THÈSE de DOCTORAT

de l'UNIVERSITÉ TOULOUSE CAPITOLE

Présentée et soutenue par

Monsieur Gyung Mo KIM

Le 23 juin 2025

Essais sur les modèles structurels d'investissement éducatif et de choix professionnels

École doctorale : **TSE Toulouse Sciences Economiques**

Spécialité : **Sciences Economiques - Toulouse**

Unité de recherche : **TSE-R - Toulouse School of Economics - Recherche**

Thèse dirigée par *Monsieur Thierry MAGNAC et Olivier DE GROOTE*

Composition du jury

Rapporteur : M. Arnaud MAUREL
Rapporteur : M. Dieter VERHAEST
Examinateur : M. Nicolas PISTOLESI
Directeur de thèse : M. Thierry MAGNAC
Co-encadrant de thèse : M. Olivier DE GROOTE

Essays on Structural Models of Educational Investment and Occupational Choices

Gyung-Mo Kim

Acknowledgements

It has been a privilege to be a student of my advisors. I am deeply grateful to Prof. Olivier De Groote for his unwavering support throughout my doctoral journey. From our regular Zoom meetings during the difficult COVID-19 period to his steady mentorship over the years, his guidance has helped me grow not only as a researcher but also as a more organized and independent thinker. I am also sincerely thankful to Prof. Thierry Magnac, who has been a role model of respectful and thoughtful scholarship. His gentle manner, along with his insightful comments and intellectual guidance, were instrumental in helping me find my place in the literature and advance my research.

My heartfelt thanks go to Prof. François Poinas, who generously coached me on how to present my work during the early stages of my PhD and offered me the valuable opportunity to contribute to a research project. I would also like to thank Prof. Matteo Bobba, Prof. Nour Meddahi, Prof. Christian Bontemps, and all the faculty who engaged with my work during workshops and the job market season, providing thoughtful and constructive feedback.

Beyond formal mentorship, I am especially indebted to Prof. Jihyun Kim, whose advice—always shared in a relaxed and open manner—helped guide me through difficult decisions. I will always remember the Chinese fish soup we shared and the grounding comfort of those moments. I am also thankful to Prof. Doh-Shin Jeon for inviting me to his gatherings and offering emotional support during my time in Toulouse.

My research was generously supported by TSE and the Fondation Jean-Jacques Laffont, whose scholarships and grants enabled me to attend international conferences and undertake a research visit to the University of Wisconsin–Madison. I am deeply grateful to Prof. Chao Fu for hosting me and to the other people with whom I interacted during that time.

Living in Toulouse since 2018, belonging to an absolute minority group, sometimes made me feel like an outsider. I was fortunate, however, to share this PhD journey with friends like Young Kwang and Chanwoo, whose companionship meant the world to me. I would also like to thank Anaïs, Lisa, Max, Mudit, Lea, Jonas, Oscar, Alfonso, and many other fellow students for making the journey such a meaningful, joyful, and memorable experience.

Pursuing a PhD abroad was far more challenging than I had anticipated. I would not have endured without the constant support of my family, Juri Kim, Hyangra Kim, and Chungu Kim, who have always believed in me. I also pay tribute to the grandparents I lost during my time in France—my memories of them have remained a source of strength. And while I hesitate to name her, I am deeply grateful and sorry to someone very special whose cheer and steadfast support sustained me for a long time, five years, and when I felt most struggling.

Reflecting on my time in France, I am filled with gratitude—not only for the academic growth and achievements, but also for the personal trials that taught me about my limits, faults, and passions. As I leave the warm, greenhouse-like environment of campus life and step forward as an independent researcher, I do so with both worry and excitement. It was a privilege to come to this stage and have this closing moment of the PhD, thanks to and with those I cherish deeply.

Summary

This doctoral thesis consists of three chapters that study the mechanisms of educational investment and occupational choices. In the first chapter, I provide structural estimates of a dynamic discrete choice model of education and labor market decisions and study heterogeneity in returns to higher education. The model explicitly accounts for different alternatives for higher education and the mismatch between the education level attained and that required for jobs, called overeducation. In the segmented labor markets, overeducation is the option inherent in college education, providing college graduates with access to the non-college sector, but not the other way around. A conditional choice probability (CCP)-based estimation method enables the model to be identified and estimated using recent Korean data, which is a short panel without assumptions about far into the later lifecycle not observed in the data. The estimation result shows that overeducation makes the wages among college graduates more dispersed, while the option secures their employment. A counterfactual policy of more stringent academic-ability-based sorting into vocational and general high schools is found to reduce academic college graduation driven by its consumption value, with those directed to vocational education lowering the overeducation rate and yielding larger total log wage income by their late twenties.

In the second chapter, my co-authors and I study educational investment decisions when students are uncertain about their own ability and preferences. We estimate a dynamic discrete choice model in which we allow agents to be uncertain about their own unobserved type. We estimate the degree of uncertainty by exploiting commonly available data on agents' self-reported most likely outcome. We apply this model to the Korean context and find large uncertainty before important high school track and effort decisions are made, particularly among the (overconfident) low-ability students. Providing more information would lead to stronger sorting as it disincentivizes low-ability students to exert effort and choose academically focused tracks, while it incentivizes high-ability students more. We also find effects in the long run, leading them to more different types of college degrees.

In the last chapter, I study the uncertainty about workers' own labor market productivity, which could have crucial implications for occupational sorting and wage dynamics. I start by documenting the key features in the US labor market that can be explained by the presence of uncertainty about workers' own skills and the learning process that unfolds depending on specific experiences. A structural model is built on these empirical patterns and characterizes that workers in their initial occupations decide each year whether to stay or quit in response to the wage outcomes while accumulating and learning about their skills. The dynamic discrete choice model, extending conventional discrete-time duration analysis, is estimated using the Kalman filter and a conditional choice probability estimator. I find supporting evidence that workers experience significant uncertainty about their skills. The results depict that the difference in the dynamics of wage distribution across occupations can be attributed to workers' sorting as well as human capital accumulation in different occupations. Counterfactual analysis shows that, in occupations with non-trivial probabilities of mismatch due to the choices based not on the true skills but on the beliefs, an information provision policy has a sorting effect. By reducing the share of the mismatched, the policy influences employment duration in the initial occupations, resulting in higher wage growth and lower wage dispersion.

Contents

Chapter 1.

Educational Choices and Heterogeneous Returns to Higher Education in the Segmented Labor Market

Gyung-Mo Kim

Abstract: This paper provides structural estimates of a dynamic discrete choice model of education and labor market decisions and studies heterogeneity in returns to higher education. The model explicitly accounts for different alternatives for higher education and the mismatch between the education level attained and that required for jobs, called overeducation. In the segmented labor markets, overeducation is the option inherent in college education, providing college graduates with access to the non-college sector, but not the other way around. A conditional choice probability (CCP)-based estimation method enables the model to be identified and estimated using recent Korean data, which is a short panel without assumptions about far into the later lifecycle not observed in the data. The estimation result shows that overeducation makes the wages among college graduates more dispersed, while the option secures their employment. A counterfactual policy of more stringent academic-ability-based sorting into vocational and general high schools is found to reduce academic college graduation driven by its consumption value, with those directed to vocational education lowering the overeducation rate and yielding larger total log wage income by their late twenties.

1.1 Introduction

Knowing the effect of education on wages or labor market productivity is a major concern to policymakers. A vast amount of worldwide evidence suggests average positive returns to higher education, which is in line with the increased investment in and expansion of the education sector. Relatively recent literature documents substantial heterogeneity in returns to higher education, delivering lower or even negative returns for some individuals (e.g., Carneiro et al. 2011; Rodríguez et al. 2016). One of the factors contributing to the heterogeneity would be the choices made in the labor market, such as blue- or white-collar jobs (Keane and Wolpin 1997) and occupations with different tasks (Yamaguchi 2012). One objective of this paper is to empirically study "overeducation" as a source of heterogeneous returns. Overeducation, a type of mismatch, defines the condition of having a higher education level than that required to adequately perform a specific job (McGuinness 2006).[1] The use of the terminology first in Richard Freeman's book "The Overeducated American" (1976) was motivated by an excessive supply of college graduates, which could reduce the returns to investing in a college education.[2] It is empirically supported that a large fraction of college graduates work in overeducated positions in the labor market and experience a significant wage penalty (e.g., Korpi and Tåhlin 2009; Barnichon and Zylberberg 2019).

The returns heterogeneity and increased wage inequality among college graduates, along with an increasing number of high school graduates seeking a college degree, raised the debate on whether too many students are going to college. From a policy perspective, particularly in countries where higher education is heavily subsidized, it may be important to understand the economic returns that students going to college expect and actually gain from the investment. The distinction is attributed to the sequential nature and uncertainties inherent in the decision-making process in which *ex-ante* expectation matter. Additionally, the increased demand for college education may stem not only from economic returns, but

[1] The terminology does not deny education functions to improve the quality of life in various aspects in addition to the role of supplying labor equipped with a certain skill level to the workplaces. Overeducation pays attention to the aspect of mismatch condition in the labor market between education level and jobs.

[2] Freeman's idea about the excessive supply of college graduates compared to labor demand became less concerned in accordance with changes in demand favoring highly skilled workers due to the skill-biased technological change (e.g. Katz and Murphy 1992; Katz et al. 1999).

also from non-pecuniary benefits associated with consuming college education and from the inherent option value for graduates to access the non-college sector when encountering difficulty in finding a job in the college sector or if such positions do not pay well enough. In this regard, I contribute to the recent debate by quantitatively studying heterogeneous economic returns to college education and the relative importance of different factors in demanding college education despite the possibility of being overeducated.

I apply dynamic discrete choice (DDC) models which are frequently used to study the mechanisms determining human capital investment and predict the effects of counterfactual policies (Keane et al. 2011; Blundell 2017). The overeducation concept is framed by assuming the segmented labor markets where college graduates can choose to work in the sector with jobs not requiring a college degree (i.e., non-college sector). This operationalizes the option that college graduates have access to the non-college sector, while, contrarily, workers without a college education do not have access to the college sector.[3] When individuals make educational decisions, they take into account the value of this overeducation option embedded in college graduation, which depends on quantities associated with uncertainties and risk aversion (Stange 2012; Trachter 2015). Lastly, the model in this paper assumes individuals do not choose the alternative that yields the highest expected lifetime earnings but lifetime utility. Besides the standard trade-off between time invested in producing academic capital and in accumulating labor market experience with immediate pecuniary payoff, the non-pecuniary utilities in colleges and workplaces are taken into account in the model.[4]

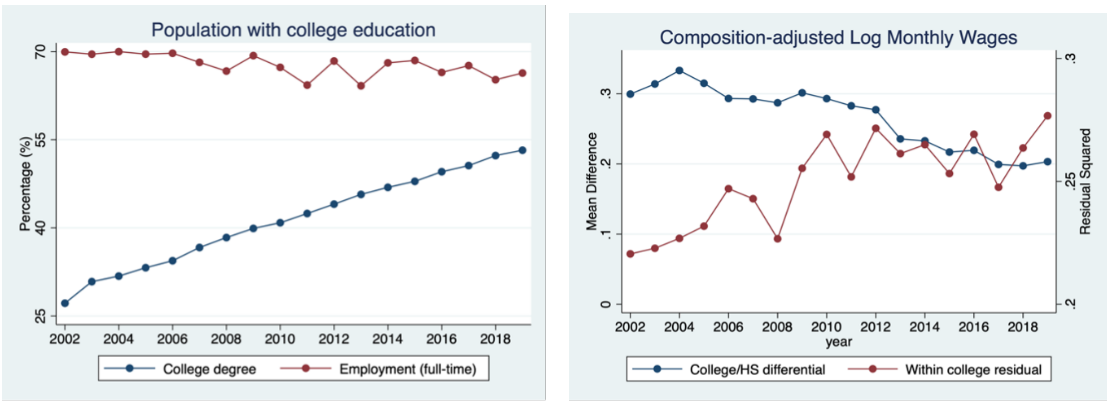
The empirical interest of this paper is in South Korea, the country which experienced the rapid expansion of the higher education sector with the highest proportion of those aged 25-34 attaining higher education worldwide (70% in 2018, OECD). Figure 1.1 shows that the share of individuals between the ages of 25 and 60 who graduated from a college sharply increased from 27% in 2002 to 52% in 2018. Meanwhile, their employment rate did

[3]The idea is that employers may use education as a screening device (for unobserved characteristics). Developing a segmented labor market model that theoretically formalizes overeducation, with workers without a college education crowded out of the college sector, is beyond the scope of this document.

