.02044)	0.46763 (0.01661)	0.43324 (0.01217)

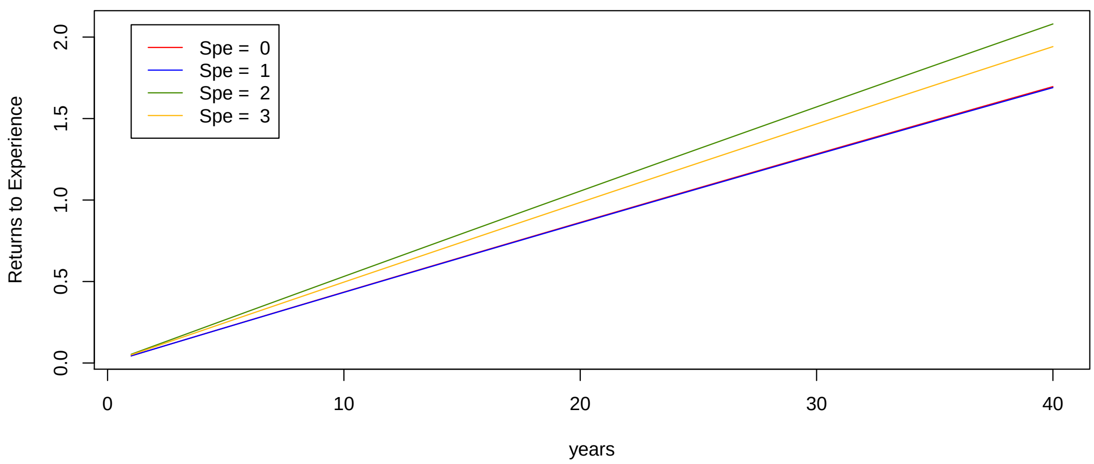
TABLE 1.18 – Probability of receiving an offer while working

	k = 0	k = 1	k = 2	k = 3
$\lambda_{1,d=1}$	0.42968 (0.00874)	. (.)	. (.)	. (.)
$\lambda_{1,d=2}$	. (.)	. (.)	0.36393 (0.00996)	0.31055 (0.00782)
$\lambda_{1,d=3}$	. (.)	0.33599 (0.01608)	0.29885 (0.00961)	0.33091 (0.00842)
$\lambda_{1,d=4}$	. (.)	. (.)	0.26651 (0.01231)	0.28642 (0.00823)
$\lambda_{1,d=5}$	. (.)	0.33775 (0.01823)	0.26955 (0.02665)	0.30171 (0.01942)
$\lambda_{1,d=6}$	. (.)	0.21463 (0.01921)	0.23185 (0.01774)	0.21693 (0.012)

TABLE 1.19 – Probability of a position to be non renewed

	k = 0	k = 1	k = 2	k = 3
$\delta_{d=1}$	0.41691 (−0.08033)	. (.)	. (.)	. (.)
$\delta_{d=2}$	. (.)	. (.)	0.41762 (0.00149)	0.41331 (0.00144)
$\delta_{d=3}$	. (.)	0.40844 (0.00301)	0.40961 (0.00185)	0.41295 (0.00148)
$\delta_{d=4}$	. (.)	. (.)	0.4111 (0.0025)	0.40827 (0.00164)
$\delta_{d=5}$	. (.)	0.41339 (0.0031)	0.40038 (0.00707)	0.4105 (0.00336)
$\delta_{d=6}$	. (.)	0.41528 (0.00413)	0.40921 (0.00355)	0.40902 (0.00296)

FIGURE 1.9 – Return to experience



Chapitre 2

Early Tracking and the Distribution of Risk Aversion and Psychological Traits : Evidence from the German Socio-Economic Panel [1]

Using recent waves of the German Socio-Economic Panel (GSOEP), we investigate the distribution of risk aversion and psychological traits (measured at age 17 and18). We construct a model that maps family background characteristics onto six unobserved underlying factors. We adopt a factor structure which accounts for measurement error. The focus of our analysis is on the distinction between those who are in the academic track and those who are found to be in the professional track. Our results indicate that differences in education track (as measured at age 17) seem to translate into sizeable differences in Intelligence and Motivation. However, risk aversion remains weakly correlated with educational track. Our findings also suggest that the usual orthogonality assumption between factors is illusive.In particular, cognitive and non-cognitive factors are highly correlated while risk aversion is much less correlated with either cognitive or non-cognitive skills.

[1]This chapter was written with Christian Belzil (Ecole Polytechnique, ENSAE, and IZA), Francois Poinas (Toulouse School of Economics) and Konstantinos Tatsiramos (University of Leicester and IZA).

Introduction and Motivation

In many western countries, educational institutions allow for the parallel existence of professional and general schooling tracks. This is particularly true in Germany, where children are typically assigned to either an academic track (*Gymnasium*), or a professional track (*Hauptschule* and *Realschule*) as early as age 10. This assignment is crucial because *Gymnasium* naturally leads to higher education (university or technical school) while very few individuals enrolled in professional high-schools eventually reach higher education. Evaluating the degree of heterogeneity between individuals assigned to different tracks is therefore crucial for the purpose of designing optimal education policies.

Our objective is to analyze how individuals in academic and non-academic tracks differ with respect to a wide class of personality traits and economic preferences. To be precise, we consider one preference parameter, risk aversion, and five psychological traits : intelligence, motivation, conscientiousness, trust and positive reciprocity.[2] We document to what extent the distribution of personality traits and risk aversion varies according to early tracking and to family background characteristics.

Our objective is fulfilled using recent waves of the German Socio-Economic Panel (GSOEP), which contain detailed information on family characteristics (family income, parents education, geographical origin, etc.) as well as self-reported psychometric measures collected when individuals are 17 years old.

The originality of our approach is that we estimate the distribution of factors without imposing orthogonality. To achieve this, we construct a model that maps family background characteristics onto six unobserved underlying factors. Adopting a factor structure permits to extract one underlying unobserved component from a multiplicity of observed measures, while accounting for measurement error. Our

[2]Conscientiousness is the degree to which a person is willing to comply with conventional rules and norms. It practically means that someone is efficient, self-organized, systematic and hardworking. Positive reciprocity and Trust are related to both preferences and beliefs about the behavior of others. Reciprocity refers to responding to a positive action with another positive action and rewarding kind behavior.

analysis focuses on the distinction between those who are in the academic track and those who are found to be in the professional track by age 17, as it represents one of the major sources of differences in access to higher education.

At the outset, it should be clear that our approach is mostly descriptive. While measuring risk aversion obviously requires introducing preferences into the analysis, our model remains silent about the causal effect of early tracking on those psychometric factors and on risk aversion. Our objective is to obtain an accurate description of those factors by age 17. Because the factors are allowed to depend on parental background, we can investigate if differences in psychological and economic traits are explained by differences in background variables, or by differences in tracks that persist even after conditioning on parental characteristics. However, we do not attempt to distinguish the track-specific causal effect on factor/preference formation from the effect of parents' expectation about the ability and preferences of their own children on the choice of a track.[3]

This paper is line with the recently emerging literature on separating the effects of cognitive and non-cognitive factors on educational outcomes.[4] As it will become clear later, our paper is complementary to a large number of recent papers in that we show how a large set of preferences and traits relate to early schooling status. Until now, applied econometricians have focused on measuring the effects of psychological traits on schooling outcomes.[5] Our work also contributes to the recent literature attempting to measure the correlation between fundamental economic preferences

[3]Because tracks are chosen as early as age 10, it is reasonable to assume that both parents and teachers are actually those who take the decision.

[4]Seminal work by Osborne, Gintis, and Bowles (2001) has launched a recent literature studying the effects of non-cognitive factors, and applied econometricians are now paying more and more attention to collection of variables that may help uncover latent non-cognitive skills, as well as other personality traits. This is exemplified in Carneiro, Hansen, and Heckman (2003), Heckman, Stixrud, and Urzua (2006), as well as in Cunha, Heckman, and Schennach (2010). Recently, Heckman, Humphries, Urzua, and Veramendi (2010) have quantified the relative importance of cognitive and non-cognitive skills for those individuals who drop-out of high school and those who obtain their diploma through the GED program.

[5]This is the case in Heckman, Stixrud, and Urzua (2006) who investigate the relative importance of cognitive and non-cognitive skills on education using a sample of young Americans. Belzil and Leonardi (2013) investigate how individual specific risk aversion affect the decision to enroll in higher education in Italy.

and psychological traits such as cognitive or non-cognitive skills.[6]

Our results indicate that differences in education track (as measured at age 17) seem to translate into sizeable differences in Intelligence and Motivation. These differences are simultaneously explained by the fact that the educational track status is largely correlated with parental human capital (education and income) and that psychological traits are also highly correlated with the same variables. However, risk aversion remains only weakly correlated with the educational track. Our findings suggest that the usual orthogonality assumption between factors is illusive. For instance, the correlation between Intelligence and Motivation is as high as 0.91. We also find that the pattern of correlation between background variables and risk aversion is quite different from the pattern observed for other factors. Risk aversion is much less correlated with cognitive skills and motivation than other factors. Our results are at odds with those reported in **?** who underline that cognitive ability remains correlated with risk aversion even after controlling for other individuals characteristics.

The paper is structured as follows. In the next section, we provide a relatively detailed description of the components of the GSOEP used in our analysis. In section 2.2, we present the econometric model and provide information about the estimation methods. Section 2.3 is devoted to the presentation of the main results. The conclusion is found in Section 2.4.

2.1 The GSOEP Data

The German Socio-Economic Panel is an annual survey which started in 1984 and is widely used by economists and sociologists. The panel data have a high degree of stability over time : private households, persons and families have been consistently

[6] Although empirical work remains in its infancies, many economists have already investigated the link between risk aversion, time preferences and cognitive skills. This is the case, for instance, in **?**, who conduct choice experiments measuring risk aversion, impatience and cognitive ability, over an annual time horizon, and for a randomly drawn sample of German adults. This branch of the literature is surveyed in Almlund, Duckworth, Heckman, and Kautz (2011).

surveyed over time on topics like household composition, employment, earnings, health, satisfaction indicators, objective living conditions, values, willingness to take risks, etc.

The GSOEP incorporates a variety of questionnaires measuring risk attitude, cognitive ability tests, non-cognitive ability measures, trust and reciprocity. Because the questions that we use have not all been asked in the same year, we consider various cohorts of 17 years old. These questions are either included in the adult questionnaire, which applies to all individuals aged 17 or more (until 2005), or in the youth questionnaire (since 2006), which is devoted only to those aged 17.

To complete our analysis, we also use data on family characteristics and psychological traits taken at age 17.[7] We use waves from 2004 to 2009. In total, 1,784 individuals are included in our sample.

2.1.1 The German Education System

In Germany, children are selected into different types of secondary schools as early as age 10. This educational system is often referred to as an "Early Tracking" system. There exists 3 types of secondary schools : *Hauptschule*, *Realschule* and *Gymnasium*.[8] They differ with respect to 3 main dimensions. The first dimension concerns the number of academic years covered : students stay 5 years in *Hauptschule* (from grade 5 to 9), 6 years in *Realschule* (grades 5 to 10) and 8 or 9 years in *Gymnasium* (grades 5 to 12 or 13, depending on the state on which they are registered). The second dimension is the content of the programs. Education in *Gymnasium* is intensive in general knowledge, whereas programs in the two other types of school is less demanding in general education and contains courses with a practical content. The third dimension is the type of education accessible after graduation. Completion of *Gymnasium* requires passing a terminal exam, called *Abitur*, that permits to enter

[7]For a portion of our sample, information about risk aversion is collected at 18 years old.