[4]Admittedly, not every channel relevant to the students' choice and overducation is not explicitly modelled so that computational tractability is achieved. Channels that are treated as exogenous and in a reduced-form way in the model include, for example, search frictions and probabilistic job offers that constrain some college graduates from having a job offer in the college sector. Marriage/fertility concerns and the labor demand effects are also considered exogenous factors.

not decrease significantly. The figure suggests that workers with a college degree received about 30 percent more on average than those whose education ended with high school or lower, but the premium decreased as the supply of college-educated potential workforce increased. In conjunction with the rise in within-group residual wage dispersion of college-educated workers, it can be inferred that there could be individuals who could have been financially better off by not going to a college.[5] Furthermore, direct evidence, for example, in Cho (2014), Lee et al. (2016), and the news article in Financial Times (Christian Oliver 2010), suggests 20∼30% of overeducation in Korea. Studying the Korean context may have implications for other countries where higher education is expanding with a rising college enrollment rate, which has been observed in many developed and developing countries over the past few decades.

Figure 1.1: [Left] share of college graduates and their employment rate, [Right] college premium and residual wage variance of college-educated workers



** Source: Economically Active Population Survey, composition-adjusted college premium and the within-group residual are computed following Autor et al. (2008).*

I estimate the structural model using the Korean Education and Employment Panel (KEEP) data. The panel data tracks Korean students in their secondary education in the initial survey until the mid- and late-twenties spanning the periods of higher education, transition to the labor market, and early job career. Two practical issues need to be addressed: measurement and short panel. To classify a worker to being overeducated, several measures have been proposed and commonly used in the literature. Hartog (2000) classifies the empirical designation of overeducation into three types: job analysis (JA),

[5]Considering the spending on supplementary private education during the secondary education years and college tuition fees would make the "internal" rate of return of college education look much less profitable, but the metric is not addressed in this paper.

realized matches (RM), and worker self-assessment (WA). Studies using the different measures generated broadly consistent evidence that a fair amount of overeducation appears to exist in labor markets. The KEEP is particularly appealing in that the WA overeducation measure is repeatedly observed, so I make use of the assessments by the same person over time. Given the workers' jobs must differ even within the same occupation code, this measure is advantageous to provide views about the "actual" job that the workers carry out[6] The other issue is attributed to short panel consisting of samples from nonstationary data generating processes where the time horizon of the agent extends beyond the length of the data. Because imposing arbitrary assumptions about future events after the end of the sample period which is the late twenties up to the end of the lifecycle is too strong, I adopt the conditional choice probability (CCP) estimation exploiting finite dependence property (Arcidiacono and Ellickson 2011). The estimated model can be employed to reconstruct some counterfactuals associated with temporary changes in utilities or choice sets within the sample period still without imposing strong assumptions beyond the sample period (Arcidiacono and Miller 2020; Kalouptsidi et al. 2021).

Empirical results in this paper confirm that overeducated workers, those who obtain college degrees but choose to work in the non-college sector, experience non-trivial wage penalties relative to college graduates employed in the college sector. The option of employment in the non-college sector for some college graduates who could otherwise remain unemployed brings about heterogeneity in returns to higher education and contributes to the observed wage dispersion. At the same time, a somewhat moderate role of the economic returns in one's indirect utility underscores the significant roles played by the consumption value and option value associated with academic college education in shaping students' educational choices. As a counterfactual policy, I consider restricting access to general high school for students with relatively low academic abilities and assigning them to vocational high school. This policy reduces the share of high school graduates

[6] Given the large amount of variation in job tasks that has been found to exist even within detailed occupation codes (e.g., see Autor and Handel (2013)), the other measures are at a disadvantage. Agopsowicz et al. (2020) is the first work to use a WA by the same person over time. Different measures of overeducation and their advantages/disadvantages are more discussed in Appendix. Ideally, to relieve the concerns stemming from the reliability of the WA measure, robustness could be checked with respect to an alternative measure of overeducation, realized matches (RM), which can be constructed based on the 4-digit occupational codes provided in the Korean data.

who enter academic college despite their higher probability of being overeducated after graduation. Instead, the marginal students redirected to vocational high schools pursue a vocational college degree and work in the college sector, decreasing the overall rate of overeducation while maintaining similar employment rates conditional on a college graduation. Notably, these marginal students earn higher log wages summed up until their late twenties that are the ages covered in the data, due to mainly early labor force participation even not being overeducation.[7]

Related literature This paper contributes to three strands of literature. First, it relates to the literature examining determinants, dynamics, and wage consequences of overeducation. A broadly consistent finding in this literature is that overeducated workers earn significantly lower wages compared to their counterparts who hold jobs aligned with their educational level (e.g., Baert et al. 2013), and such wage penalties persist over the long term (e.g., Clark et al. 2017). Interpretations of overeducation can differ based on theoretical perspectives. Some suggest that overeducation does not necessarily indicate an undesirable mismatch between workers and their jobs. For instance, overeducation might be rationalized by college graduates acquiring the qualification due to the preferences from schooling itself (called consumption value or psychic costs/benefits, Lazear 1977; Oosterbeek and Van Ophem 2000; Jacob et al. 2018) and then working in non-college sector because of the non-pecuniary amenities (Gottschalk and Hansen 2003) or considerations about occupational mobility (Sicherman and Galor 1990; Sicherman 1991). Allen and Van der Velden (2001) and Agopsowicz et al. (2020) emphasize sorting based on differing abilities among similarly educated workers in a well-functioning labor market. In contrast, Albrecht and Vroman (2002) and Gautier (2002) highlight search frictions as a cause of inefficient allocations leading to overeducation. Unlike most prior work, this paper explicitly models overeducation as a choice option, incorporating the option value that college graduation can provide by securing the college graduates' employment against negative shocks associated with finding and working in the college sector jobs.

Second, the paper contributes to the empirical literature analyzing heterogeneous returns to education. Understanding the economic return to education is central to policy discus-

[7] There remains a possibility that their lifetime earnings could be lower in the counterfactual scenario, as workers with four-year college degrees may exhibit different wage trajectories, including distinct patterns in wage growth over their mid- to late-career stages.

sions, and the large treatment effects literature has focused on the heterogeneity of these returns at the point of educational choice. For example, Carneiro et al. (2011) and Nybom (2017) estimate distributions of the treatment effects of college education in the U.S. and Sweden, respectively. The latter documents positive expected lifetime earnings gains even for individuals who did not attend college, challenging predictions of the classic Roy model that assumes self-selection based only on lifetime earnings. Further expanding the set of alternatives considered, Rodríguez et al. (2016) evaluate returns across different post-secondary education options in Chile, finding that about 31% of five-year college graduates could have increased their earnings by selecting different educational pathways. Subsequent employment choices following education also influence heterogeneity in returns to educational decisions, and Kinsler and Pavan (2015) demonstrate that college graduates working in jobs unrelated to their fields of study face around 30% lower wages. Building upon these insights, this paper explicitly incorporates overeducation as a dimension of labor market mismatch to explore heterogeneous returns jointly across educational and employment decisions.

Lastly, this paper adds to the literature on modeling educational decisions and conducting ex-ante evaluations of educational policies. Dynamic structural models have become important tools for understanding human capital accumulation by capturing the mechanisms underlying observed choices and dynamic selection, and for evaluating counterfactual policies regarding lifetime earnings and utility. Influential studies such as Keane and Wolpin (1997) and Arcidiacono (2005) simulate how college decisions respond to changes in the cost of college attendance. However, Altonji et al. (2012a) emphasizes the importance of jointly analyzing high school and college decisions. Navarini and Verhaest (2023) highlight the broad relevance of different types of education when analyzing overeducation and emphasize the role of vocational education in reducing overeducation and the resulting wage penalties. Aligned with the findings in this paper, they illustrate how targeted educational policy could mitigate mismatches and enhance labor market outcomes. I consider, similar to the counterfactual policy simulated in Declercq and Verboven (2018), an admission policy to assign general high school students to vocational high school based on their academic ability. This policy addresses the question within the Korean context, asking whether ”too many” students pursue academic college and whether alternative sorting of the students at the high school stage could yield socially preferable outcomes.

The rest of the paper is structured as follows. Section 2 presents the institutional context in Korea and the data, and section 3 describes the dynamic discrete choice model of schooling and work decisions. Section 4 discusses the estimation method and estimation results obtained from the model. In section 5, the simulation procedure and counterfactual policies are outlined to carry out the counterfactual analysis. The summary of this paper and its limitations at this moment is presented in section 6.

1.2 Background and Data Description

1.2.1 Institutional Background: The Education System in Korea

The Korean education system consists of six years of primary education in elementary schools, six years of secondary education that is split into three years of middle school and three years of high school, and higher education. Education up until the lower-secondary education in middle school (i.e., grade 9, usually age 15), is mandatory and comprehensive, with all students receiving the same standardized curriculum free of charge. Upper-secondary education in high school is more diversified, neither compulsory nor free, allowing students to select from different high school types.[8] High school types are divided broadly into “general high school” (prioritizing academic education), “vocational high school” (offering vocational training), and the “other types” (specialized education in specific areas like foreign languages, science, and arts or sports). Among the population of the high-school entrants in 2005, 69.2% entered general high schools, 28.3% entered vocational high schools, and 2.5% entered other types.[9]

[8]While high school education is not compulsory, it is nearly universal with an enrollment rate higher than 99%. The high school fees vary across the type of high school, but the amounts are relatively not very large; in 2010, 22% of budget for secondary education but 73% for higher education was borne by private households.

[9]Students attending other types of high schools, who make up only a small portion, are classified as general high school students from here on. Admission policies differ according to school type and location. In large cities, general high schools follow the high-school equalization policy that assigns students irrespective of academic performance, while general high schools outside equalization zones usually in smaller cities, vocational high schools, and specialized high schools select students based on their own criteria.

Students in general high school study a common core curriculum in grade 10. At the end of the academic year, the students choose between a natural science-focused track and a liberal arts-focused track. Similarly, students in vocational high school are taught a standard academic core curriculum in grade 10 before specializing in a vocational field offered differently depending on the characterization of the school, such as business, agriculture, engineering, technology, fishery. Regardless of the type of education that high school students complete, those who obtain a high school diploma can enter a college for higher education. Senior high school students can choose whether to take College Scholastic Ability Test (CSAT) administered every year before graduation season. Since the grades of the test are often a crucial criterion in the admission policies set by colleges, many students who intend to go to college take this high-stakes examination.[10] Depending on their desired college programs, students select among subjects largely divided into mathematics, Korean verbal, English, social sciences/natural sciences/vocational education, and second foreign languages. Students, in addition, choose one of the two types of mathematics exams, one of which is more comprehensive (advanced) in terms of material addressed than the other. The CSAT can be retaken after their high school graduation to improve their grades.