[8]In some states, there exists a fourth type of secondary school, called *Gesamtschule*. This is a comprehensive school that groups the three types of curricula.

higher education (universities or technical colleges). After exiting from *Hauptschule* or *Realschule*, individuals can enter an apprenticeship program or a vocational school. They also have the possibility to get a qualification in order to enter a *Gymnasium* at an advanced stage.

In our analysis, we focus on two categories based on the observed track at age 17. To be in the academic track, one must be enrolled in *Gymnasium* at age 17. This may include individuals who started their secondary education in *Gymnasium* and individuals who started in another type of school and switched to *Gymnasium*. The professional (non-academic) category is composed of the rest of the population, i.e.individuals who started secondary education in *Realschule* or *Hauptschule* and did not switch to *Gymnasium*.[9] According to this classification, in our sample, 47% of individuals belong to the academic track, whereas 53% belong to the non academic track (see Table 2.5 in Appendix 2.B).

2.1.2 Risk Aversion Measurement

The GSOEP contains a number of questions which provide measures of risk preference. In this paper, we use the question that pertains to a potential investment in a risky environment.[10] We do so for the following reason. Because beliefs are held constant in the investment question, differences in responses are more clearly attributable to risk aversion alone, as compared to the other measures, which are potentially incorporate by risk aversion as well as risk perceptions.[11]

Respondents are faced with the following question :

---

[9] Individuals enrolled in *Gesamtschule* who did not switch to an academic track are also included in the non academic category. The main reason is that their outcomes are more in lines with individuals who were enrolled in *Hauptschule* or *Realschule*. This concerns 48 cases in our sample (3%).

[10] See Dohmen, Falk, Huffman, Sunde, Schupp, and Wagner (2011), for a detailed description. Dohmen, Falk, Huffman, and Sunde (2012) have used the same data in order to compare the predictive power of all risk measures available in the data set.

[11] The GSOEP contains other measures which consist of self-reported ordered responses for the willingness to take risk in specific environments (driving, sports, work, and others). All of these measures are characterized by ambiguity as they leave it up to the respondent to imagine the relevant probabilities and stakes.

"*Imagine you had won 100,000 Euros in a lottery. Almost immediately after you collect, you receive the following financial offer from a reputable bank, the conditions of which are as follows : There is the chance to double the money within two years. It is equally possible that you could lose half of the amount invested. What fraction of the 100,000 Euros would you choose to invest ?*"

Respondents are allowed six possible responses : 0, 20,000, 40,000, 60,000 80,000, or 100,000 Euros. This measure shares the common feature of other lottery measures in that it presents respondents with explicit stakes and probabilities, and thus holds risk perceptions constant across individuals. As documented in Dohmen, Falk, Huffman, and Sunde (2010), this type of question is very useful to measure risk aversion.

The exact wording of the general question (translated from German) is :

"*How do you see yourself : Are you generally a person who is fully prepared to take risks or do you try to avoid taking risks ? Please tick a box on the scale, where the value 0 means : 'risk averse' and the value 10 means : 'fully prepared to take risks'.*"

2.1.3 Non Cognitive Traits

A substantive importance is attributed to subjective indicators in the design of questionnaires for the GSOEP. These questions have been used in psychology and political investigations. They are constructed to measure personality differences in perception, social behavior, values and motivation. Individuals answer on a seven point Likert scale ranging from "strongly agree" to "strongly disagree" or "does not apply to me at all" to "applies to me perfectly" . It is important to note that the questions administered in the GSOEP have been designed so to reduce the natural tendency for individuals to answer in the center of the scale, in order to conform to social desirability.

Starting with the 2005 wave, a block of questions on the Big Five personality dimensions and another one on the Locus of Control were introduced so to increase the number of personality questions in the GSOEP. Psychologists have shown that differences in personality in Western societies can be traced back to five personality dimensions : neuroticism, extrovertedness, openness to experience, agreeableness and conscientiousness. The Big Five index in the GSOEP is an abbreviated version of the common Costa McCrae 240 item index.[12] The Locus of Control is a theory referring to individuals' beliefs about their ability to control events that affect them in life. Psychologists consider these beliefs as important in understanding individuals' motivation. Those with a high internal locus of control believe it is their behavior and actions that determine results, while individuals with a high external locus of control believe that others, fate or chance are more important in determining events.

In our analysis, and in term of the big five personality dimensions, we focus on Conscientiousness which is documented as the most predictive of the usual personality measures, especially in a schooling context (see Heckman and Kautz (2012), and Almlund, Duckworth, Heckman, and Kautz (2011)). The questions we use relate to how one sees him or herself on "doing a thorough job" , on "tending to be lazy" and on "doing things effectively and efficiently". The specific questions and their associated sets of possible answers are found in Appendix 2.A.

Because motivation is also likely a key component of schooling decisions, we use a set of ten questions from the Locus of Control, which are related to different attitudes towards life and the future. These are also described in Appendix 2.A.

2.1.4 Cognitive Skills

Starting from 2006, the GSOEP collects data on the general cognitive abilities of the 17-years-old respondents participating in the survey for the first time. The measurement of cognitive skills is based on the IST-2000R test developed by Amthauer,

[12]See Borghans, Duckworth, Heckman, and ter Weel (2008) and Almlund, Duckworth, Heckman, and Kautz (2011).

Brocke, Liepmann, and Beauducel (2001). The IST-2000R draws on three general cognitive ability components : (a) verbal abilities, (b) numerical abilities, and (c) figural abilities. Each content factor is measured by means of three subscales (each consisting of 20 items) : verbal abilities (sentence completion, verbal analogies, finding similarities), numerical abilities (arithmetic operators, number series, arithmetic problems), and figural abilities (figure selection, cube task, matrices). The total score, consisting of the three content factors, reflects what can be termed reasoning (or general cognitive) abilities. The total time for the administration of this (basic) cognitive ability module is approximately 90 minutes. For the purpose of the GSOEP youth questionnaire, only a subset of the components of the test were selected. Verbal analogies were used to measure verbal cognitive potential, numerical series to measure numerical cognitive potential, and matrices to measure figural cognitive potential. All three measures of cognitive abilities are reported on a scale ranging from 1 to 20.

2.1.5 Social Preferences : Trust and Reciprocity

Social preferences such as Trust and Reciprocity are also measured in the GSOEP. Trust is measured in the Youth Questionnaire since 2006 by asking individuals about their opinion on a number of statements such as : "On the whole, one can trust people", "On the whole, one can't rely on anyone", and "On the whole, one should be careful before trusting strangers". The answers are on a four point Likert scale and are described in Appendix 2.A.[13]

Individuals are also asked in year 2005 to respond to various questions or statements that are meant to measure their degree of reciprocity. In the paper, we focus on positive reciprocity. We use three questions on whether one is prepared to return a favor, help those who were kind or helped the person in the past. More details are found in Appendix 2.A.

[13] The first two questions were also asked in the Adult Questionnaire in 2004.

2.1.6 Parental Background Variables

Aside from the factors mentioned earlier, we also condition our analysis on a wide range of family background variables. In particular, we create indicators equal to 1 when the father and/or the mother has graduated from Gymnasium. We also include indicators for gender, for those living in the east or the south of Germany and for those who have a foreign nationality. Our analysis also incorporates a variable measuring the number of siblings below 14 years of age. Finally, we also condition on household income. More details may be found in the section devoted to the econometric model.

2.2 The Econometric Model

In this section, we present the main components of our estimation strategy : the statistical specification of the distribution of factors and the measurement equations. Because the risky investment question used to infer risk aversion admits a structural interpretation, we also devote a specific sub-section to risk aversion.

2.2.1 Defining Risk Aversion

As indicated earlier, the risky investment-lottery question allows for six possible responses for the amount invested : 0, 20,000, 40,000, 60,000, 80,000, or 100,000 Euros. In order to model individual choices, we first assume that preferences are represented by Constant Relative Risk Aversion (CRRA) utility function, denoted $U(.)$, and set the benchmark wealth level equal to 0.

Because the GSOEP does not provide information on the benchmark consumption level and on life-cycle consumption, it is unrealistic to model investment shares using an intertemporal framework. To resolve this issue, we interpret individual decisions as being generated from a comparison of the expected utilities of discounted income (at rate $\beta = 0.95$) for each investment shares. So, the expected payoff associated to a windfall gain, g, and a potential investment share s, which is denoted $EV(.)$, is simply :

$$EV(.) = 0.5\frac{\{(1-s)\cdot g + \beta^2(2\cdot s\cdot g)\}^{1-\theta_i}}{1-\theta_i} + 0.5\frac{\{(1-s)\cdot g + \beta^2(0.5\cdot s\cdot g)\}^{1-\theta_i}}{1-\theta_i} \quad (2.1)$$

To model the investment share probabilities, we also introduce an additional error which may be potentially interpreted as an optimization error, or a utility shock that affects individual choices. This is an important feature of our model as this is what allows us to treat risk aversion context specific. However, it would be impossible to model the choice of the optimal share as a standard ordered probit simply because the optimal share is not a monotone function of risk aversion.

To proceed, we represent the probabilistic optimal choice (from the perspective of the econometrician) as follows :

$$EV_i^*(s_l;\theta_i) = EV(s_l;\theta_i) + \kappa_i^l \quad (2.2)$$

where the index l represents all the possible shares (0.0, 0.2, 0.4, 0.6, 0.8, and 1.0). The share specific error term, κ_i^l, follows an extreme-value distribution with parameter τ. The choice probabilities are therefore given by :

$$\begin{aligned}\Pr(S_i^* = s_l) &= \Pr\{EV_i^*(s_l;\theta_i) > EV_i^*(s_k;\theta_i)\} \qquad \forall k, k \neq l \\ &= \frac{\exp(EV(s_l;\theta_i)/\tau)}{\sum_k \exp(EV(s_k;\theta_i)/\tau)} \end{aligned} \quad (2.3)$$

2.2.2 Modeling Psychological Factors and Risk Aversion

2.2.2.1 Psychological Traits

In order to estimate the model, we assume the existence of 5 distinct factors.[14] Each factor, F_j, is the sum of two parts : a deterministic component and a random

[14] In terms of the usual terminology used in factor analysis, our analysis is more "confirmatory" than "exploratory".

(unobserved to the econometrician) component which follows a Normal distribution :

$$F_j = \overline{F_j} + F_j^* \text{ for } j = 1, 2, \ldots, 5$$

The factors F_j are cognitive ability (denoted I), conscientiousness (CO), motivation (MO), trust (TR), and reciprocity (RE).