Alike to the U.S. higher education system, the standard structure of higher education in Korea includes associate degrees awarded by vocational colleges, bachelor's degrees awarded by academic colleges, and advanced degrees including master's degrees and doctoral degrees. Vocational colleges cover junior colleges, polytechnics, and industrial universities, and they offer two- or three-year vocationally oriented post-secondary programs. Many programs in fields like engineering, business administration, health care, secretarial studies, or agriculture are two years in length, but three-year programs also exist in fields like nursing, rehabilitation therapy, or early childhood education. Academic colleges are mostly multi-disciplinary institutions that comprise multiple departments (i.e., university), but there are also several mono-specialized academic colleges in the fields like engineering or education. Programs in standard academic disciplines are four years in length, while bachelor's programs in professional disciplines like architecture, pharmacy,

[10] Among the population of high school graduates in 2008, 71.8% took the CSAT in that year. There are also those who entered college without taking the exam, and some high school graduates did not enter college after taking the exam.

or medicine require five or six years of study. Military service, which is mandatory in South Korea for all men starting at age 18, is typically fulfilled in their early twenties in the middle of college years or right after high school or college graduation.[11]

1.2.2 Data: Korean Education and Employment Panel (KEEP)

The KEEP is a dataset provided by the Korea Research Institute for Vocational Education and Training, a government-funded research institute. Three groups of samples—senior middle-school students, senior general high-school students, and senior vocational high-school students—were randomly selected within each group with a sample size of 2,000 per group (i.e., stratified sampling). The students were first surveyed in 2004, and annual follow-up surveys were conducted until 2015. The middle-school cohort sample (i.e., third-year middle-school students in the initial survey) is largely nationally representative, considering that around 95% of Koreans attend middle school. The vocational high-school cohort is oversampled, as fewer than 30% of middle school graduates enter vocational high school. To address sample composition and attrition in the middle-school cohort, an additional 600 senior general high school students and 1,000 senior vocational high school students were added in the fourth survey wave. Panel structure and sample sizes are detailed in Appendix A.

The dataset includes variables related to educational and labor market choices, individual characteristics, parental background, and CSAT grades. Notably, since CSAT is observed only for those who took the test, there are individuals whose scores of the test are missing. To utilize the grades as a control for academic ability in the main analysis, missing values are predicted based on regressions by leveraging additional variables in the panel and treated equivalently to observed selection-free measurement.[12] The data contains detailed information about educational outcomes (e.g., GPA, major, dropout status) and labor outcomes (e.g.,occupational codes, wages, hours worked). Importantly, workers self-report how well their job matches their education level through a specific question: *"What*

[11]Duration of the military service is roughly two years, varying by military branch. Service can be deferred under specific circumstances, and exemptions are possible only for health or other valid reasons.

[12]The regressions show high explanatory power, but another way that relies more on the structural model would be to use EM-algorithm. Detailed specification for regression-based prediction and predictors are explained in Appendix B.

do you think about the level of education your current job requires compared to your education level?"[13] The response, capturing within-education and within-occupation variation more effectively, enables the construction of an overeducation indicator.

For the main analysis using a dynamic discrete choice model, I aggregate schooling and occupation choices at the yearly level. Observations are defined by academic years that run from March to February, and a few individuals making choices inconsistent with the model introduced in Section 3 are excluded.[14] The final sample features that the college graduates who answer that their current job requires lower or much lower level of education than they have are regarded as working in the non-college sector. However, the opposite that the less educated have access to the more educated's labor market which is segmented by education would not be plausible. Therefore, while present in the survey, high school graduates and college dropouts responding that their job requires high or much higher level of education are not considered working in college sector. Another feature is that working part-time while in school is not treated as an available choice option. Either schooling or working full-time as a main activity is considered as the choice made in the year. The constructed balanced panel with valid schooling choices and labor market outcomes contains 1159 and 1132 individuals in the middle school and high school cohort, respectively.

1.2.3 Descriptive Statistics

This section provides descriptive statistics that highlight the features of observed educational and occupational choices. First, Table 1.1 presents individual characteristics by the type of high school in senior year: vocational or general high schools. Overall, students in general high schools tend to have a stronger parental background in terms of their education and income, invest more in supplementary education, and are more likely to take the CSAT in which higher grades are achieved.

Figure 1.2 illustrates the distribution of subsequent choices followed by graduation from different high school types. College enrollment is high across the two groups, yet patterns

[13] The response falls into five categories: (a) much higher than my level, (b) higher than my level, (c) comparable to my level, (d) lower than my level, and (e) much lower than my level.

[14] Details of sample restriction and data construction procedures are described in Appendix C.

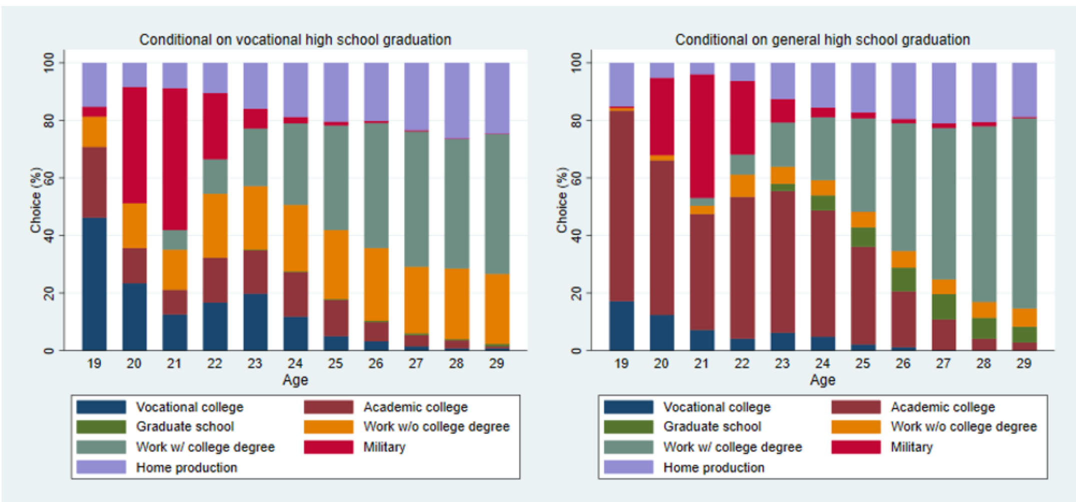
Table 1.1: Schooling Statistics (Secondary Education Choice)

	Vocational High School			General High School		
	Mean	S.D.	N	Mean	S.D.	N
Female (==1)	0.41	0.49	913	0.46	0.50	1370
Urban (==1)	0.39	0.49	913	0.46	0.50	1370
Parent's Years of Schooling	10.76	2.96	913	12.88	3.03	1370
College Educated Parent (==1)	0.10	0.30	913	0.36	0.48	1370
Household Income Quartile	1.98	1.07	913	2.68	1.10	1370
Private Tutoring (==1)	0.43	0.50	913	0.71	0.45	1370
Spending on Private Tutoring	8.76	14.96	913	28.08	34.79	1370
CSAT Korean	2.85	1.42	254	5.24	1.89	1212
CSAT English	2.71	1.17	251	5.22	1.93	1199
CSAT General Math	3.35	1.22	198	5.28	1.88	828
CSAT Advanced Math	2.00	1.02	26	5.04	1.86	314

Notes: Only the CSAT taken in the senior high school year is taken into account to compute the statistics. Grades are on a scale from 1 to 9. Spending on private tutoring is expressed in units of 10,000 won (2015 value).

differ: approximately 70% of vocational high school graduates enter college at age 19, with over two-thirds attending vocational colleges. In contrast, over 80% of general high school graduates continue to higher education, but only about one-quarter choose vocational colleges. Male students' mandatory military service causes non-steady college attendance patterns in their early twenties. The transition into the labor market primarily occurs in the mid-twenties.

Figure 1.2: Choice Distributions after High School Graduation



I further study the employment and wage outcomes of the college graduates. Table 1.2 displays the employment status, both in extensive and intensive margins, over the years after graduation from either vocational or academic colleges without further postgraduate education. In the early career stage, unemployment rate decreases but seems to slightly bounce back after several years. Conditional on being employed, the last column indicates a non-trivial degree of overeducation, with academic college graduates more likely to experience both overeducation and unemployment.

Table 1.2: Share across Employment Status after College Graduation

	Unemployed	**Employed**		
		A. matched	B. overeducated	B/(A+B)
Vocational College Graduates				
1 year after graduation	22.8	66.7	10.5	13.6
2 years after graduation	20.6	68.6	10.8	13.6
3 years after graduation	18.6	69.7	11.7	14.4
4 years after graduation	18.5	71.4	10.1	12.4
5 years after graduation	21.3	68.2	10.5	13.3
6 years after graduation	19.7	70.8	9.5	11.8
7 years after graduation	27.2	61.7	11.1	15.3
8 years after graduation	22.0	65.9	12.1	15.5
Academic College Graduates				
1 year after graduation	34.9	52.0	13.1	20.2
2 years after graduation	28.9	56.9	14.3	20.0
3 years after graduation	22.2	63.2	14.6	18.7
4 years after graduation	19.6	66.7	13.7	17.1
5 years after graduation	26.0	58.7	15.3	20.7
6 years after graduation	23.8	63.5	12.7	16.7

Notes: Overeducated employment refers to the share of workers who answer that their current job requires a lower level of education than they have.

To investigate the relationship between wages and overeducation, I estimate extended Mincerian wage equations, and the results are reported in Table 1.3. The variable of interest is overeducation dummy that indicates whether the college educated worker says his or her current job requires a lower education level. The results show overeducation is associated with a large and significant wage penalty, which amounts to 12%, and the magnitude is robust with respect to the specification. The insigificant interaction between overeducation dummy and years of work experience suggests that the penalty

Table 1.3: Estimation Results of Extended Mincer Wage Equations

Variables	(1)	(2)	(3)
Dependent variable: Log (monthly wage)			
Overeducation	-0.109***	-0.122***	-0.115***
	(0.015)	(0.023)	(0.023)
Overeducation × Work Experience		0.007	0.005
		(0.007)	(0.007)
Schooling, Experience, Year Dummies	Yes	Yes	Yes
Academic ability measure (CSAT grade)	No	No	Yes
Adjusted R-squared	0.187	0.196	0.217

persists over time. I find that, even though controlling for academic ability enhances the explanatory power by 10 percentage as measured by adjusted R-squared, the negative wage effect does not disappear.

1.3 The Baseline Model

1.3.1 Overview

I build a dynamic discrete choice model to analyze educational and employment decisions together with their labor market returns. Forward-looking individuals in each period make a yearly decision from a finite set of mutually exclusive alternatives. In the initial period, students choose their high school type. Having a high school diploma awarded, individuals make schooling or work choices from the second period on. They have schooling options including whether to attend a vocational college or an academic college. Academic college graduates can enroll in graduate school. As for work decisions, individuals who did not graduate from either vocational or academic college can work only in the non-college sector. In contrast, college graduates can choose employment between the college sector and the non-college sector. Males have military obligation, so they need to serve in the military for at least one period. Those not in education, employment, and military are classified as staying home.

Individuals may not choose the alternative that yields the highest economic return but, similar in spirit to Cameron and Heckman (2001), maximizes the utility that includes such factors as the consumption value of college education and frictional costs faced in the labor market. In the initial period, they choose the type of high school they graduate from, which affects the future through its switching costs on the subsequent choice. Vocational high school graduates may experience the utility costs of switching educational tracks if they choose an academic college, and vice versa. Attending a college for higher education may provide students with a direct utility (called consumption value or psychic cost/benefit) and a higher earning potential, at the cost of foregone earnings and an increase in labor market experience. I also suppose that working in the labor market yields non-pecuniary benefits, receiving more attention in the literature recently (e.g., Kaplan and Schulhofer-Wohl 2018), as well as pecuniary payoff. The additional gain associated with college graduation in the labor market segmented by education is that it opens up the option for employment in the college sector while still providing the option for working in non-college sector. In other words, if a college graduate has difficulty finding a job that requires a college degree or realizes that the available jobs in the college sector pay poorly due to economic factors, a job that has a lower education requirement instead would be an available option.