Formally, we have :

$$\begin{aligned} I = \bar{I} + I^* &\quad \text{with} \quad I^* \sim \mathcal{N}(0, \sigma_I^2) \\ CO = \overline{CO} + CO^* &\quad \text{with} \quad CO^* \sim \mathcal{N}(0, \sigma_{co}^2) \\ MO = \overline{MO} + MO^* &\quad \text{with} \quad MO^* \sim \mathcal{N}(0, \sigma_{mo}^2) \\ TR = \overline{TR} + TR^* &\quad \text{with} \quad TR^* \sim \mathcal{N}(0, \sigma_{tr}^2) \\ RE = \overline{RE} + RE^* &\quad \text{with} \quad RE^* \sim \mathcal{N}(0, \sigma_{re}^2) \end{aligned}$$

All deterministic parts have an identical linear structure :

$$\overline{F}_j = \sum_{r=1}^{8} \phi_j^r \cdot x_r + \phi_j^T \cdot Track$$

where x_r denotes the components of the vector of family characteristics and the variable $Track$ is an indicator for being in an academic track at age 17. The full vector, denoted X, includes the following variables :

- a gender indicator (1 for females and 0 for males),
- two indicators for father and mother education (equal to 1 if the parent has finished *Gymnasium*),
- two indicators of the region of residence, one equal to 1 if the individual lives in a state belonging to the former East Germany and the other one equal to 1 if the individual lives in southern Germany (either in Bavaria or Bade-Wurtemberg),

- an indicator for migration status (equal to 1 if the individual's parents are not German citizens),
- the number of siblings between 0 and 14 years old when the individual is aged 17,
- family income, measured as the yearly per capita income (in tens of thousands of 2005 euros).

2.2.2.2 Risk Aversion

For risk aversion, we adopt a slight modification so to eliminate risk loving behavior. If we allowed for a set of risk loving individuals, this would complicate the interpretation of the results. So, we assume that :

$$\theta_i = \exp\left\{\theta^0 + \sum_{r=1}^{8} \theta^r \cdot x_r + \theta^T \cdot Track_i + \theta_i^*\right\} \text{ with } \theta_i^* \sim \mathcal{N}(\mu_{ra}, \sigma_{ra}^2)$$

This equality is used in equation (2.3) to write the likelihood.

To simplify notation, we group all the factors (their latent part) into a vector, $\Omega^* = \{\theta^*, I^*, CO^* MO^*, TR^*, RE^*\}$ and denote its distribution $G(\Omega^*)$. The distribution function $G(\Omega^*)$ is essentially a multi-variate normal of dimension 6, with all of its components being mutually independent.

2.2.3 Measurement Equations

In total, we model 22 measurement equations. Most of the measurements (19) are reported as discrete-ordered variables, while only three of them (the cognitive tests) are measured by a continuum of values. Explicit questions are reported in Appendix 1.

For the discrete-ordered variables, we assume that each measurement, denoted

M_k, is driven by a continuous latent variable, M_k^*, which is represented as follows

$$M_k^* = \alpha^k + \sum_{j=1}^{5} \alpha_j^k \cdot F_j + \varepsilon_{Mk} \text{ for } k = 1, 2, \ldots, 19 \tag{2.4}$$

where ε_{Mk} is a pure measurement error term. In line with factor analysis, we assume that $\varepsilon_{Mk} \sim \mathcal{N}(0, \sigma_{Mk})$. As measurements are reported as ordered discrete variables (out of W possibilities), we normalize σ_{Mk} to 1 and we estimate a set of $W-1$ measurement specific thresholds, $\tau_1, \tau_2, ..., \tau_{W-1}$. For each measurement, the contribution to the likelihood is simply the probability that a measurement outcome, M_k, takes a particular modality m_k. It is evaluated by the probability that the latent variable, M_k^*, lies between a lower and a upper threshold.

For the continuous cases (verbal, numerical and figural abilities), the model is simply :

$$T_{k'} = \gamma^{k'} + \sum_{j=1}^{5} \gamma_j^{k'} \cdot F_j + \varepsilon_{Tk'} \text{ for } k' = 1, 2, 3 \tag{2.5}$$

where T_k' is the score obtained to the test and $\varepsilon_{Tk'} \sim \mathcal{N}(0, \sigma_{Tk'})$.

For an individual i, we denote the contribution to the likelihood of a discrete measurement by $\Pr(M_{ki} = m_{ki})$ and the contribution of a continuous measurement by $\Pr(T_{k'i})$.

2.2.4 Identification of the Factor Structure

Following Cunha and Heckman (2008), we impose several restrictions on the contribution of each factor to potential measurements. For instance, many measurements are naturally devoted to specific factors. The restrictions imposed to the contribution of factors to measurements are summarized in Table 2.1.

Finally, in order to estimate the model, the following restrictions also need to be imposed. First, we normalize loading parameters to 1 for the first measurement questions related to cognitive tests, locus of control, conscientiousness, trust and reciprocity. Second, the variance of the measurement error term for discrete-ordered

Table 2.1 – Exclusion Restrictions in the Model of Factors

Type of question	# of questions	Factors				
		I	MO	CO	TR	RE
Cognitive tests	3	X	·	·	·	·
Locus of control	10	·	X	·	·	·
Conscientiousness	3	·	·	X	·	·
Trust	3	·	·	·	X	·
Reciprocity	3	·	·	·	·	X

Note : Entries marked by an "X" denote when a factor (in the column) enters the measurement equation identified in the row.

measurement equations (σ_{Mk}) is set to 1.

2.2.5 Estimation Method and Likelihood Function

The model is estimated by simulated maximum likelihood methods. The likelihood function is written conditionally on the Track indicator and other covariates and reflects the distribution of the measurements used in our analysis, conditional on each schooling track. It depends on the unobserved components of the factors ($\theta^*, I^*, CO^* MO^*, TR^*, RE^*$). Formally, the likelihood function is given by the following expression :

$$\int \cdots \int \prod_i \left\{ \Pr(S_i^* \mid \theta_i^*) \cdot \prod_k \Pr(M_{ki} \mid \Omega^*) \cdot \prod_{k'} \Pr(T_{k'i} \mid I^*) \cdot \right\} \, dG(\Omega^*) \tag{2.6}$$

where $\Omega^* = \{\theta^*, I^*, CO^* MO^*, TR^*, RE^*\}$. All components of Ω^* are mutually independent, even though the total factors are arbitrarily correlated since they all depend on a common set of parental and geographical background variables.

To integrate with respect to the distribution of the unobserved components of risk aversion and all psychometric traits, we use 20 draws per individual. The model is estimated using a Fortran procedure.

2.3 Empirical Results

As a first step, we estimated the full model as described in Section 2.2. However, our estimates of $G(\Omega^*)$ indicated that the latent portion of all factors (included risk aversion) converged to a degenerate distribution. This is most likely explained by the fact that both the factors and risk aversion depend on a wide set of observed regressors and that many of those regressors are only weakly correlated with each other, as indicated in Table 6 found in appendix. So, to avoid identification issues that arise when the standard deviation of the latent part of the factor approaches 0, we re-estimated the model after removing all those irrelevant components. The results presented below are therefore those obtained when removing the latent parts of all factors and risk aversion.

As a first step, we show how the factors are distributed in the whole population and in each sub-population separately. Then, we investigate how parents' background variables are related to risk aversion, skills and personality traits after conditioning on the Track indicator.

2.3.1 The Distribution of Factors and Differences between Tracks

There are two ways to regard differences across tracks. The first one is to measure the mean of all factors for each track, while letting background variables vary accordingly. This provides us with an unconditional (on background variables) measure. It captures both the effect of track and the differences in background variables across tracks. This approach is summarized in Table 2.2.

This result discloses important differences between academics and professionals in terms of cognitive ability (Intelligence). Indeed, the mean value of Intelligence is 6 times as large for individuals in the academic track (3.60) as it is for individuals in a non academic track (0.61). This is the only factor which discloses such a sizeable

Table 2.2 – Summary Statistics of Factors

		Whole population	Non academic track	Academic track
Risk aversion	Mean	4.3726	4.2950	4.4610
	S.D.	(0.2914)	(0.2661)	(0.2940)
Trust	Mean	0.0317	-0.0051	0.0735
	S.D.	(0.0704)	(0.0538)	(0.0633)
Intelligence	Mean	2.0044	0.6069	3.5963
	S.D.	(1.7647)	(0.9451)	(0.9402)
Motivation	Mean	0.0343	-0.0013	0.0748
	S.D.	(0.0504)	(0.0327)	(0.0335)
Conscientiousness	Mean	0.0131	0.0659	-0.0471
	S.D.	(0.1584)	(0.1410)	(0.1557)
Reciprocity	Mean	-0.2614	-0.1813	-0.3526
	S.D.	(0.2179)	(0.2010)	(0.1999)
Individuals		1,784	834	950

Note : Estimated factors are computed for each individual and unconditional on covariates. Statistics represented in this table are computed on the whole population and separately for individuals enrolled in the academic track and individuals not in academic track at age 17.

gap between academics and professionals. However, and as indicated by the standard deviations, there is a substantial level of dispersion within each group.

Although differences in Trust and Motivation appear less spectacular, those in the academic track also dominate the non-academic in terms of those factors. The opposite is observed for Reciprocity and Conscientiousness.

Interestingly, we find a very small difference between academics and professional in terms of measured risk aversion. The average degree of relative risk aversion is 4.46 for academics and 4.30 for professionals and is also commensurate with those values reported in the microeconometric literature.[15] This is therefore consistent with the idea that attitudes toward risk may not be influenced by past schooling level and more precisely that risk aversion is a true economic primitive.

At this stage, one question arises naturally. To what extent, is the difference between academics and non-academics driven by differences in parental background only. To answer this question, we must consider a conditional approach and examine the parameter estimates capturing the effect of track (the ϕ_j^T's and θ^T), after conditioning on all background variables. Those estimates, reported in Table 3 indicate that differences in psychological traits between academics and professionals (after conditioning on background variables) remain after conditionning on a wide set of characteristics, although the differences are smaller than those obtained when we do not condition on background variables. The academic-professional gap is positive for Intelligence, Motivation and Trust, and negative for the rest.

In light of the correlatons between the track indicator and other background variables reported in appendix (Table 6), those persistent effect of early tracking status is not that surprising. For instance, the correlation between early tracking status and father's Gymnasium completion (the variable most correlated with early tracking status) is only 0.37. The correlation with household income is only 0.23. All other correlations are really low.

[15] Most estimates that allow for cross-sectional heterogeneity are obtained from laboratory experiments, or various field experiments.