I model the discrete choice as an optimization solution by an economic agent facing a finite-horizon dynamic problem. Individuals at age t make a choice d_t to maximize expected lifetime discounted utility. There are two types of state variables: observed state variables s_t and idiosyncratic preference shocks ε_t observed by the individual but unobserved by the econometrician. δ is the yearly discount factor of which value is set to 0.95 . $U(d_t, s_t, \varepsilon_t(d_t))$ denotes the per-period flow utility from making the choice d_t, which depends on s_t and ideosyncratic shock $\varepsilon_t(d_t)$ realized before the decision. The state variables evolve according to a Markov transition process $Q(s_{t+1} \mid s_t, d_t, \varepsilon_t(d_t))$ which is known to the decision-makers. To make the individual maximization problem relatively simple like Rust (1987), I impose the standard assumptions: additively separable utility $U(d_t, s_t, \varepsilon_t(d_t)) = u(d_t, s_t) + \varepsilon_t(d_t)$, and conditional independence $Q(s_{t+1} \mid s_t, d_t, \varepsilon_t(d_t)) = Q(s_{t+1} \mid s_t, d_t)$. The value function, denoted by $V(s_t, \varepsilon_t)$, corresponds to the expected present discounted value of lifetime utility from following the optimal decision rule conditional on s_t and ε_t. Then, by Bellman's principle of optimality:

$$V(s_t, \varepsilon_t) = \max_{d_t} \{u(d_t, s_t) + \varepsilon_t(d_t) + \delta\, \mathbb{E}_t \left[V(s_{t+1}, \varepsilon_{t+1}) \mid s_t, d_t\right] \} \tag{1.1}$$

where the expectation is taken with respect to the uncertain components which are the future state variables. I further assume that $\{\varepsilon_t(d_t)\}$ are distributed independently and identically following Type I Extreme Value distribution. Every observed state variables, including schooling/work experience, previous choice, and age, evolve deterministically, except for college graduation, which is treated as probabilistic in the model.[15]

The assumption that the unobserved shocks ε_t are independent over time could be restrictive when there exist persistent state variables that are observed to the decision-maker but not to the econometrician. An example would be the ability and taste endowed to the individuals from the beginning. The persistent unobservables captured by ε_t lead to serial correlation, making the independence assumption invalid. Despite the well-known importance of controlling for the dynamic and self-selection through the unobserved heterogeneity component, I build a simpler model without the unobserved heterogeneity. Instead, for the moment, I use an observed measure of academic ability (i.e., CSAT grades) to control for dynamic selection through an ability channel, which can be generalized for extension.[16]

1.3.2 Flow utility

Secondary education

In the initial period of normalized age $t = 1$, individuals choose which type of high school they graduate from. The secondary education choice is treated as endogenous, rather than an initial condition, so it could vary under counterfactual scenarios. The choice set consists of two high school types: $J = \{v, g\}$ where v and g refer to vocational and general high school respectively. Note that I exclude the small number of high school dropouts from

[15]For notational simplicity, I denote $u(d_t, s_t)$ by $u_d(s_t)$. The state variable s_t can be divided into Z_{it} and X_{it}, which are not mutually exclusive.

[16]The standard approach in dynamic discrete models is to employ finite mixture distributions following Heckman and Singer (1984). I plan to follow Arcidiacono and Miller (2011) to implement the approach in estimating the dynamic discrete choice model while handling the unobserved heterogeneity.

the sample. Therefore, the initial choice set does not include the option corresponding to stopping education before graduating from high school education. The flow payoff of choosing a high school type, $d_{it} = j(\in J)$, is parameterized by

$$u_s(Z_{1it}) = \alpha_s + Z_{1it}\beta_s \,, \tag{1.2}$$

where Z_{1it} denote the characteristics influencing the flow utility of completing secondary education.[17] The list of the observed covariates includes the measure of (academic) ability, household income discretized into quartiles, and dummies indicating college-educated parent, gender and living in an urban area.[18] Importantly, I impose the restriction that, once an individual attends a college, the type of high school education does not play a role as a state variable anymore. In other words, secondary education does not have a permanent effect, for example, through the accumulated human capital, conditional on having college education.

Higher education

The flow utility is defined for each of the college choices from the choice set $H = \{a, b\}$ for those who have graduated from a high school and $H = \{gs\}$ for those who have graduated from an academic college. Indices a, b, and gs respectively refer to vocational college (associate), academic college (bachelor), and graduate school. Let Z_{2it} denote the variables that affect the utility of higher education, and then the utility for choosing a schooling option $d_{it} = h(\in H)$ is given by

$$u_h(Z_{2it}) = \alpha_h + Z_{2it}\beta_h \,. \tag{1.3}$$

[17] Magnac and Thesmar (2002) prove the exact degree of underidentification of dynamic discrete choice models under a standard setup. They show that choice-specific utilities are not nonparametrically identified without setting the distribution function of unobserved preference shocks, the discount rate, and the utilities of one reference choice. For identification, in the initial period, the vocational high school choice ($d_{i1} = v$) is taken as the reference choice of which utility is normalized to 0.

[18] Admittedly, there are several relevant issues to using CSAT grades as a measurement of innate ability. For instance, students are evaluated after they have chosen and been exposed to the different types of education. Also, whether to take the test and what subjects to take are self-selected. To simplify the specification, I specify the grades as exogenously given from the beginning, rather than modeling everything (e.g., effort choice to prepare the test, unobserved ability of which distribution can be inferred from a factor model) to be treated within the model.

Z_{2it} include the indicator corresponding to the high school education in the previous period to captures switching costs. Household income, discretized into quartiles, also appears in the list of covariates because the variable might be directly related to credit constraints making payment for college less attractive. The flow utility u_h is interpreted as the consumption value of college attendance representing the value attached to schooling and learning.

Employment

I assume that there are two sectors in the labor market: college sector indexed by c and non-college sector indexed by nc. The choice set of the individuals choosing a work option depends on their college graduation status. Before college graduation, they are supposed to work only in the non-college sector, so the choice set is a singleton $L = \{nc\}$. College graduates, either from vocational or academic colleges, choose their employment between the two labor sectors from the choice set $L = \{nc, c\}$. The workers who have studied, at least, two years in graduate school are assumed to work only in the college sector. The flow utility for working in a labor sector $d_{it} = l \in L$ is given by

$$u_l(Z_{3it}) = \alpha_l + Z_{3it}\beta_l + \phi \mathbb{E}_t[w_{lit}] , \tag{1.4}$$

where w_{lit} are log wages in each sector.[19] At the beginning of period t when individuals make a choice, expectations are formed about the wages they will receive at the end of the period. The coefficient ϕ transforms the economic return into the utility, so it is natural to restrict the parameter to be the same across the labor sectors. On top of the (expected) pecuniary payoff, there exist nonpecuniary benefits which depend on Z_{3it}.

[19] It is not arbitrary to assume that the flow utility of work is logarithmic in earnings. See, for instance, Belzil and Hansen (2002) and Pavan (2011) who estimate models of human capital accumulation where the utility of work is assumed to be logarithmic in earnings.

Military service

Military service is mandatory for Korean males who do not have an exceptional reason.[20] Until being discharged from the military, serving in the military ($d_{it} = m$) is in their choice set after the initial period (i.e., $t > 1$). In reality, there are complex rules which determine the upper limit of age before when the military obligation has to be fulfilled, but it is infeasible to reflect all the individual-specific conditions. It is usually the early twenties when most men start to serve in the military, and college students mostly complete military service before their third year in college. In most cases, two consecutive years are devoted to the military, but there are some variations that amount to one year to nine years of military service observed in the data within the sample period. The flow utility of serving the military is specified as

$$u_m(Z_{4it}) = \alpha_m + Z_{4it}\beta_m \,. \tag{1.5}$$

Z_{4it} include age, squared age, and the years in the military to reproduce the observed choice pattern. The flow utility of military service would vary over age, so males have a dynamic incentive to join the military at some point, especially at an early age. Once the military choice is made, he can choose a schooling or work option from the next period on. The (dis)utility of serving additional periods in the military is captured by the effect of the years in the military, with the set of dummies indicating one year, two years, and more than two years of military experience. Even though the flow utility of forced military service does not depend directly on the other observable characteristics, opportunity costs (i.e., the values of the choices other than military service) are different across individuals through the characteristics and unobserved preference shocks. Lastly, it is assumed that the stock of human capital which had been accumulated remains constant without human capital depreciation during the military service.

Home production

I set the choice of not being in education, employment, or military at $t > 1$ to be the reference choice $d_{it} = 0$ which is often referred to as home production. The corresponding

[20] It is assumed that those who are exempted from the military service perfectly predicted their exemption from the initial period.

flow utility is normalized to zero. Well-established nonparametric identification studies in the dynamic discrete choice literature (e.g., Magnac and Thesmar 2002) demonstrate the underidentification result that the number of free flow utilities is the same as the dimension of the finite state space. Hence, some restrictions need to be imposed, and it is usual to normalize the flow utility associated with one of the alternatives in every state. Then, the flow utility parameters should be interpreted relative to the normalized alternative.

1.3.3 Belief

Since individuals account for the future impact of their decisions, their beliefs about the future objects that are uncertain at the time of making a decision need to be specified. The transition of the time-varying state variables other than college graduation follows deterministic processes: years of experience corresponding to different schooling and work options, age, calendar year, previous choice. Beliefs about the potential wages as well as college graduation, which are not realized when the current choice is made, are formed as described in the following. They are assumed to be objectively correct with respect to future realizations (i.e., rational expectation).

College graduation

College graduation is treated as probabilistic as did in Declercq and Verboven (2018) and Arcidiacono et al. (2025). The probabilistic college graduation embraces several components missed in the model. For instance, vocational college study could be two- or three-year associate degree program, and academic college study could be four-, five-, or six-year bachelor’s degree program. Also, college students could endogenously choose to attend additional years than required for graduation. In the model, college students are at risk of graduation when the minimum years of schooling of the corresponding college type are completed. Academic college students are at risk of graduation if they had attended at least three years and are currently attending the fourth year. Similarly, vocational college students are at graduation risk after completing one year in the college and when attending the second year.[21] A probability of graduation after their τ-th period of college enrollment depends on a set of characteristics $X_{ih\tau}$ including time-variant or -invariant components. It is worth mentioning that $X_{ih\tau}$ is not t-specific, so the graduation probability is.[22] The specification implies that $Z_{ih\tau}$ does not include age or year dummy variables but involves years in college. The probability of college graduation conditional on $Z_{2i\tau}$ is further assumed to take a logit form and parameterized by a vector ψ:

$$\Pr(gradu^h_{i\tau+1} = 1 | X_{ih\tau}) = \frac{\exp(X_{ih\tau}\psi)}{1 + \exp(X_{ih\tau}\psi)}\,, \tag{1.6}$$

for $h \in \{a \text{ (academic college)},\ b \text{ (vocational college)}\}$. The state variable $gradu^h_{i\tau+1}$ indicates graduation from college type h in the next period.