Table 2.3 – Parameter Estimates of Individual Characteristics on Factors

	Risk Aversion	Trust	Intel-ligence	Motiva-tion	Conscien-tiousness	Recipro-city
- Intercept	1.4169***	-	-	-	-	-
	(0.0253)					
- Academic track	0.0261***	2.4495***	0.0516***	-0.0435***	0.0429***	-0.2126***
	(0.0050)	(0.0473)	(0.0036)	(0.0032)	(0.0039)	(0.0097)
- Female	-0.0331***	-0.5556***	-0.0108***	0.2276***	-0.0139***	0.2384***
	(0.0046)	(0.0185)	(0.0027)	(0.0057)	(0.0028)	(0.0146)
- Father finished	0.0776***	0.4264***	0.0179***	-0.0905***	0.0353***	0.1843***
Gymnasium	(0.0057)	(0.0177)	(0.0038)	(0.0042)	(0.0044)	(0.0107)
- Mother finished	-0.0637***	0.7871***	0.0329***	-0.1058***	0.0709***	0.0305***
Gymnasium	(0.0066)	(0.0325)	(0.0040)	(0.0047)	(0.0047)	(0.0074)
- Living in east	-0.0615***	0.7144***	-0.0226***	0.0573***	-0.0660***	-0.1432***
Germany	(0.0047)	(0.0253)	(0.0028)	(0.0033)	(0.0036)	(0.0080)
- Living in south	0.0089	0.8793***	-0.0052	0.0031	0.0354***	-0.2397***
Germany	(0.0062)	(0.0345)	(0.0034)	(0.0042)	(0.0038)	(0.0156)
- Migrant	0.0336***	-1.7531***	-0.0832***	0.1286***	-0.1055***	0.0423***
	(0.0109)	(0.0905)	(0.0072)	(0.0119)	(0.0116)	(0.0158)
- Number of siblings	0.0476***	0.0144***	0.0019	-0.0056**	0.0037	-0.0815***
Aged 0-14	(0.0049)	(0.0041)	(0.0020)	(0.0027)	(0.0034)	(0.0051)
- Per-capita house-	0.0339***	0.5316***	0.0148***	-0.0510***	0.0070***	-0.1594***
hold income	(0.0024)	(0.0137)	(0.0014)	(0.0022)	(0.0021)	(0.0062)

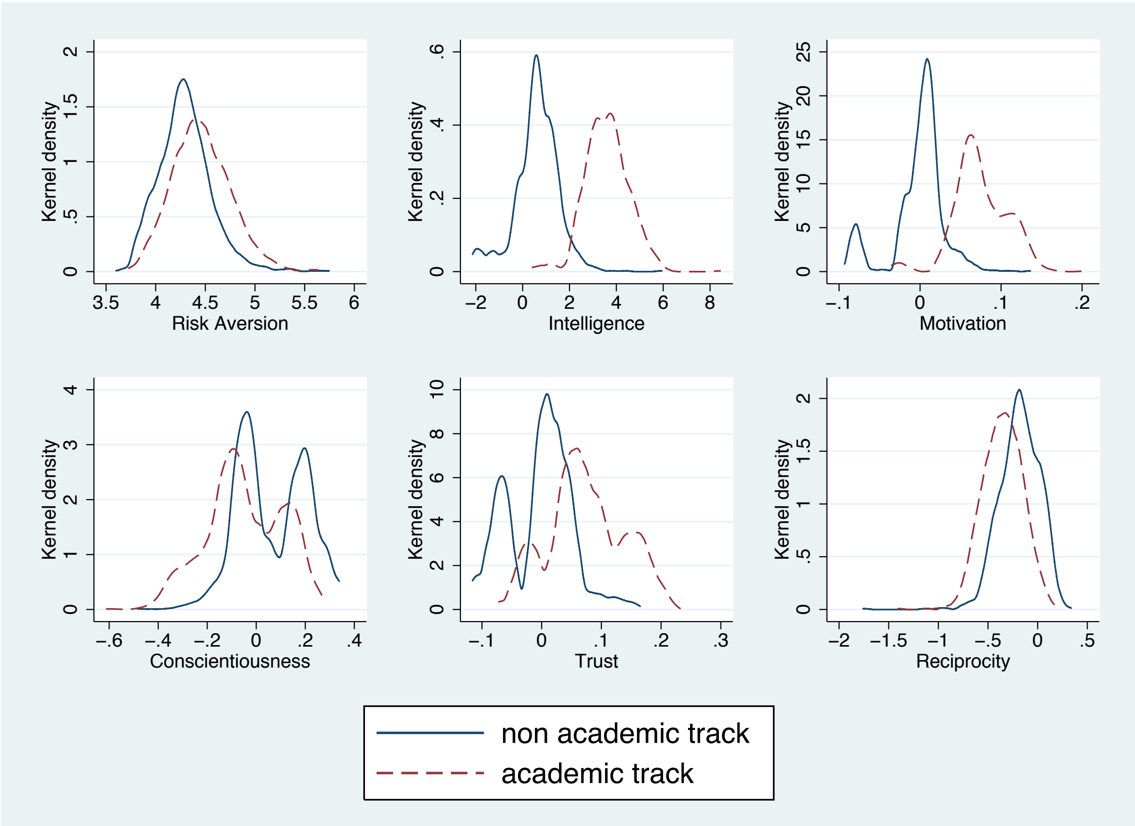
Note : Standard errors in parenthesis. Significant levels : *** 1% ; ** 5% ; * 10%. Sample size is 1784.

As noted earlier in the unconditional analysis, early tracking has a modest (although significant) effect on measured risk aversion. The effect, equal to 0.0261, represents less than 0.6% the average value of risk aversion.

Although this does not necessarily imply that risk aversion is independent of parental background variables, at least it suggests that the link between background variables and risk aversion may be different than the link between background variables and Intelligence and Motivation.

The regression estimates of Table 3 allows us to separate the effects of each variable for each specific factor. Father's and mother's education, as well as household income, are regressors that have attracted special attention in the human capital literature. For this reason, we examine their effects on Risk Aversion and Intelligence (two factors that raise particular interest for economists).

FIGURE 2.1 – Distribution of factors



Note : For each factor, the Kernel density function is represented, based on the estimated factors.

Our estimates indicate that both parental income, father's education, raise risk aversion and Intelligence, but that mother's education reduce risk aversion and raises Intelligence. At the opposite, being a female and having been raised in East-Germany reduces both risk aversion and cognitive skills. Finally, family size (number of siblings) raises risk aversion but has no impact on Intelligence.

In order to provide a global picture of the distribution of factors, we plot the distribution of risk aversion and psychological factors (Figure 2.1).

The figure indicates that the distributions of factors in the two sub-populations seem to have similar shapes, except for intelligence and motivation. One interesting pattern that emerges from these distributions is that the population of academic

TABLE 2.4 – Correlation between Factors

	Risk Aversion	Trust	Intel-ligence	Motiva-tion	Conscien-tiousness	Recipro-city
Risk Aversion	1.0000					
Trust	0.4648	1.0000				
Intelligence	0.2944	0.7723	1.0000			
Motivation	0.3582	0.8915	0.9124	1.0000		
Conscien.	-0.4638	-0.7024	-0.6526	-0.7151	1.0000	
Reciprocity	-0.4105	-0.3581	-0.5926	-0.3929	0.5729	1.0000

Note : The correlation matrix reports the coefficient of correlation between estimated factors.

enrollees seems to be more heterogenous than the non academic enrollees one. Indeed, the dispersion is larger and the density associated to the mode is lower. To picture the effect of the track after conditioning on background variables, we also include a plot of the distribution of factors obtained when the background variables are held constant. This plot is in Appendix 2.C.

2.3.2 The Correlation betweens Factors and Risk Aversion

We now investigate the degree of correlation between all factors and risk aversion. The correlations are found in Table 2.4.[16]

There is overwhelming evidence against the usual orthogonality assumption made about latent factors. Although our approach does not allow us to infer anything about the correlation between factors in early childhood or at birth, the results indicate that most factors appear to be highly correlated at age 17. Upon examining the correlations, it is relatively clear that Motivation, Intelligence and Trust form a single block of factors that are highly (positively) correlated. For instance, the correlation between Intelligence and Motivation is as high as 0.91. We also note that Reciprocity and Conscientiousness are positively related and tend to form a second

[16]The structure of correlation is very similar in the two subsamples (academics and non academics), so we just present it for the whole population.

block.

The correlations also point out to the specificity of risk aversion. Risk aversion is correlated positively with Intelligence, Motivation and Trust but those correlations (0.29, 0.36 and 0.46) are much weaker than the overall correlation between cognitive and non-cognitive factors. This result is at odds with the negative correlation between risk aversion and cognitive skills reported in **?**[17] Finally, we also find that risk aversion is negative correlated with Reciprocity and Conscientiousness (-0.41 and -0.46).

2.4 Concluding Remarks

Our results indicate that differences in education track (as measured at age 17) seem to translate into sizeable differences in Intelligence and Motivation. These differences are partly explained by the fact that educational track status is correlated with parental human capital (education and income) and that psychological traits are also highly correlated with the same variables. However, differences in Intelligence and Motivation according to track tend to persist after controlling for family background variables.

Another major finding is that risk aversion is only weakly correlated with educational track. This result is therefore consistent with the view that risk aversion (like the rate of time preference) is an economic primitive.

Our findings also suggest that the usual orthogonality assumption between factors is illusive. In particular, cognitive and non-cognitive factors are highly correlated while risk aversion is much less correlated with either cognitive or non-cognitive skills.

Finally, it should be noted that the weak correlation between risk aversion and tracking status is not incompatible with the existence of a causal effect of risk aver-

[17] However, our approach which is based on a structural representation of risk attitudes and which allows for measurement error, renders comparisons quite difficult.

sion on subsequent schooling choices such as the decision to enter higher education. For instance, differences in risk aversion between individuals who have been selected into the Academic track may still act as a fundamental determinants of individual choices if higher education entails facing a high level of risk. This is an issue that is currently the object of ongoing research.

2.A Psychometric Questions in the GSOEP

2.A.1 Locus of control

The following statements apply to different attitudes towards life and the future. To what degree to you personally agree with the following statements ?

Please answer according to the following scale : 1 means : “disagree completely”, 7 means : “agree completely”.

- How my life goes depends on me
- Compared to other people, I have not achieved what I deserve
- What a person achieves in life is above all a question of fate or luck
- If a person is socially or politically active, he/she can have an effect on social conditions
- I frequently have the experience that other people have a controlling influence over my life
- One has to work hard in order to succeed
- If I run up against difficulties in life, I often doubt my own abilities
- The opportunities that I have in life are determined by the social conditions

- Inborn abilities are more important than any efforts one can make
- I have little control over the things that happen in my life

2.A.2 Conscientiousness

Our every-day actions are influenced by our basic belief. There is very limited scientific knowledge available on this topic.

Below are different qualities that a person can have. You will probably find that some apply to you perfectly and that some do not apply to you at all. With others, you may be somewhere in between.

Please answer according to the following scale : 1 means "does not apply to me at all" ; 7 means "applies to me perfectly". With values between 1 and 7, you can express where you lie between these two extremes.

I see myself as someone who

- does a thorough job
- tends to be lazy
- does things effectively and efficiently

2.A.3 Trust

What is your opinion on the following three statements ?

Please answer according to the following scale : 1 means : “disagree completely”, 7 means : “agree completely”.

- On the whole one can trust people
- Nowadays one can’t rely on anyone
- If one is dealing with strangers, it is better to be careful before one can trust them

2.A.4 Positive Reciprocity

To what degree do the following statements apply to you personally ?