Expected wages

Workers either in the college sector or non-college sector receive a log wage w_{ilt}. At the beginning of period t, expectations about wages which will be paid at the end of the period and in the future are formed based on the true relationship between state variables and wages:

$$w_{lit} = \tau_{lt} + X_{ilt}\gamma_l + e_{ilt}\,, \tag{1.7}$$

Log wages in sector l at time t linearly depend on the sector-specific time fixed effect δ_{lt} and a set of the individual characteristics X_{ilt}. Idiosyncratic wage shocks, e_{ilt}, follow normal distribution $N(0, \sigma_l^2)$. They are independent across individuals, sectors, and time, and independent of the other state variables including preference shocks ε_{ilt}. Log wages depend on individuals' characteristics and the highest obtained diploma in X_{ilt}, but not on the time spent in college. The nonlinear sheepskin effect implies that attending college without obtaining a degree does not lead to higher wages relative to high school graduates. I model the wage returns to the variables in X_{ilt} to be sector-specific, which incorporates the idea that accumulated human capital is appreciated differently in the different labor

[21]While available in practice, the model does not incorporate college graduation when the schooling choice (i.e. taking a course and earning credits) is not made.

[22]This assumption enables the finite dependence property to hold and be leveraged for estimation, which is better explained in the next section. It is worth mentioning at this stage that, once the graduation probability depends on t through policy changes over time or effects of aging, the property turns out to be invalid.

sectors. As for the calendar year dummies τ_{lt} that account for sector-specific aggregate labor market shocks and business cycle effects, individuals observe the current values before making a choice and form beliefs over the variables for periods $t+1$ and beyond. I assume that they exactly know the future values (i.e., perfect foresight).

Lastly, to identify the coefficient of the expected log wage in the utility function, it is necessary to have, at least, one variable included in the wage equation but excluded from the non-pecuniary benefit from working: $X_{ilt} \setminus Z_{3it}$. The variation in wages through the excluded variables helps to identify the utility attributed to an additional pecuniary payoff, providing the money metric measure of utility ϕ. In this paper, I assume the business cycle, work experience, and the dummy indicating fulfilled military service to satisfy the exclusion restriction.

1.4 Estimation

1.4.1 Sequential approach

Under the assumption that the idiosyncratic shocks are mutually and serially independent, the model can be estimated sequentially. The validity of sequential maximum likelihood (MLE) estimation rests on the log-likelihood being separable in the contributions of the choices, probabilistic graduation, and log wages. The sequential approach proceeds in two stages. In the first stage, one can consistently estimate the vector of parameters in the college graduation probability θ_h and in the wage equation θ_w. The second stage is to estimate the vector of flow utility parameters θ_u via a CCP estimation method, taking the first stage estimates as given.[23]

[23]While the two-step estimator provides a relatively simple way to obtain consistent parameter estimates, the asymptotic variance of the two-step estimator needs to be adjusted for valid statistical inference by incorporating the statical errors in the first step. In this paper, I provide only the standard errors that have not been adjusted yet.

First stage: college graduation and the wage parameters

In the absence of unobserved heterogeneity using a finite mixture approach, it is straightforward that each vector of college graduation and the wage parameters can be consistently estimated separately from all of the other parameters since every idiosyncratic component is assumed independent. Therefore, data on college graduation and wages is pooled across individuals over time. I estimate the parameters governing the graduation probabilities via maximum likelihood and the wage parameters via ordinary least squares which is equivalent to MLE. Given that individuals' beliefs are correctly specified, the expected graduation probabilities and wages for each individual are constructed based on the first stage estimates and are used in the next stage.

Second stage: utility parameters

With the first stage estimates taken as given, estimation of the flow payoffs uses the conditional choice probability (CCP) estimation method developed by Hotz and Miller (1993) and further refined by Arcidiacono and Miller (2011) to account for unobserved heterogeneity. There are several advantages of this estimation strategy. First, when finite dependence holds, the dynamic discrete choice models with no analytical solution do not have to be numerically solved, which minimizes the curse of dimensionality and reduces the computational burden. Another important advantage is that arbitrary assumptions regarding beliefs about future events beyond the length of the data do not need to be made because everything about the future is captured by the CCPs. To explain the CCP method, I dropped the subscript i for simplification of the notation, and define the conditional value function, $v_t(Z_t, d_t)$ as the present discounted value net of current preference shock ϵ_{td_t} when choosing d_t and behaving optimally from period $t+1$ on:

$$v_t(s_t, d_t) = u(s_t, d_t) + \beta V_{t+1}(s_{t+1}) dQ(s_{t+1}|s_t, d_t), \tag{1.8}$$

where $V_t(s_t) \equiv E_{\epsilon_t}[V_t(s_t, \epsilon_t)]g(\epsilon_t)d\epsilon_t$ is called the ex-ante value function. Then, the individual's maximization problem each time t can be written as

$$\max_{d_t}\{v_t(s_t, d_t) + \varepsilon_t\}. \tag{1.9}$$

Provided by the i.i.d. type I extreme value distribution assumption, the problem yields the closed-form choice probability for choice d_t and *Emax* expression of $V(s_t) = \mathbb{E}[\max_{d_t}\{v_t(s_t, d_t) + \epsilon_t\}]$ (McFadden 1973):

$$p_t(d_t|s_t) = \frac{\exp(v_t(s_t, d_t))}{\sum\limits_{d'_t}\exp(v_t(s_t, d'_t))}, \tag{1.10}$$

$$V_t(s_t) = \ln\left(\sum_{d_t}\exp(v_t(s_t, d_t))\right) + \gamma, \tag{1.11}$$

where γ is Euler's constant. Combining the two equations lets $V_t(s_t)$ be rewritten with respect to the conditional value function associated with an arbitrarily selected choice, d_t^*:

$$V_t(s_t) = -\ln(p_t(d_t^*|s_t)) + v_t(s_t, d_t^*) + \gamma. \tag{1.12}$$

Using the expression obtained, $v_t(s_t, d_t)$ is re-expressed:

$$\begin{aligned} v_t(s_t, d_t) = \; & u(s_t, d_t) \\ & + \beta\int\{-\ln(p_{t+1}(d_{t+1}^*|s_{t+1})) + v_{t+1}(s_{t+1}, d_{t+1}^*)\}dQ(s_{t+1}|s_t, d_t) + \beta\gamma, \end{aligned} \tag{1.13}$$

where d_{t+1}^* is an arbitrary choice in period $t+1$. Note that the data consist of agents' actions at different states, which implies the CCPs, $p_t(d_{t+1}^*|s_t)$, and the state transition $Q(s_{t+1}|s_t, d_t)$ can be recovered directly from the data. It is still necessary to characterize $v_{t+1}(s_{t+1}, d_{t+1}^*)$ to form the likelihood contribution of the observed choices in the data. The next subsection is devoted to explaining the technicality to address the remaining conditional value function.

1.4.2 Finite dependence

Two-period dependence There is finite dependence when two choice sequences with different initial decisions lead to the same distribution of states after a few periods. Consider a choice $d_t = j$ and the reference choice $d_t = 0$. By construction of the model described in Section 3, there exist two sequences of choices such that the probability distributions of the state variables after two periods are the same after following the choice sequences:

$$\{d_t = j, d_{t+1} = 0, d_{t+2} = 0\}, \{d_t = 0, d_{t+1} = j, d_{t+2} = 0\}.$$

In other words, the two sequences of choices lead individuals with the same state variables in period t to the same age, having an equal number of years of schooling, work experience and the same previous choice afterward. Then, the following property holds:

$$\begin{aligned} &E_t[V_{t+3}(s_{t+3}) \mid s_t, d_t = j, d_{t+1} = 0, d_{t+2} = 0] \\ &\qquad = E_t[V_{t+3}(Z_{t+3}) \mid Z_t, d_t = 0, d_{t+1} = j, d_{t+2} = 0]. \end{aligned} \tag{1.14}$$

This finite dependence property helps to formulate the choice probabilities in terms of utility functions and reduced-form objects: conditional choice probabilities and state transition probabilities. Similar to the standard logit or probit models, estimating dynamic discrete choice models rests on differences in conditional value functions, not the conditional value functions themselves:

$$p_t(d_t = j|s_t) = \frac{\exp(v_t(s_t, d_t = j) - v_t(s_t, d_t = 0))}{\sum_{j'} \exp(v_t(s_t, d_t = j') - v_t(s_t, d_t = 0))}. \tag{1.15}$$

I rewrite the expression for $v(s_t, d_t = j')$ and $v(s_t, d_t = 0)$ such that the future utility term is expressed relative to the choices in the finite dependence sequence, respectively. Then, the finite dependence property simplifies the differenced conditional value function $v(s_t, d_t = j) - v(s_t, d_t = 0)$ because the three-period ahead value terms cancel out when differences are taken:

$$
\begin{aligned}
& v_t(s_t, d_t = j') - v_t(s_t, d_t = 0) \\
&= u(s_t, d_t = j') - \beta u(s_{t+1}, d_{t+1} = j') \\
&\quad + \beta \int \ln(p_{t+1}(d_{t+1} = j'|s_{t+1})) dQ(s_{t+1}|s_t, d_t = 0) \\
&\quad - \beta \int \ln(p_{t+1}(d_{t+1} = 0|s_{t+1})) dQ(s_{t+1}|s_t, d_t = j') \\
&\quad + \beta^2 \iint \ln(p_{t+2}(d_{t+2} = 0|s_{t+2})) dQ(s_{t+2}|s_{t+1}, d_{t+1} = j') dQ(s_{t+1}|s_t, d_t = 0) \\
&\quad - \beta^2 \iint \ln(p_{t+2}(d_{t+2} = 0|s_{t+2})) dQ(s_{t+2}|s_{t+1}, d_{t+1} = 0) dQ(s_{t+1}|s_t, d_t = j').
\end{aligned}
$$

Note that the differenced conditional value function which is relevant to estimation consists only of flow utility and the adjustment terms which can be precalculated. As demonstrated in the denominator, the finite dependence must hold for each possible choice at time t. There are two cases where the standard two-period dependence does not hold because the choice set is restricted in some states. The finite dependence paths corresponding to the exceptions are described below.

One-period dependence Two exceptional states require attention to construct finite dependence paths because the two-period dependence described above is not applied to them. First, consider the secondary education choice which can be made only at $t = 1$ but not at $t \geq 1$. It is not available, by construction, to establish the path $\{d_1 = 0, d_2 = j, d_3 = 0\}$ for any secondary education alternative $j \in \{v, g\}$. However, equipped with the assumption that the highest education level matters once a student enter college, the sequences of two choices below satisfy the finite dependence:

***Case 1*:** $\{d_1 = v, d_2 = j'\}, \{d_1 = g, d_2 = j'\}$ for higher education alternative $j' = b$. That is, despite the switching costs incurred depending on the type of high school, academic college enrollment leads to the same state at $t = 3$. I set j' to the choice of academic college (i.e., $j' = a$), so the difference in the value functions yields

$$\begin{aligned}
&v_1(s_1, d_1 = v) - v_1(s_1, d_1 = g) \\
&= u(s_1, d_1 = v) + \beta u(s_2|d_1 = v) + \beta \int \{u(s_2, d_2 = a) + \ln p_2(d_2 = a|s_2)\} dQ(s_2|s_1, d_1 = g) \\
&\quad - \beta \int \{u(s_2, d_2 = a) + \ln p_2(d_2 = a|s_2)\} dQ(s_2|s_1, d_1 = v),
\end{aligned}$$

where $u(s_1, d_1 = g)$ is normalized to 0. Two-period ahead value terms are canceled out because of the finite dependence property:

$$E_1[V_3(s_3)|s_1, d_1 = v, d_2 = a] = E_1[V_3(s_3)|Z_1, d_1 = g, d_2 = a].$$

Similarly, once males are discharged from military duty (i.e., $d_{t-1} = m, d_t \neq m$), the military choice cannot be made from the next period on. This implies that the choice sequence $\{d_t = 0, d_{t+1} = m, d_{t+2} = 0\}$ is not feasible given $d_{t-1} = m$. Because the effect of completing military service is modeled through a dummy variable in the wage equation regardless of the periods spent in the military, two choice sequences along which finite dependence hold can be established:

Case 2: $\{d_t = m, d_{t+1} = j'\}, \{d_t = 0, d_{t+1} = j'\}$ for $j' = 0$ given $d_{t-1} = m$. That is, extended compulsory military service provides the flow utility but does not the future value through changes in state variable. Then, the differenced value function is

$$\begin{aligned}
&v_t(s_t, d_t = m) - v_t(s_t, d_t = 0) \\
&= u(s_t, d_t = m) + \beta \int \ln p_{t+1}(d_{t+1} = 0|s_{t+1}) dQ(s_{t+1}|s_t, d_t = 0) \\
&\quad - \beta \int \ln p_{t+1}(d_{t+1} = 0|s_{t+1}) dQ(s_{t+1}|s_t, d_t = m).
\end{aligned}$$

1.4.3 Estimation in practice

By exploiting the finite dependence property, estimation of the flow utility parameters reduces to a static conditional logit formula with a precalculated adjustment term. Estimation of the flow utility parameters involves the following steps:

1. Estimate the CCPs $p_t(d_t|s_t)$ via a flexible conditional logit model.

2. Compute the differenced future value terms along the finite dependence paths using the predicted values of CCPs.

3. Estimate the flow utility parameters which maximize the choice probabilities expressed in terms of the utility functions and the adjustment terms.

Note that the CCPs estimated in a reduced-form way in the first step are included in the expression of the second step. In practice, with enough data points, CCPs could be estimated using nonparametric bin estimators. However, as the number of possible state combinations is large, CCPs are estimated using a flexible logit, which is a usual approach in empirical studies using the CCP method (e.g., De Groote and Verboven 2019; Arcidiacono et al. 2025).

1.5 Empirical Results

1.5.1 Parameter Estimates

Estimates of the wage equations are presented in Table 1.4. In the non-college sector, vocational college graduation (associate degree) is not significantly better rewarded relative to high school graduation, while the return to academic college graduation (bachelor's degree) is around 14%. In the college sector which is accessible only to the college graduates, academic college graduation yields 5% higher wage than vocational college graduation. However, focusing solely on within-sector comparisons may mask the heterogeneity in returns to college education induced by occupational choice. For example, imagine that a vocational college graduate with all other explanatory variables equal to zero chooses a sector in the labor market. The difference in the expected wages between the two labor sectors is captured by the difference in the constants. Similarly, academic college graduates can expect around 37% higher wages by working in the college sector if all covariates are zero.

Returns to college-sector and non-college-sector work experience are of a similar magnitude in the non-college sector. However, experience in the non-college sector does not translate into higher wages in the college sector. This highlights the importance of accounting for sector-specific labor market experience, as experience is rewarded differently across sectors. I find significant returns to a one-grade improvement in the ability measure, CSAT score, only in the college sector. Additionally, military experience appears to be rewarded in both sectors, and females receive substantially lower average wages in the college sector.

Table 1.4: Estimates of Log Wage Equations

VARIABLES	Non-college Sector	College Sector
no degree	0.008 (0.027)	
academic college degree	0.142*** (0.030)	0.046*** (0.012)
graduate school experience		0.022 (0.014)
non-college work experience	0.087*** (0.010)	-0.000 (0.007)
(non-college work experience)^2	-0.005*** (0.001)	
college work experience	0.084*** (0.021)	0.067*** (0.007)
(college work experience)^2	-0.005 (0.004)	-0.004*** (0.001)
ability measure	-0.005 (0.005)	0.031*** (0.003)
college educated parent	0.058** (0.023)	0.039*** (0.013)
female	0.010 (0.037)	-0.094** (0.039)
military completed	0.209*** (0.037)	0.104*** (0.040)
constant	4.712*** (0.051)	5.091*** (0.043)
Time fixed effect	YES	YES
Adjusted R-squared	0.184	0.223
Number of observations	3133	4116

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Table 1.5 reports the structural parameter estimates associated with the flow utility of schooling. The estimated coefficients on previous general high school (compared to vocational high school) point to the existence of relatively large switching costs from vocational high school to academic college. The parameter on vocational college indicates that, as expected, the vocational college option is more attractive to vocational high school graduates than to general high school graduates. The results suggest that sorting students into high schools differently would lead to different college attendance. It can be found from the coefficient on the ability that higher ability individuals prefer to attend general high school and academic college. The particular importance of the ability in the utility of the academic tracks suggests a lower cost of effort at learning when the measure is high. The effect of extra years of schooling implies that the longer students attend college the lower cost of effort spent into schooling.

Table 1.5: Flow Utility of Schooling Alternatives

VARIABLES	General High School	Vocational College	Academic College	Graduate School
ability measure	1.432*** (0.066)	-0.161 (0.247)	0.470*** (0.163)	0.202 (0.504)
female	-0.163 (0.139)	6.233*** (0.823)	3.795*** (0.705)	-1.562 (1.728)
college educated parent	0.135 (0.184)	0.155 (1.028)	-0.251 (0.652)	-0.898 (1.849)
parental income quartile 1	-0.636*** (0.214)	-2.018* (1.161)	-6.206*** (1.108)	-0.384 (2.808)
parental income quartile 2	-0.054 (0.219)	-0.155 (1.272)	-3.230*** (1.154)	-1.139 (2.564)
parental income quartile 3	0.194 (0.215)	0.736 (1.337)	-2.137* (1.151)	1.292 (2.222)
urban region	-0.595*** (0.145)			
years in vocational college		3.163*** (0.597)		
vocational college graduation		-11.267*** (2.556)	-0.342 (1.414)	
years in academic college			1.581*** (0.311)	
academic college graduation			-44.409 (18489.230)	-14.450*** (5.106)
years in graduate school				-2.814 (2.706)
previous general high school		-0.402*** (0.125)	0.450*** (0.109)	
previous vocational college		0.876*** (0.103)	-1.395*** (0.230)	
previous academic college		-1.906*** (0.314)	1.370*** (0.078)	0.001 (0.192)
previous graduate school				5.267*** (0.647)
constant	3.791*** (0.255)	-0.926 (1.194)	-2.779*** (0.806)	-1.108 (2.502)

Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Observations = 95,917.

The estimated flow utility parameters of work and military options are presented in Table 1.6. The parameter on expected log wage indicates that individuals do care about pecuniary payoffs. Graduates from either vocational or academic colleges are less attracted to employment in the non-college sector than high school graduates. Ability seems to have a direct effect on the utility gains from working, in addition to the indirect effect through expected wages. Lower ability makes employment in the non-college sector more attractive to college graduates, suggesting that overeducation reflects sorting among individuals with similar education levels but different abilities.

The coefficients on previous activities indicate substantial switching costs associated with changing sectors. Interestingly, for working in the college sector, previous employment in the non-college sector entails a lower switching cost than staying at home. The large and significant coefficients emphasize the importance of the first labor market entry after college graduation. Heterogeneous entry costs into the college sector explain why some college graduates accept overeducation by working in the non-college sector if they cannot afford high entry costs. No evidence is found that household income levels affect entry

Table 1.6: Flow Utility of Employment Alternatives and Military

VARIABLES	Non-college Sector	College Sector	Military
ability measure	-0.351* (0.187)	-0.248 (0.210)	
female	1.469** (0.688)	2.307*** (0.774)	
college educated parent	-0.128 (0.830)	-0.126 (0.862)	
vocational college graduation	-3.470*** (1.062)		
academic college graduation	-1.852* (1.065)	1.218 (0.994)	
extra years in vocational college	4.934*** (0.976)	5.158*** (0.756)	
extra years in academic college	3.565*** (0.989)	2.792*** (0.742)	
years in graduate school		4.287*** (1.417)	
expected log wage	4.888*** (1.460)	4.888*** (1.460)	
previous general high school	-2.657*** (0.307)		
previous non-college sector	1.121*** (0.068)	0.557*** (0.117)	
previous college sector	1.013*** (0.117)	1.543*** (0.091)	
previous military			-3.355*** (0.395)
age			0.292* (0.170)
age^2			-0.035** (0.016)
constant	-40.252*** (8.782)	-35.000*** (9.117)	-1.108 (2.502)

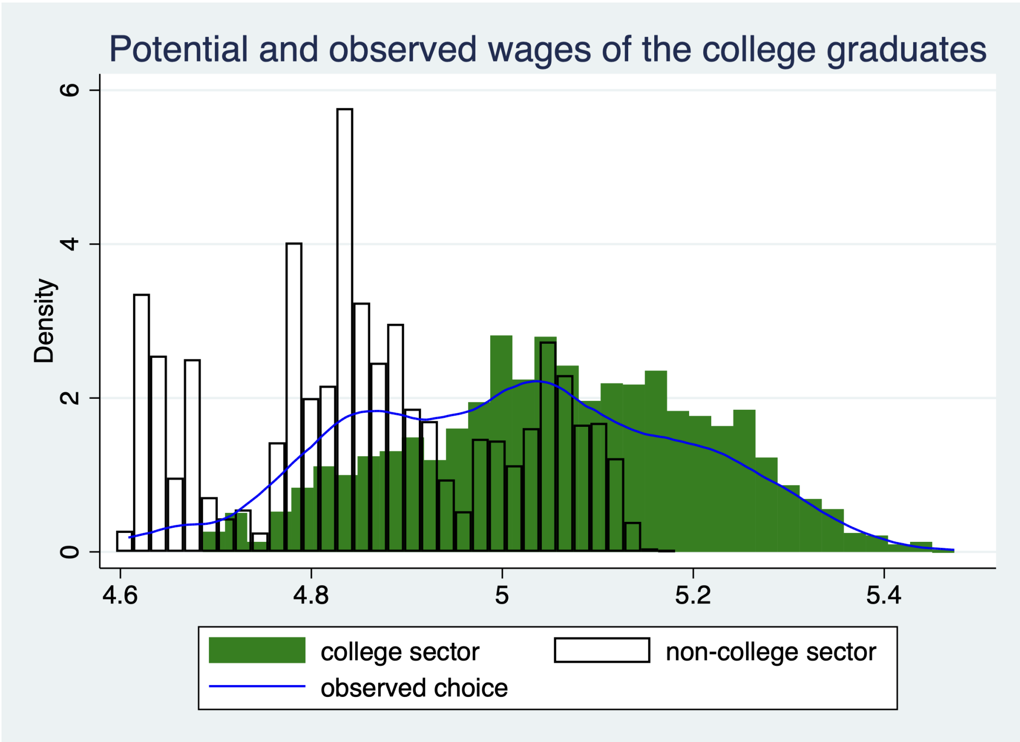
Notes: Robust standard errors in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$. Observations = 95,917.

costs, rejecting the hypothesis that less wealthy graduates are disadvantaged in this regard. Lastly, military utility increasing with age captures the realistic aspect that older males are unlikely to avoid military service.

1.5.2 Heterogeneity and Wage Dispersion

To see the contribution of the overeducation option on the potential wage distribution, I simulate the wage distribution of the college graduates and illustrate it in Figure 1.3. More specifically, I simulate the wages of the workers who choose to work in the presence of only the college sector (green histogram), which is the wage outcomes in the absence of access for college graduates to the non-college sector. To compare the wage determination, I also simulate the same workers' wages conditional on working in the non-college sector (uncolored histogram). They demonstrate different wage determinations in the segmented labor market. Lastly, I simulate the wages of the workers who choose to work either in the college or non-college sectors (blue contour). Here, the main takeaway is that, since some college graduates would not work if they don't have access to the non-college sector, the overeducation option secures their employment while the realized wages are more dispersed in the segmented labor market.