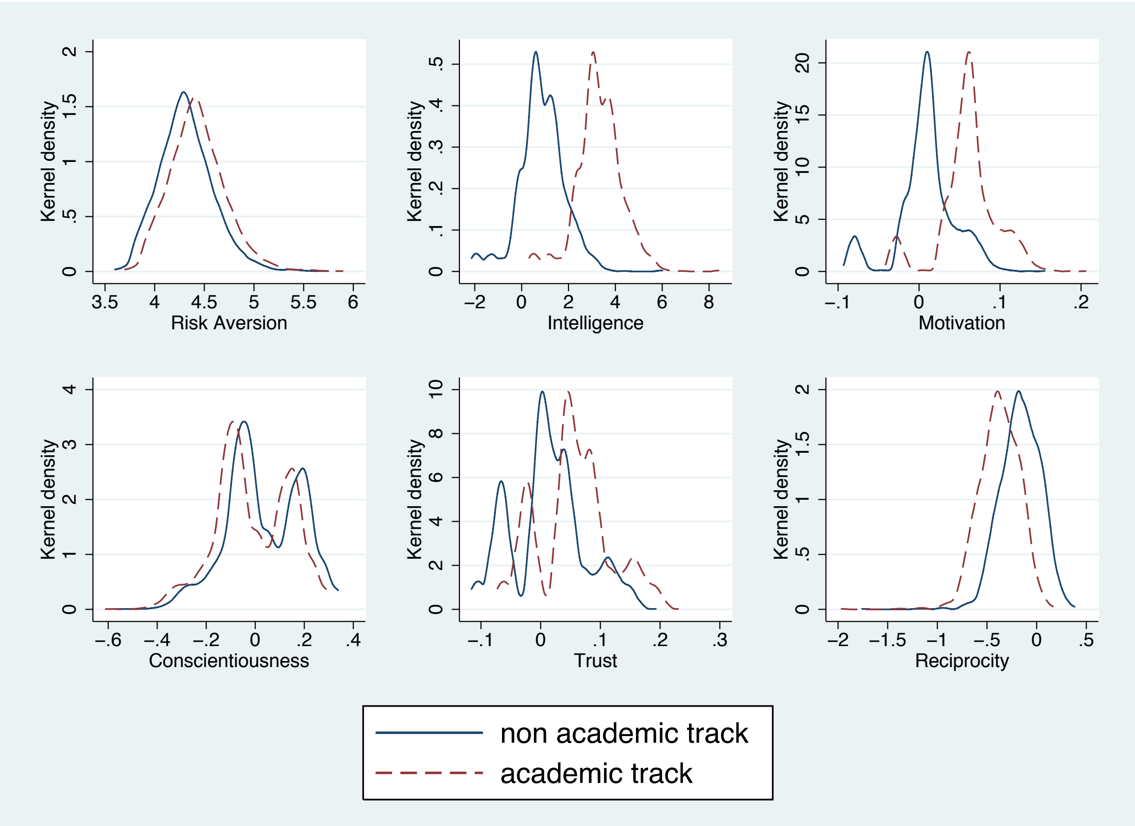
Please answer according to the following scale : 1 means : “does not apply to me at all”, 7 means : “applies to me perfectly”.

- If someone does me a favor, I am prepared to return it
- I go out of my way to help somebody who has been kind to me before
- I am ready to undergo personal costs to help somebody who helped me before

2.B Descriptive Statistics

2.C Counterfactual Distributions

FIGURE 2.2 – Counterfactual distribution of factors



Note : For each individual of the sample, the counterfactual factor associated to a given track is the value of the factor, considering that the individual belongs to the corresponding track, whatever his/her observed track. Then, the Kernel density function is represented for each factor.

Table 2.5 – Descriptive Statistics of the Variables used in the Analysis

	Whole population		Non academic track		Academic track	
	Mean	S.D.	Mean	S.D.	Mean	S.D.
Academic track*	0.4675	-	0.0000	-	1.0000	-
Female*	0.4849	-	0.4642	-	0.5084	-
Father finished *Gymnasium**	0.2091	-	0.0674	-	0.3705	-
Mother finished *Gymnasium**	0.1850	-	0.0716	-	0.3141	-
Living in east Germany*	0.2214	-	0.2242	-	0.2182	-
Living in south Germany*	0.2814	-	0.3116	-	0.2470	-
Migrant*	0.0706	-	0.1074	-	0.0288	-
Number of siblings aged 0-14	0.5443	0.8164	0.5726	0.8418	0.5120	0.7858
Per-capita household income	1.1369	0.6753	0.9888	0.6292	1.3056	0.6867
Individuals	1,784		950		834	

Note : Variables denoted with a * are indicator $(0, 1)$ variables. Per-capita household income is measured in tens of thousands of 2005 euros.

Table 2.6 – Correlation between Covariates

	Academic track	Female	Father Educ.	Mother Educ.	Living East	Living South	Migrant	Siblin 0-14
Academic track*	1,0000							
Female*	0,0441	1,0000						
Father finished *Gymn.**	0,3719	0,0197	1,0000					
Mother finished *Gymn.**	0,3117	0,0115	0,4935	1,0000				
Living in east Germ.*	-0,0072	-0,0095	-0,0451	-0,0141	1,0000			
Living in south Germ.*	-0,0717	0,0040	0,0063	-0,0092	-0,3337	1,0000		
Migrant*	-0,1531	0,0390	-0,1364	-0,1257	-0,1470	0,0756	1,0000	
Siblings aged 0-14	-0,0371	0,0003	0,0067	0,0343	-0,0496	0,0348	0,1003	1,000
Per-capita HH income	0,2341	-0,0221	0,3025	0,3013	-0,1461	0,0922	-0,1374	-0,224

Note : Variables denoted with a * are indicator $(0, 1)$ variables. Per-capita household (HH) income is measured in tens of thousands of 2005 euros.

Chapitre 3

Risk Tolerance and Life Cycle Income Profiles [1]

This paper investigates the relationship between risk attitudes (risk tolerance) and individual life cycle income profiles. Theoretically, individuals tends more to take risks in order to face higher income growth. As of now, most of the debates taking place in the earnings dynamics literature is centered on the relative importance that persistent shocks, and individual heterogeneity may play, although the theoretical foundations of these alternative views are practically never examined. In particular, literature remains silent about the relative importance of potential increases in returns to skills, and increases in risk exposure (ex-ante).
Using various waves of the Panel Study of Income Dynamics (PSID), we use individual measures of risk tolerance, to estimate a reduced-form model of individual income profiles, in which individual persistent heterogeneity is decomposed into two sources ; differences in risk attitudes and differences in residual heterogeneity.
We show that individual heterogeneity is crucial to study earnings profiles and that risk preferences has an explicit impact on individual labor income growth. Risk tolerance has a positive impact on income distribution, offering a wage premium around 2%. Risk attitudes seem determinant at early stage of careers but vanishes for oldest individuals.

Introduction

This paper investigates the relationship between risk attitudes (risk tolerance) and individual life cycle income profiles. The time series properties of individual income

[1]This chapter was written with Christian Belzil (Ecole Polytechnique, ENSAE, and IZA) and Marco Leonardi (University of Milan and IZA).

processes have been the object of much empirical work over the past 20 years. Most of the debates taking place in the earnings dynamics literature is centered on the relative importance that persistent shocks, and individual heterogeneity may play, although the theoretical foundations of these alternative views are practically never examined.[2]

In line with the documented increase in wage inequality experienced in the US economy (over a period going from the 1970's until the early 1990's), another branch of the literature has focussed on the degree of instability in earnings. Gottschalk and Moffitt (1994b) have documented an increased in the variance of the transitory components of earnings. Without assuming perfect financial markets, idiosyncratic labor market risk will affect individual labor income.

Because the literature on earnings dynamics has remained statistical in nature, it has ignored the nature of heterogeneity characterizing heterogenous income profiles. The effects of risk aversion and human capital risk on human capital investment is a long-standing question in economics. It goes back to Levhari and Weiss (1974) who examined theoretical conditions under which more risk averse agents invest less in human capital, and under which the rate of return to human capital exceeds the rate of return on safer assets. They show that when consumption (or earnings) risk increase with accumulated human capital, risk aversion is detrimental to investment.

As of now, there exist no empirical evidence on whether or not those conditions are fulfilled empirically, and more specifically, to what extent either schooling or post-schooling investments may depend on risk aversion.

The microeconometric literature has for the most part ignored the role of heterogeneity in risk aversion in important lifecycle decisions. Belzil and Leonardi (2013) have estimated the effect of measured risk aversion on schooling attainments in Italy using the bank of Italy panel data but were forced to use a measure of risk aversion

[2]Guvenen (2007) refers to age-earnings profile dominated by heterogeneity as Heterogenous income profiles (HIP) and those dominated by persistent shocks as Restricted Income Profile (RIP). Browning, Ejrneas, and Alvarez (2010) present an overview of the literature.

that was measured posterior to schooling investments.

Another strand of the microeconometric literature literature has focussed on separating heterogeneity from ex-ante risk. Carneiro, Hansen, and Heckman (2003) introduce factor models within the Willis and Rosen classical model of college attendance decision to identify counterfactual distributions and use their model to quantify the amount of information available to individuals at the time college decisions are made. However, to do so, Carneiro, Hansen, and Heckman (2003) assume logarithmic preferences and ignore the role of heterogeneity in risk aversion[3].

Using various waves of the Panel Study of Income Dynamics (PSID), we use individual measures of risk tolerance, to estimate a reduced-form model of individual income profiles, in which individual persistent heterogeneity is decomposed into two sources ; differences in risk attitudes and differences in residual heterogeneity. Because data on risk tolerance is only available in 1994, we estimate the distribution of risk tolerance conditional on some characteristics, as well as accumulated human capital, as of 1994. For the same reason, the distribution of residual unobserved heterogeneity affecting life-cycle income profiles is defined as conditional on what has been achieved by 1994.

The first section presents how we model and estimate individual risk tolerance. Risk tolerance is a direct transformation of risk aversion and is extracted through the lottery question of the PSID. The second part gives the statistical model of income profiles. In order to obtain parameter of interest, we then maximize the joint likelihood of measured individual tolerance levels and observed earnings path between 1994 and 2009. Results of estimation and conclusion are given afterwards.

We show that individual heterogeneity is crucial to study earnings profiles and that risk preferences has an explicit impact on individual labor income growth. Risk

[3]A similar approach is used in Cunha, Heckman, and Navarro (2005). Guvenen (2007) uses consumption choices to separate heterogeneity from risk in age earnings profiles. Attanasio and Kaufmann (2009) rely on subjective expectations to analyze the decision to invest into schooling using data from a household survey on Mexican junior and senior. Finally, Polachek, Das, and Thamma-Apiroam (2013) have introduced psychometric factors within a Ben-Porath model but ignore ex-ante risk.

tolerance has a positive impact on income distribution, offering a wage premium around 2%. Risk attitudes seem determinant at early stage of careers but vanishes for oldest individuals.

3.1 Modeling Risk Tolerance

3.1.1 The Lottery and CRRA Assumption

We use individuals' responses to questions about hypothetical risky choices in order to measure risk tolerance. The questions were asked in the 1996 wave of the PSID. As will become clear later, individuals have to choose between a job with a certain lifetime income, and a job with a random income with a higher mean. The questions are very similar to those introduced and analyzed by Barsky, Juster, Kimball, and Shapiro (1997). They are worded as follows :

"Suppose you had a job that guaranteed you income for life equal to your current, total income. And that job was [your/your family's] only source of income. Then you are given the opportunity to take anew, and equally good job, with a 50-50 chance that it will double your income and spending power. But there is a 50-50 chance that it will cut your income and spending power by a third. Would you take the new job ?"

Depending on the answer, the respondent is asked similar questions with job prospects that always double income with a 50 percent probability and cut income by a changing fraction $1-u$ (with $1-u$ equal to 10, 20, 50 or 75 percent, respectively). For example, if a participant answers "yes" to the first question (with an income loss of one third), the next question presents a scenario with a possible 50 percent cut in income. However, if the participant answers "no" to the first question, the income loss is reduced to just 20 percent in the next lottery question. The answers allow us to group individual in 6 categories, depending on how far the individual actually gets before rejecting the current offer.

As do Kimball, Sahm, and Shapiro (2008), we assume that individuals are endowed with a constant relative risk aversion (CRRA) utility function, over lifetime income. Precisely, we assume that :

$$U(y) = \frac{y^{1-\frac{1}{\theta}}}{1-\frac{1}{\theta}}$$

By assumption, the CRRA implies a constant (relative) degree of risk tolerance, $\theta = \frac{-U'}{U''y}$. Using the lottery structure of the PSID, individual responses may be used to bound individual risk tolerance (see Barsky, Juster, Kimball, and Shapiro (1997)).