Figure 1.3: Wage distribution



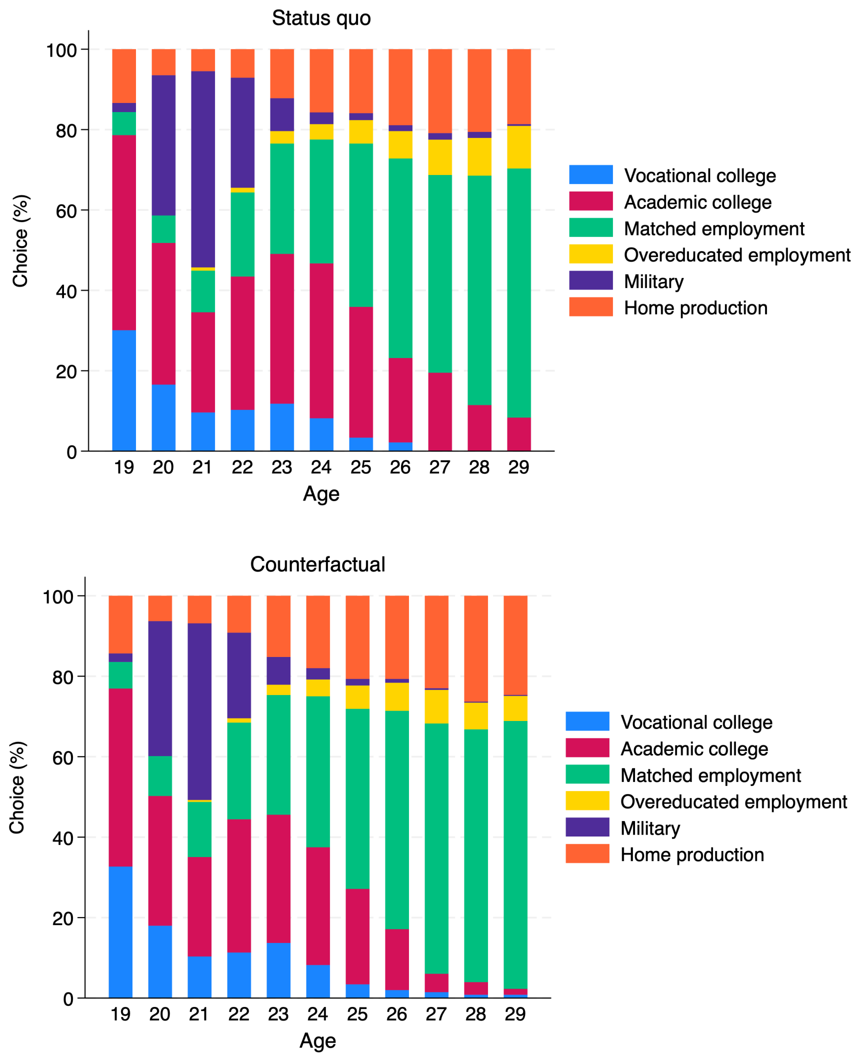
1.5.3 Counterfactual Policies

In the structural model with the segmented labor markets, academic college graduates with lower academic ability have a lower preference to work in college-sector jobs, leading them to decide to work in the non-college sector. Aimed at addressing the overeducation rate in society, I consider a counterfactual policy to sort students based on their academic ability at the high school level. Particularly, general high school students with relatively lower academic ability, corresponding to the bottom 15% of CSAT grade within the general high school, are assigned to vocational high school.[24] The status quo and counterfactual choices under this policy are described in Figure 3.6.

[24]This counterfactual policy can be studied without solving the dynamic programming problem. As described in the previous section, the finite dependence assumption imposes the restriction that the policy only alters the choice in the first period along with its impact on the college entrance in the next period. The resulting value functions are unchanged after the second period, then future value terms are still available to be recovered from the original CCPs from that period on.

This counterfactual choice distribution in the lower panel of the figure shows that the share of high school students who advance to vocational college in the next year of high school graduation increases by around 3%. This policy moderately reduces the share of academic college students, but the share of those who attend academic college until their late twenties decreases significantly. In the labor market, as the overall share of academic college graduates who have low ability decreases, the overall rate of overeducation diminishes. The higher rate of home production is driven by those who do not have a college degree, which occupies a larger share under the counterfactual policy.

Figure 1.4: Status Quo and Counterfactual Choice Distributions after High School Graduation



1.6 Conclusion

This paper presents a dynamic discrete choice framework to investigate the educational decisions and heterogeneous returns to higher education in the segmented labor market, using Korean panel data. By incorporating the concept of overeducation as a labor market option, the model highlights how college graduates may self-select into non-college sector employment when faced with limited payoffs in the college sector. The findings underscore that while overeducation secures employment, it contributes to the dispersion in wages among college-educated individuals. Crucially, the structural estimates reveal that the utility derived from higher education extends beyond pecuniary returns, encompassing significant consumption and option values that influence educational choices. Sector-specific experience and switching costs also play a critical role in shaping labor market outcomes, suggesting that mismatches could reflect complex trade-offs in individual optimization.

The counterfactual policy simulation—reallocating lower-ability students from general to vocational high schools—demonstrates the potential of early sorting to reduce overeducation and enhance early-career wage outcomes without sacrificing employment rates. However, the analysis is bounded by the use of a short panel, limiting insight into long-term effects, including mid- and late-career wage trajectories. While the model excludes unobserved heterogeneity and some market frictions, future work could extend the framework to incorporate individual unobserved heterogeneity and richer labor demand dynamics. Nevertheless, the study provides empirically grounded insights into how educational policy and labor market structure jointly shape the economic returns to education, with implications for countries undergoing rapid educational expansion.

Chapter 2.

Information about ability and educational investment mistakes: Evidence from Korea

Olivier De Groote, Gyung-Mo Kim, François Poinas

Abstract: We study educational investment decisions when students are uncertain about their own ability and preferences. We estimate a dynamic discrete choice model in which we allow agents to be uncertain about their own unobserved type. We estimate the degree of uncertainty by exploiting commonly available data on agents' self-reported most likely outcome. We apply this model to the Korean context and find large uncertainty before important high school track and effort decisions are made, particularly among the (overconfident) low-ability students. Providing more information would lead to stronger sorting as it disincentivizes low-ability students to exert effort and choose academically focused tracks, while it incentivizes high-ability students more. We also find effects in the long run, leading them to more different types of college degrees.

2.1 Introduction

Schooling investment and human capital acquisition are prime factors for subsequent higher income and other important positive life outcomes (e.g., Krueger and Lindahl 2001; Heckman et al. 2018). Becker's canonical model of human capital and Willis and Rosen's (1979) model, which applies the Roy model (1951) to the economics of education, view education as an investment, where costs are compared to the discounted stream of expected future benefits. Since it is the ex-ante returns perceived by students that will influence actual schooling decisions, understanding their beliefs and perceptions would serve as the necessary foundation for understanding how educational decisions are made.

This paper studies the role of beliefs in educational decisions. We consider that individuals could be uncertain about their skill endowment or ability, which in turn leads to uncertainty about preferences and outcomes. Individuals, especially in their early life-cycle, are often found to be over-confident about their ability (Moore and Healy 2008; Bobba and Frisancho 2022). This misperception can lead to mistakes in choices such as the high school track, hours of study effort, or college enrollment. For example, an over-confident, low-ability student might want to obtain a college degree but would invest too little in study effort in high school to achieve this. However, they could also be exerting too much effort if the required effort to achieve this is too costly for a low-ability student.

To quantify the degree of uncertainty, we propose a dynamic discrete choice model equipped with underlying uncertainty as to one's own true endowment. This is a state of nature that characterizes ability, preferences, and match quality. Individuals form beliefs about the state, which is flexibly estimated in the model. It nests the usual assumption in the literature that individuals are sorted into different choices based on the state that is fully known to the decision-makers but unobserved to the econometrician (see e.g. Keane and Wolpin 1997).[1]

[1] The standard assumption in the literature is that individuals have perfect information about the deterministic features in the decision-making process and rational expectations over the stochastic features. In contrast, in this paper, we relax the assumption of perfect information about their own state, one of the key deterministic features. Even we allow that the beliefs about the state can be objectively incorrect with respect to its true distribution, violating the rational expectation assumption.

However, since dynamic discrete choice models are under-identified (Magnac and Thesmar 2002), it is often hard to relax the assumptions regarding individuals' information and beliefs given typically available data consisting of choices and the corresponding outcomes. In order to drop the particular assumption that one's latent state is perfectly known from the beginning, we bring in additional data reporting the choice that is expected to be made with the highest probability in the future.[2] First, we specify a behavioral model that approximates sequential choices and outcomes, incorporating a discrete and time-invariant latent state representing an agent's unobserved heterogeneity. We show that, by leveraging prediction data, the beliefs about own latent state can be identified. The fundamental intuition for identification is as follows: if agents with different latent states exhibit distinct optimal behaviors, then discrepancies between the predicted educational outcomes eliciting their beliefs and the predictions based on their true latent state reveal the information they have about their latent state. We formalize this identification strategy by applying the straightforward statistical specification, which results in a system of linear equations that allows us to recover the probabilistic beliefs.

We estimate the identified model using Korean panel data on high school students and their career advancement. South Korea presents a relevant context, characterized by high educational attainment relative to other OECD nations, alongside significant challenges coupled with intense academic pressure. Notably, long study hours and substantial household expenditure on supplementary private education, aimed at securing admission to selective colleges, are widespread concerns. [3]

We obtain the following three main findings. First, we find a clear distinction between a high and low unobserved academic ability type, with the latter being almost twice as likely to choose a vocational track in high school. Second, agents have little information about

[2] We, in this paper, call it "prediction" data. To disentangle the influence of preferences versus beliefs on individual behavior, observed choice and outcome data is supplemented from time to time with additional data eliciting one's beliefs. However, the specific types of belief data utilized and the methodologies employed for their integration vary considerably across this body of research.

[3] OECD data from 2012 indicated that 72% of middle school and 53% of high school students participated in such private education, with the associated financial burden taking up 11.6% and 13.4% of average household income per student, respectively. Studying long hours to go to a better college has become so excessive that Korean authorities even imposed curfews, usually at 10 p.m., on students having private tutoring. In 2012, the Programme for International Student Assessment (PISA) reported that, among 65 surveyed countries, Korea had the lowest proportion of 15-year-old students reporting happiness at school. The heavy burden and pressure associated with academic pursuits alongside competition is considered a primary factor contributing to this statistic.

this type. The low type agents believe there is a 60% chance they are a low type. As the population average is 41%, it shows they have more information than the econometrician, but it also rejects the common assumption they know it with certainty. The high type is also imperfectly informed, believing they are a high type with a 70% probability. Third, a counterfactual simulation that provides all agents with perfect information about their type reveals stronger sorting patterns. It increases the academic level of the track and study effort of high-ability students, and decreases it for low-ability students. In the long run, it leads to more high-ability students having a college degree from a selective, academic institution, while low-ability students graduate more often from a vocational college.

Related literature This paper contributes to a body of research that examines the drivers of economic decisions and the importance of belief. Economic decisions with dynamic considerations can be characterized in three dimensions: preference, time, and belief. In the literature on dynamic decisions, it is well known that dynamic selection bias attributed to omitted individual effects that capture different initial skill endowment, ability, and/or match quality across individuals needs to be controlled (Cameron and Heckman 1998). One of the standard ways to account for the unobserved heterogeneity in the models of dynamic choice is to apply the random effects approach along with the assumption that economic agents are perfectly aware of their own latent state endowed from the beginning but not observed by the econometrician.[4] On top of the unobserved heterogeneity influencing preference, many evidences empirically support that beliefs about uncertain future returns are subjective and play a significant role.[5] We study subjective belief within the dynamic discrete choice model framework such as Cameron and Heckman (2001); Arcidiacono (2005), and Heckman et al. (2018).

This paper addresses the crucial, but usually not tested, assumption of perfect information about unobserved heterogeneity (e.g., Heckman and Singer 1984; Cunha et al. 2005), under which unobserved heterogeneity is found to account for a substantial portion

[4]The two most popular ways to facilitate the unobserved heterogeneity are to apply the random effects approach applying the finite mixture model as in Heckman and Singer (1984) or to proxy the unobservables using multiple measurements as in Carneiro et al. (2003).