Our empirical model therefore uses the fact that individual responses imply a range for risk tolerance, as opposed to point values, and that individual answers are likely to be measured with errors.

The risk tolerance writes :

$$\log\theta_i = \tau_i = \mu_0 + \mu' X_i + \varepsilon_i^\theta \qquad \text{with} \qquad \varepsilon_i^\theta \sim N(0, \sigma_\theta^2) \tag{3.1}$$

and where μ_0 is an intercept term, μ is a vector of parameter and X is a vector of individual specific regressors.

3.1.2 Risk Tolerance Categories

Because risk attitudes are measured after many individual choices have already been exercised, we interpret the distribution of risk tolerance as conditional on choices. The vector X includes schooling (in three categories : dropout, high school graduate and college graduate), age, number of children, marital status, region of living, unemployment rate and initial earnings in the labor market. Geweke and Keane (2000) show that early earnings can be strong predictors of later position, even conditioning on observed individual characteristics. The majority of variance in earnings are due to unobserved heterogeneity, this paper is a try to look deeper individual primitives of choices.

We assume that individual choices reported in the PSID are affected by effective (log) tolerance, $\tilde{\tau}_i$, which is the sum of true log tolerance, τ_i, and a classical measurement error term ω_i :

$$\tilde{\tau}_i = \tau_i + \omega_i \qquad \text{with} \qquad \omega_i \sim N(0,1) \tag{3.2}$$

The lottery question is a multimodal and sequential question, so that risk parameters only have finite numbers of category (see details in Appendix). Given this, the likelihood function of observed individual choices is formulate as follows. For a given category C_j of risk tolerance, defined by a low bound $\tau_{j,low}$ and $\tau_{j,high}$, the conditional (on ε_i^θ) probability of observing an individual, choosing category j, is

$$\Pr(\tau_{j,low} < \tilde{\tau}_i \leq \tau_{j,high} \mid \tau_i) = \Phi(\tau_{j,high} - \tilde{\tau}_i) - \Phi(\tau_{j,low} - \tilde{\tau}_i) \tag{3.3}$$

The unconditional probability is obtained after integrating out the distribution of ε_i^θ :

$$\Pr(i \in C_j) = \int_{-\infty}^{-\infty} [\Phi(\tau_{j,high} - \tau_i) - \Phi(\tau_{j,low} - \tau_i)] dH(\varepsilon_i^\theta) \tag{3.4}$$

where $H(.)$ is the distribution function of ε_i^θ. To estimate the model, we use a simulated version of this integral following usual numerical methods.

3.2 A Model of Life-Cycle Income Profiles

Our empirical model if life-cycle income profiles offers a compromise between the purely statistical representation used in the earlier literature on earnings dynamics, and a more economically interpretable specification in which the role of risk tolerance may be easily inferred.

The model has some similarities with the specification used by Guvenen (2007). It is composed of four different additively separable functions capturing i) a relationship between income and a polynomial inexperience common to everyone, and denoted

$f_1(h_t^i)$, ii) the effect of individual specific risk tolerance on earnings levels and growth rates, denoted $f_2(\tau^i, h_t^i)$ iii) the effect of individual specific residual heterogeneity on earnings levels and growth rates, and denoted $f_3(.) = \alpha_0^i + \alpha_1^i \cdot h_t^i$, and iv) the effect of calendar time (business cycle) common to everyone, denoted $f_4(t)$.

Formally, the income profile equation is given by :

$$y_{h,t}^i = f_1(h_t^i) + f_2(\tau^i, h_t^i) + f_3(\alpha_0^i, \alpha_1^i, h_t^i) + f_4(t) + \varepsilon_{h,t}^i \tag{3.5}$$

where

- $y_{h,t}^i$ is logarithm of labor earnings of individual i, at date t, with h years of labor market experience
- h_t^i is accumulated experience if i at date t
- $f_1(h_t^i) = \delta_0 + \delta_1 \cdot h_t^i + \delta_2 \cdot (h_t^i)^2 + \delta_3 \cdot (h_t^i)^3$ is a polynomial function of experience
- $f_2(\tau^i, h_t^i) = \beta_0 \cdot \tau^i + \beta_1 \cdot \tau^i \cdot h_t^i$ is a linear function in risk tolerance with interaction term with experience. This term is usually implicit in the litterature
- $f_3(\alpha_0^i, \alpha_1^i, h_t^i) = \alpha_0^i + \alpha_1^i \cdot h_t^i$ is an individual heterogeneity term (shifts in the intercept and slope parameters)
- $f_4(t)$ is a function of calendar time (yearly unemployment rate)
- $\varepsilon_{h,t}^i \sim N(0, \sigma_t^i)$ where $\sigma_t^i = \exp(\sigma_0 + \sigma_1 \cdot \tau^i + \sigma_2 \cdot h_t^i)$ is an heteroscedastic residual term

Individual heterogeneity on wage is modeled through the f_3 term and interaction between τ_i and h_t^i. This is where the heterogenous profile comes from : individual heterogeneity plays a direct role on income growth. For estimation reasons, we assume that the vector $\{\alpha_0^i, \alpha_1^i\}$ follows a bivariate distribution with $m = 1, 2, ...M$ points of support (types). Because of identification constraints, the mean of α_0^i and α_1^i are unrestricted, we normalize α_1^j to 0 for type $j = 1$.

As we do for risk tolerance, we model the distribution of risk tolerance as conditional on schooling, earnings growth rate, age and number of siblings. To do so we use unobserved heterogeneity *à la Heckman Singer* with M discrete types of support $\{\theta_m\}_{1\leq m\leq M}$. Following Keane and Wolpin (1997), we assume that the type m probability is given by :

$$\Pr(\theta_i = \theta_m) = \frac{\exp(p'_m X_i)}{1 + \exp(p'_m X_i)} = \Lambda_m(X_i) \tag{3.6}$$

where p_m is a vector of parameters to be estimated.

This specification is general enough to capture individual heterogeneity and estimate labor market characteristics. Guvenen's works show that this heterogenous income profile are consistent with data on consumption, that we do not have here. The flexibility on the risk aversion parameter is quite general in order to capture its interaction with individual incomes evolution[4].

To estimate this model, we form the likelihood of observed degree of risk tolerance ($\Pr(i \in C_j)$ and the likelihood of observed panel data on earnings, denoted $L(y^i_{h,1}, ...y^i_{h,T})$. Note that each component depends on the underlying (true) risk tolerance.

The density of observed earnings, $L(y^i_{h,1}, ...y^i_{h,T}, \theta_m)$, for a type m is :

$$\Pr(i \in C_j) \cdot \prod_{t=1}^{T} \frac{1}{\sigma^i_t} \phi[\frac{1}{\sigma^i_t} \cdot (y^i_{h,t} - f_1(h^i_t) + f_2(\tau^i, h^i_t) + f_3(\alpha^i_0, \alpha^i_1, h^i_t) + f_4(t))] \mid \theta_i = \theta_m]$$

The full likelihood is $L(.) = \prod_i \sum_{m=1}^{M} \Lambda_m(X_i) \cdot L(y^i_{h,1}, ...y^i_{h,T}, \theta_m)$

[4]We suppose schooling as given in the labor market : according toBelzil and Hansen (2007), tastes for schooling plays a small role for individual schooling choices.

3.3 Results

3.3.1 Sample description

We use a sample comparable to the usual literature of this field. From the PSID data from 1968 to 2009, we keep only white men, head of households, aged between 25 and 55 (13 % of the initial whole sample). We then restrict to reliable informations excluding inconsistent education, gender and earnings[5] variables. Keeping non missing risk aversion (lottery question) and never self-employed gives us a final sample of 1 139 individuals with 15 442 observations. Detailed statistics of sample selection are given step by step in appendix. We use actualized annual log wages in thousand dollars, in order to be comparable to other works with these data.

The following table incorporates information relating to risk parameters' categories.

Table 3.1 – Risk tolerance categories

Category	Risks considered		Risk Tolerance Bounds		Frequencies
	accepted	rejected	lower	higher	
1	none	0.10	0.00	0.13	254
2	0.10	0.20	0.13	0.27	254
3	0.20	0.33	0.27	0.50	187
4	0.33	0.50	0.50	1.00	188
5	0.50	0.75	1.00	3.27	161
6	0.75	none	3.27	∞	77

The pooled earnings profiles show that risk lovers face significantly higher incomes, especially in early careers (see figure 3.3). After 20 years of experience, differences statistically vanish but higher risk tolerance is still correlated with higher earnings. The story seems to be that initial conditions on the labor market, such as schooling, matters and in labor market evolutions, the lowest risk tolerance group catches up part of the difference.

[5]There is an issue how outliers in income are trimmed : Gottschalk and Moffitt (1994b) drop the first and last percentile and Meghir and Pistaferri (2004) trim people with large wage changes. We use both ways in this work.

TABLE 3.2 – Descriptives statistics by Risk Tolerance Group

Variables	Low RT	Medium RT	High RT
Mean Earnings	33467.13	36657.60	32298.36
College	21 %	30 %	29 %
Graduate	69 %	62 %	65 %
Dropout	10 %	7 %	6 %
Married	80 %	80 %	74 %
Age	33.09	32.68	31.72
Average Number of children	1.17	1.16	0.90

Note : Risk Tolerance categories are grouped in this table : *Low RT* means category 1 and 2 of table 1, *Medium RT* means category 3 and 4, *High RT* means category 5 and 6.

3.3.2 Model estimates

The model fits well the risk tolerance distribution according to the table 3.4, with an higher variance if risk tolerance is estimated with the full likelihood and not the unconditional one (see appendix page 111 for further details on unconditional risk tolerance estimation).

The figure 3.1 gives observed log-earnings in three groups of estimated risk tolerance (left panel) and the error term for each group (right panel). High risk tolerance is linked with higher income growth at early stages of career. Risk averse individuals (low risk tolerance group) has a lowest initial condition on the labor market but highest values after 30 years of working. And after some decades, between 10 and 20 years of experience, predicted profiles face highest error term. At the end of career, after 25 years of experience, this model with heterogeneous income profiles has much better fit.

3.3.3 Effect of the Estimated Risk Tolerance

Running a pooled regression with estimated values of risk tolerance indicates a significant and positive impact of risk preferences on income profiles (see table 3.10 in appendix), in direct line of model estimates (see table 3.8 in appendix). The R^2 of the regression of $\hat{y}$ on $\hat{tau}$ is 0.85 (0.90 with controls and human capital). With $\hat{\varepsilon}$, this is 0.55.

Table 3.3 – Pooled Mincer estimates (with survey risk tolerance)

Variables	Estimates
(Intercept)	-0.48^{***} (0.03)
Experience	0.12^{***} (0.00)
Experience^2	-0.002^{***} (0.0001)
College	0.83^{***} (0.03)
Graduate	0.31^{***} (0.02)
Number of children	0.02^{***} (0.00)
Non married	-0.20^{***} (0.01)
Medium Risk Tolerance	0.09^{***} (0.01)
High Risk Tolerance	0.08^{***} (0.01)
R^2	0.29
Adj. R^2	0.29
Num. obs.	15708

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Table 3.4 – Risk Tolerance Parameters

	Mean	Standard Deviation
Unconditional	-1.27	1.579
Estimated	-1.17	2.292

The table 3.5 gives the same coefficient for specific sample, at different age of individuals. At the beginning of carrer, risk tolerance increase labor income with a 2.5% effect and of 2.0% for 35-year old individuals. But at the end of the career, this impact vanishes. This shows that risk tolerance plays a positive role on income profiles, especially at early stage of career and the effect is decreasing.

This decreasing effect of risk tolerance is coherent with decreasing marginal returns of experience that are plotted in the figure 3.2. Lower risk tolerant group has higher and stable marginal returns of experience. Risk lovers experience the lowest experience returns that even have negative values. Comparing that to average value in the whole sample shows that marginal returns are less decreasing when taking into account risk preferences.

FIGURE 3.1 – Income Profiles by Risk group

(a) Observed (b) Predicted

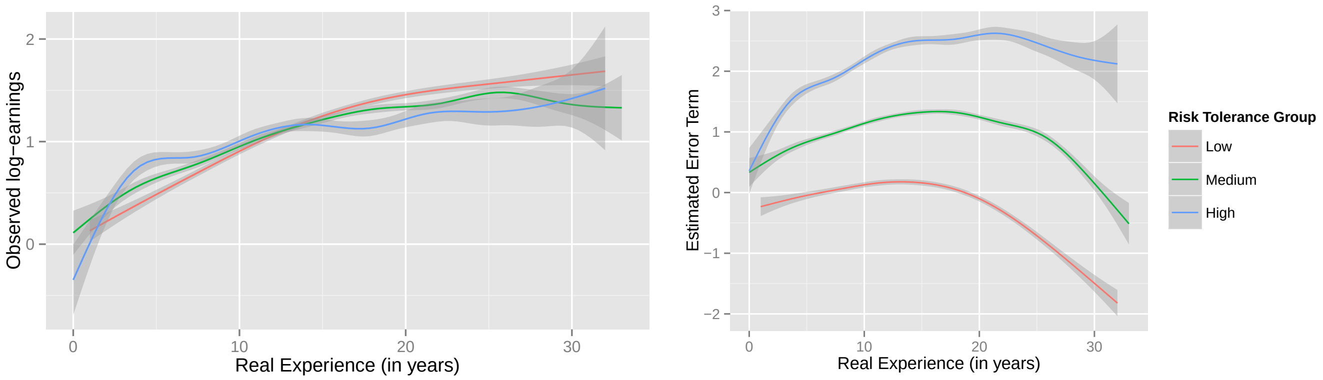


FIGURE 3.2 – Average Marginal Returns of Experience

(a) Whole sample (b) By Risk Tolerance Group

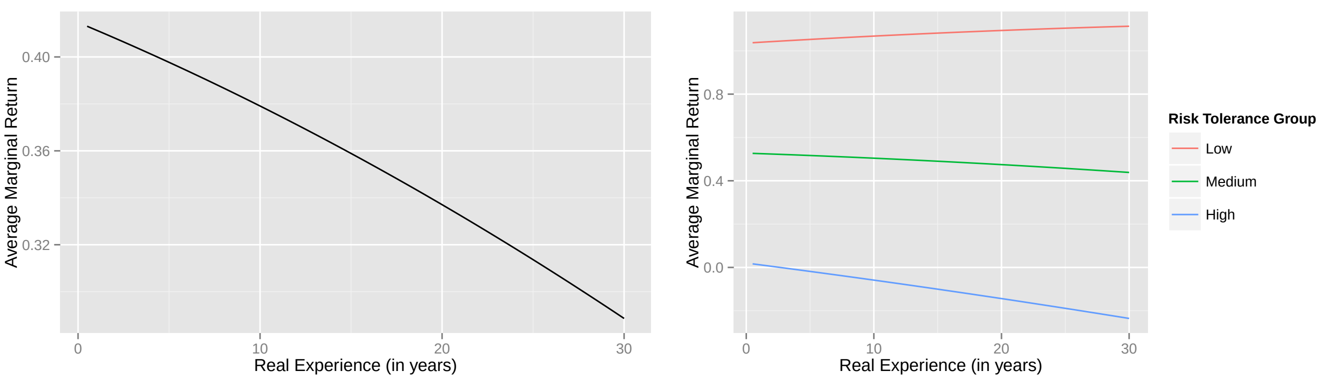


Table 3.5 – Mincer Regression in Levels at different age

Variables	Risk Tol. Coefficient	R^2
y_{25}	0.025*** (0.004)	0.06
y_{35}	0.020*** (0.004)	0.19
y_{45}	−0.012 (0.007)	0.20

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Concluding remarks

This paper uses various waves of PSID data to build a model of heterogeneous income profiles where risk tolerance play a specific role. Extracting the values trough a lottery question, we showed that individual heterogeneity is crucial to study earnings profiles and that risk preferences has an explicit impact on individual labor income growth. Risk tolerance has a positive impact on income distribution, offering a wage premium around 2%. Risk attitudes seem determinant at early stage of careers but vanishes for oldest individuals.

Further research has mainly two directions. First, one can add flexibility on the specification of the model, especially using an ARMA process for the error term of the income equations, closer to the restricted income profiles (RIP). A second major improvement would be to add a selection equation on self employment.

3.A Estimating Risk Tolerance From the PSID

Questions M1-M5 asked the extent of willingness to take jobs with different prospects. All choices are 50-50 chance to double income or to cut income in different proportions. According to answers to these questions, we can group people into 6 groups with exact risk tolerance range and 4 groups with larger ranges due to lack of answers in some of the questions.

It is usual in this literature to assume a CRRA utility function : $U(y) = \frac{y^{1-\frac{1}{\theta}}}{1-\frac{1}{\theta}}$. The goal is to estimate θ. If θ is also assumed as log-normally distributed conditional to individual heterogeneity (observed or not), then $\tau = \log(\theta)$ has a normal distribution. However, τ is unobservable. What we observe is $\tilde{\tau}$, which is in one of the ten groups determined by the design of the survey questions : $\tilde{\tau} \in C_j \quad \text{if} \quad \tau_{j,low} < \tilde{\tau} \leq \tau_{j,high}$.

The likelihood function of individual risk categories is then the product of each individuals probability of being in that particular group. We estimate unconditional mean μ and standard deviation σ by maximizing the likelihood function. And we then can recover θ for each group by computing the expected e^{τ} conditional on being in that group.

Without observed earnings and other data, the raw estimation gives bounds values that can be compared to the literature and used in the full model of labor market profiles. The estimated μ is -1.27 while the estimated σ is 1.579 (estimation on the whole sample[6]). The conditional mean of θ are thus calculated for each group :

Table 3.6 – Risk Categories and bounds estimations

Lottery group	Sample Size	HRS sample	$E[\theta \mid \tilde{\tau}]$	Range of θ
6	365 (6.5 %)	3.8 %	8.30	[3.27, ∞]
5	760 (13.5 %)	7.1 %	1.76	[1, 3.27]
4	829 (14.8 %)	14.2 %	0.71	[0.5, 1]
3	861 (15.3 %)	12.9 %	0.37	[0.26, 0.5]
2	1009 (17.0 %)	17.4 %	0.19	[0.13, 0.26]
1	1741 (31 %)	44.5 %	0.06	[0, 0.13]

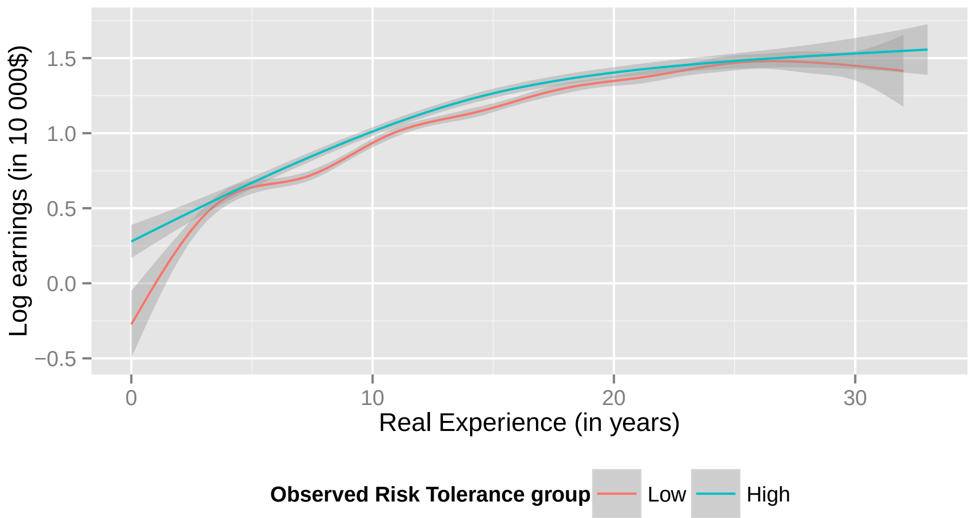
3.B From Initial sample to proper data

[6] We also compare to HRS Wave II sample, where the lottery question is similar.

TABLE 3.7 – Step by step sample selection (size and number of individuals)

Selected variable	Sample size	Observations deleted	Individuals deleted
Initial	60 705	0 %	0
Head of households	23 962	53 %	35 743
Age between 25 and 55	19 076	29 %	4 886
White Men	7 809	50 %	11 267
Non missing Education	7 496	3 %	313
Never Self-employed	5 348	38 %	2 148
Age < 45 in 1996	1 413	62 %	3 935
Other non missing variables	1 139	18 %	274

FIGURE 3.3 – Earnings by Risk Tolerance group (survey question)



Lecture : Red curve (below) is the lowest risk tolerance group ; Green curve (above) is the highest risk tolerance group.

3.C Full Model Estimates

TABLE 3.8 – Full model estimates : Risk tolerance parameters

Variables	Full
μ_0	10.0351 (0.00403)
$\mu_{college}$	−2.87419 (0.08564)
$\mu_{graduate}$	−3.19853 (0.20351)
μ_{age}	−1.88418 (<10e−5)
$\mu_{children}$	−2.01802 (6e−05)
$\mu_{nonmarried}$	0.66747 (<10e−5)
$\mu_{northcentral}$	1.61327 (0.20516)
μ_{south}	−1.07159 (0.06536)
μ_{west}	−1.02802 (0.00163)
$\mu_{initial_inc}$	−0.59965 (<10e−5)
σ_θ	2.292 (<10e−5)

TABLE 3.9 – Full model estimates : Income equations parameters

Equation	Parameter	Full	Basic
f_1	δ_0	1.61801 (<10e−5)	1.64204 (0.23925)
f_1	δ_1	−1.61929 (<10e−5)	−0.92411 (0.13838)
f_1	δ_2	−1.07201 (0.00467)	−0.39058 (0.16775)
f_1	δ_3	0.9012 (0.0987)	−0.33491 (0.20171)
f_2	β_{01}	−0.9613 (0.01541)	−0.8193 (0.09018)
f_2	β_{11}	−1.05244 (0.32057)	−0.34605 (0.67759)
f_3	α_0^1	0.2529 (0.00813)	0.148 (0.1549)
f_3	α_1^1	0 (.)	0 (.)
f_3	α_0^2	−7.20121 (0.06562)	−3.24893 (0.30539)
f_3	α_1^2	−3.80005 (0.16358)	.
f_4	γ	0.89999 (0.17828)	0.18032 (0.30585)
Var.	σ_0	−2.33338 (0.02064)	−1.64159 (0.15913)
Var.	σ_1	−5.13482 (0.01966)	.
Var.	σ_2	−0.09956 (0.04278)	1.91945 (0.00862)
Covariates	μ	Yes	No

TABLE 3.10 – Pooled Mincer estimates with Estimated Risk Tolerance

Variables	Estimates
(Intercept)	−0.47*** (0.03)
Experience	1.14*** (0.03)
Experience^2	−0.021*** (0.01)
College	0.90*** (0.02)
Graduate	0.37*** (0.02)
Estimated Risk Tolerance	0.02*** (0.00)
R^2	0.30
Control	Yes
Num. obs.	15708

$^{***}p < 0.001$, $^{**}p < 0.01$, $^{*}p < 0.05$

Bibliographie

Acemoglu, D., and D. Autor (2011) : *Skills, Tasks and Technologies : Implications for Employment and Earnings*vol. 4 of *Handbook of Labor Economics*, chap. 12, pp. 1043–1171. Elsevier.

Adda, J., C. Dustmann, C. Meghir, and J.-M. Robin (2013) : "Career Progression, Economic Downturns, and Skills," NBER Working Papers 18832, National Bureau of Economic Research, Inc.

Almlund, M., A. L. Duckworth, J. J. Heckman, and T. Kautz (2011) : "Personality Psychology and Economics," IZA Discussion Papers 5500, Institute for the Study of Labor (IZA).

Altonji, J. G., E. Blom, and C. Meghir (2012) : "Heterogeneity in Human Capital Investments : High School Curriculum, College Major, and Careers," NBER Working Papers 17985, National Bureau of Economic Research, Inc.

Amthauer, R., B. Brocke, D. Liepmann, and A. Beauducel (2001) : "Intelligenz-Struktur-Test 2000 R," *Göttingen : Hogrefe*, 2.

Arcidiacono, P. (2004) : "Ability sorting and the returns to college major," *Journal of Econometrics*, 121(1-2), 343–375.

Arcidiacono, P., V. J. Hotz, and S. Kang (2012) : "Modeling college major choices using elicited measures of expectations and counterfactuals," *Journal of Econometrics*, 166(1), 3 – 16, Annals Issue on "Identification and Decisions", in Honor of Chuck Manski's 60th Birthday.

Attanasio, O., and K. Kaufmann (2009) : "Educational Choices, Subjective Expectations, and Credit Constraints," NBER Working Papers 15087, National Bureau of Economic Research, Inc.

Autor, D. (2011) : "The polarization of job opportunities in the U.S. labor market : implications for employment and earnings," *Community Investments*, (Fall), 11–16, 40–4.

——— (2014) : "Polanyi's Paradox and the Shape of Employment Growth," NBER Working Papers 20485, National Bureau of Economic Research, Inc.

Barsky, R. B., F. T. Juster, M. S. Kimball, and M. D. Shapiro (1997) : "Preference Parameters and Behavioral Heterogeneity : An Experimental Approach in the Health and Retirement Study," *The Quarterly Journal of Economics*, 112(2), 537–79.

Bean, C., and C. Pissarides (1993) : "Unemployment, consumption and growth," *European Economic Review*, 37(4), 837–854.

Becker, G. S. (1962) : "Investment in Human Capital : A Theoretical Analysis," *Journal of Political Economy*, 70, 9.

Beffy, M., D. Fougère, and A. Maurel (2012) : "Choosing the Field of Study in Postsecondary Education : Do Expected Earnings Matter ?," *The Review of Economics and Statistics*, 94(1), 334–347.

Belzil, C. (2007) : "The return to schooling in structural dynamic models : a survey," *European Economic Review*, 51(5), 1059–1105.

Belzil, C., and J. Hansen (2002) : "Unobserved Ability and the Return to Schooling," *Econometrica*, 70(5), 2075–2091.

Belzil, C., and J. Hansen (2007) : "A structural analysis of the correlated random coefficient wage regression model," *Journal of Econometrics*, 140(2), 827–848.

Belzil, C., and M. Leonardi (2013) : "Risk Aversion and Schooling Decisions," *Annales d'Economie et de Statistique*, (111-112), 3.

Ben-Porath, Y. (1967) : "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, 75, 352.

Borghans, L., A. L. Duckworth, J. J. Heckman, and B. ter Weel (2008) : "The Economics and Psychology of Personality Traits," *Journal of Human Resources*, 43(4).

Browning, M., M. Ejrneas, and J. Alvarez (2010) : "Modelling Income Processes with Lots of Heterogeneity," *Review of Economic Studies*, 77(4), 1353–1381.

Cameron, S. V., and J. J. Heckman (2001) : "The Dynamics of Educational Attainment for Black, Hispanic, and White Males," *Journal of Political Economy*, 109(3), 455–499.

Card, D. (1999) : "The causal effect of education on earnings," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3 of *Handbook of Labor Economics*, chap. 30, pp. 1801–1863. Elsevier.

——— (2001) : "Estimating the Return to Schooling : Progress on Some Persistent Econometric Problems," *Econometrica*, 69(5), 1127–60.

Carneiro, P., K. T. Hansen, and J. J. Heckman (2003) : "Estimating Distributions of Treatment Effects with an Application to the Returns to Schooling and Measurement of the Effects of Uncertainty on College Choice," 44(2), 361–422.

Carneiro, P., J. J. Heckman, and E. J. Vytlacil (2011) : "Estimating Marginal Returns to Education," *American Economic Review*, 101(6), 2754–81.

Cunha, F., J. Heckman, and S. Navarro (2005) : "Separating uncertainty from heterogeneity in life cycle earnings," *Oxford Economic Papers*, 57(2), 191–261.

Cunha, F., and J. J. Heckman (2008) : "Formulating, Identifying and Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Journal of Human Resources*, 43(4).

Cunha, F., J. J. Heckman, and S. M. Schennach (2010) : "Estimating the Technology of Cognitive and Noncognitive Skill Formation," *Econometrica*, 78(3), 883–931.

Dohmen, T., A. Falk, D. Huffman, and U. Sunde (2010) : "Are Risk Aversion and Impatience Related to Cognitive Ability ?," *American Economic Review*, 100(3), 1238–60.

——— (2012) : "The Intergenerational Transmission of Risk and Trust Attitudes," *Review of Economic Studies*, 79(2), 645–677.

Dohmen, T., A. Falk, D. Huffman, U. Sunde, J. Schupp, and G. G. Wagner (2011) : "Individual Risk Attitudes : Measurement, Determinants, And Behavioral Consequences," *Journal of the European Economic Association*, 9(3), 522–550.

Geweke, J., and M. Keane (2000) : "An empirical analysis of earnings dynamics among men in the PSID : 1968-1989," *Journal of Econometrics*, 96(2), 293–356.

Gottschalk, P., and R. Moffitt (1994a) : "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 25(2), 217–272.

——— (1994b) : "The Growth of Earnings Instability in the U.S. Labor Market," *Brookings Papers on Economic Activity*, 25(2), 217–272.

Guvenen, F. (2007) : "Learning Your Earning : Are Labor Income Shocks Really Very Persistent ?," *American Economic Review*, 97(3), 687–712.

Hanushek, E. A., L. Woessmann, and L. Zhang (2011) : "General Education, Vocational Education, and Labor-Market Outcomes over the Life-Cycle," NBER Working Papers 17504, National Bureau of Economic Research, Inc.

Hartog, J., and L. Díaz-Serrano (2007) : "Earnings risk and demand for higher education : A cross-section test for Spain," *Journal of Applied Economics*, 0, 1–28.

Heckman, J., and B. Singer (1984) : "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data," *Econometrica*, 52(2), 271–320.

Heckman, J. J. (1990) : “Varieties of Selection Bias,” *American Economic Review*, 80(2), 313–18.

Heckman, J. J. (2008) : “Econometric Causality,” *International Statistical Review*, 76(1), 1–27.

Heckman, J. J., J. E. Humphries, S. Urzua, and G. Veramendi (2010) : “The effects of educational choices on labor market, health, and social outcomes,” Working Papers 2011-002, Human Capital and Economic Opportunity Working Group.

Heckman, J. J., and T. Kautz (2012) : “Hard evidence on soft skills,” *Labour Economics*, 19(4), 451–464.

Heckman, J. J., L. J. Lochner, and P. E. Todd (2006a) : *Earnings Functions, Rates of Return and Treatment Effects : The Mincer Equation and Beyond*vol. 1 of *Handbook of the Economics of Education*, chap. 7, pp. 307–458. Elsevier.

——— (2006b) : *Earnings Functions, Rates of Return and Treatment Effects : The Mincer Equation and Beyond*vol. 1 of *Handbook of the Economics of Education*, chap. 7, pp. 307–458. Elsevier.

Heckman, J. J., J. Stixrud, and S. Urzua (2006) : “The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior,” *Journal of Labor Economics*, 24(3), 411–482.

Kasahara, H., and K. Shimotsu (2009) : “Nonparametric Identification of Finite Mixture Models of Dynamic Discrete Choices,” *Econometrica*, 77(1), 135–175.

Keane, M. P., and K. I. Wolpin (1997) : “The Career Decisions of Young Men,” *Journal of Political Economy*, 105(3), 473–522.

——— (2001) : “The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment,” *International Economic Review*, 42(4), 1051–1103.

Kimball, M. S., C. R. Sahm, and M. D. Shapiro (2008) : “Imputing Risk Tolerance From Survey Responses,” *Journal of the American Statistical Association*, 103(483), 1028–1038.

Levhari, D., and Y. Weiss (1974) : “The Effect of Risk on the Investment in Human Capital,” *American Economic Review*, 64(6), 950–63.

Magnac, T., and D. Thesmar (2002a) : “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, 70(2), 801–816.

——— (2002b) : “Identifying Dynamic Discrete Decision Processes,” *Econometrica*, 70(2), 801–816.

Malamud, O. (2010) : “Breadth versus Depth : The Timing of Specialization in Higher Education,” *LABOUR*, 24(4), 359–390.

Meghir, C., and L. Pistaferri (2004) : "Income Variance Dynamics and Heterogeneity," *Econometrica*, 72(1), 1–32.

Mincer, J. (1958) : "Investment in Human Capital and Personal Income Distribution," *Journal of Political Economy*, 66, 281.

——— (1994) : "The Production of Human Capital and the Lifecycle of Earnings : Variations on a Theme," NBER Working Papers 4838, National Bureau of Economic Research, Inc.

Mincer, J. A. (1974) : *Schooling, Experience, and Earnings*, no. minc74-1 in NBER Books. National Bureau of Economic Research, Inc.

Mortensen, D. T., and C. A. Pissarides (1994) : "Job Creation and Job Destruction in the Theory of Unemployment," *Review of Economic Studies*, 61(3), 397–415.

Osborne, M., H. Gintis, and S. Bowles (2001) : "The Determinants of Earnings : A Behavioral Approach," *Journal of Economic Literature*, 39(4), 1137–1176.

Polachek, S., T. Das, and R. Thamma-Apiroam (2013) : "Heterogeneity in the Production of Human Capital," IZA Discussion Papers 7335, Institute for the Study of Labor (IZA).

Roy, A. D. (1951) : "Some thoughts on the distribution of earnings," *Oxford Economic Papers*, 3(2), 135–146.

Spence, A. M. (1973) : "Job Market Signaling," *The Quarterly Journal of Economics*, 87(3), 355–74.

Willis, R. J., and S. Rosen (1979) : "Education and Self-Selection," *Journal of Political Economy*, 87(5), S7–36.