[5]It has been found that subjective beliefs matter to drive individuals' behavior is provided in different contexts such as voting (DellaVigna and Kaplan 2007; Chiang and Knight 2011), sexual behavior (Delavande and Kohler 2016), and financial decisions (Hudomiet et al. 2011).

of the variations. That is, endowed with the latent variable known to the agents, they make decisions and have outcomes based a lot on the value of the variable. Among others, Keane and Wolpin (1997), who studies schooling, work, and occupational choice, find that around 90% of the total variance in lifetime utility is explained by unobserved heterogeneity. However, in the presence of uncertainty about the latent state, the explained variations would be partly attributable to the subjective beliefs over it. If this is the case, the considerable explanatory power of the unobserved heterogeneity boils down to an additional and different policy implication that indicates the importance of information provision rather than the attention just to the initial endowment. We develop a dynamic discrete choice model that treats perfect information as a special case, while the decision-makers are assumed to generally form beliefs over their own latent variable. By identifying and estimating the beliefs, we conduct the counterfactual analysis that alters beliefs so that the role of the belief can be examined.

Despite the evidence supporting the significance of subjective beliefs, the economic analysis of dynamic decisions with the beliefs identified at the same time is difficult. When it comes to dynamic discrete choice models, Rust (1994) and Magnac and Thesmar (2002) discuss and provide non- and under-identification results. Abbring (2010) broadly reviews the extent of identification in dynamic discrete choice processes, of which much of the analysis shares the rational beliefs and perfect information assumptions imposed by researchers who have access only to choice, outcome, and covariate panel. Supplementary data has been exploited to disentangle and infer the beliefs. Popular data is the one eliciting the full distribution representing one's belief through well-posed survey questions in a probabilistic way. The advantage of using such subjective belief data is stressed by Manski (2004), and some authors paid attention to this sort, for example, Stinebrickner and Stinebrickner (2014) and Kapor et al. (2020). However, directly measuring the full distributions of beliefs is often not an available option to researchers because such rich and reliable data is hard to obtain in practice. Hence, other authors leverage alternative and accessible data, such as the perceived mean outcomes (Zafar 2011; Hoffman and Burks 2020) and stated hypothetical choices (e.g., Mas and Pallais 2017; Wiswall and Zafar 2018).[6] Koşar and O'Dea (2023) provide an extensive review of this strand of literature.

[6] d'Haultfoeuille et al. (2021) develop a test for rational expectations, which can be used when a researcher has data on perceived and actual outcomes.

We contribute to the expanding literature on the use of elicited beliefs. In this paper, we incorporate the predicted outcome that is perceived to be most likely realized at the time of belief elicitation. This allows us to relax the standard assumptions of perfect information and rational expectations in dynamic discrete choice models. While earlier studies have leveraged such prediction data primarily to infer preferences, they have not directly used it to characterize beliefs. For instance, Manski (1990) demonstrates how it binds the probability to behave in a certain way. Van der Klaauw (2012) integrates expected future occupation data with actual choice data to estimate a dynamic model of teacher career decisions. Adding the moment associated with the predicted outcome enhances the precision of the parameter estimates, providing better identification of the unobserved heterogeneity under the assumption of perfect information. We leverage similar predicted outcome data which is relatively readily available in surveys to account for probabilistic beliefs formed possibly under imperfect information.[7] Our key methodological innovation lies in a new choice-based framework for inferring potentially biased beliefs, which combines a tractable statistical specification with prediction data alongside standard choice and outcome information. This approach enables us to disentangle variations in predicted outcomes across economic agents into components due to beliefs stemming from uncertainty about their own latent state and heterogeneity in the latent state that is known from the beginning to them at the time the prediction is made.

Lastly, we contribute to the literature that links subjective beliefs and educational decisions. Experimental evidence indicates that students' beliefs are formed on the basis of imperfect information and biased (e.g., Wiswall and Zafar 2015; Bobba and Frisancho 2022). The research examining the role of belief by exploiting survey that elicits expectations and embedding them in structural models is as follows. Arcidiacono et al. (2012) take into account the subjective assessments of one's own abilities as well as expectations of future earnings associated with alternative college major choices. Attanasio and Kaufmann (2014) additionally consider the subjective distribution of future earnings to account for perceived risk in the labor market. More recently, Delavande and Zafar (2019) further

[7] Exemplary surveys that contain expectations data are National Longitudinal Survey of Youth-1979 Cohort (NLSY79), National Education longitudinal Study, Youth in Transition Survey (YITS). Unlike many other large-scale surveys, the NLSY97 additionally asks respondents to report the actual percent probability corresponding to future events; for example, "What is the percent chance that you have a four-year college degree by the time you turn 30?"

investigate the beliefs about dropout and non-pecuniary aspects in the university choice, finding a relatively important role of these beliefs compared to the small role of labor market prospects.

Several studies, such as James (2011); Stange (2012); Sanders (2014), and Arcidiacono et al. (2025), highlight the significance of uncertainty about one's latent state for educational attainment and labor market outcomes. However, they still assume that students' initial beliefs are rational, ruling out the possibility of systematic misperception. Gong (2019) is an exception further dropping the rational expectations assumption, but only in settings where the students' decisions are repeatedly observed with their rules being time-invariant. In our paper, each of the sequential and different educational decisions is characterized using a flexible discrete choice model, with the initial beliefs identified rather than assumed.[8] We apply the identification strategy to a Korean data set to empirically estimate the beliefs and investigate educational investment in the country. Counterfactual educational choices under perfect information suggest the implications for Korean students of uncertainty about their own comparative advantage and beliefs formed in their early life cycle for heavy investment in education.

The remainder of the paper is organized as follows. In Section 2, we present the institutional context in Korea and the dataset containing individuals' predictions that we rely on to relax the conventional assumption in dynamic discrete choice models. In Section 3, we introduce the general set-up and the main econometric specifications. Section 4 illustrates the identification strategy and estimation method. Section 5 shows the application of our model and identification strategy to the Korean dataset. We describe the estimation results and investigate the impact of a counterfactual policy to provide information from the beginning of high school. Finally, Section 6 concludes.

[8] Methodologically, the model in this paper is in between the reduced-form and the fully structural dynamic discrete choice approach in the following sense. As in the structural literature, we try to estimate causal effects at clearly identified margins of choice for counterfactual policy analysis, but we approximate agent decision rules by remaining agnostic about the exact distinction between the costs and gains assessed for each alternative. In the absence of repeated data to infer the change in beliefs over time, we still need to model the belief process. Bayesian learning is one example.

2.2 Institutional background and data

2.2.1 Korean education system

Korean tracking starts in secondary education, which has both rigid and flexible features. On the one hand, at the end of mandatory and comprehensive lower-secondary education in middle school (i.e., grade 9), students choose a type of high school to attend: "general high school", "vocational high school", and "other types".[9] Different tracks are offered in separate high schools, and track-switching occurs infrequently. After the first-year of studies in a common core curriculum, high school students specialize in a more detailed curriculum. Academic high school students choose between a natural science-focused track and a liberal arts-focused track, and similarly, vocational high school students choose a vocational field offered differently depending on the characterization of the school, such as business, agriculture, engineering, technology, and fishery. On the other hand, regardless of the type and track of the education that high school students complete, those who have obtained a high school diploma can enter a college for higher education. Only a few very selective college majors (e.g., medical) require applicants to have studied certain subjects during the high school years.[10]

Before graduating from high school, third-year high school students can opt for taking the annual College Scholastic Ability Test (CSAT). Its score is a decisive criterion in many colleges' admission processes, so the majority take the test.[11] Again, alike to the U.S. higher education system, the standard structure of postsecondary education in Korea includes associate degrees awarded by vocational colleges and bachelor's degrees awarded by academic colleges. Vocational colleges offer two- or three-year vocationally oriented programs, while academic colleges provide academic disciplines that are from four to six years in length depending on the major. The bachelor's programs with five or six years of

[9]"Other types" includes high schools offering specialized education in areas like foreign languages, science, and arts/sports.

[10]College admissions in Korea are, similarly to the U.S., decentralized in that each college sets its own admission policy and makes binding and/or exploding admission offers to applicants, while the admitted students accept or reject the offers within a short time.

[11]Some students can apply and get admission through a selection process whose admission standard does not require a CSAT score. In general, those who choose to take the test do their best on the exam. High school graduates can retake the test to improve their CSAT scores.

the curriculum include professional disciplines like architecture, pharmacy, or medicine. Postgraduate education consists of master's degrees and doctoral degrees.

2.2.2 Korean Education and Employment Panel (KEEP)

The Korean Education and Employment Panel is provided by the Korea Research Institute for Vocational Education and Training, which is a government-funded research institute. In 2004, the initial year of the survey, three groups of students-senior middle school students, senior general high school students, and senior vocational high school students—were randomly selected within each group with a 2000 respective sample size (i.e., stratified sampling). With survey conducted in an in-person interview in principle, they are annually followed up until 2015. The middle school cohort sample (i.e., third-year middle-school students in the initial survey) are more or less nationally representative considering that most Korean attend and complete middle school. The vocational high school cohort sample is oversampled in that, conditional on high school enrollment, less than 30% of middle school graduates go to vocational high school in the population. To supplement the small number of students from the original middle-school cohort who transitioned into various tracks in high school, the survey expanded from the fourth wave onward to include an additional 600 senior general high-school students and 1,000 senior vocational high-school students.

The KEEP includes variables of schooling and work choices, their outcomes, individual characteristics, parental background, and the CSAT grades of the students who took the exam. Detailed rules to restrict the samples and construct the data set are explained in Appendix. Importantly, we pay attention to a survey item administered to parents of senior high-school students in the initial survey that elicits their prediction of the highest qualification their child will eventually obtain. The exact wording is: "What is the level of education you predict your child will be able to attain?" Parents select one of five mutually exclusive categories: high-school diploma, vocational college, academic college, master's degree, or doctorate. Unlike the probabilistic measures often considered desirable to capture subjective beliefs about an uncertain object (see, e.g., Stinebrickner and Stinebrickner 2014; Delavande and Zafar 2019), this item (a) records only the single alternative believed most likely, rather than a full probability distribution over all altanatives, and (b) is conditional on the child's current circumstances, not evaluating on a set of hypothetical states. We treat the parent and child as a single decision-maker, sharing identical information and preferences, thus abstracting from potential complications arising from asymmetric information or divergent valuations of education between parents and children.[12]

Table 2.1 provides the descriptive statistics of the variables in the panel across the different high school tracks. *effort* measures weekly average hours for self-study in the senior year of high school, and it is discretized into three levels. CSAT grades are provided after normalization, which divides the entire distribution of student scores into nine distinct ranks.[13] *prediction* indicates the elicited prediction of the final education level in the survey.

Table 2.1: Summary statistics: full sample

highest education	**High school**			**Vocational college**			**Academic college**		
	count	mean	sd	count	mean	sd	count	mean	sd
female	134	0.201	0.403	350	0.594	0.492	1447	0.464	0.499
household income quartile	134	2.269	1.091	350	2.169	1.050	1447	2.483	1.104
college-educated parent	134	0.209	0.408	350	0.194	0.396	1447	0.375	0.484
urban region	134	0.493	0.502	350	0.480	0.500	1447	0.482	0.500
high school track (1: science)	134	0.291	0.456	350	0.283	0.451	1447	0.411	0.492
effort (1-3)	134	1.470	0.733	350	1.603	0.829	1447	1.805	0.898
prediction (1-5)	134	3.015	0.736	350	2.871	0.658	1447	3.368	0.704
CSAT_Korean (1-9)	134	3.858	1.717	350	4.166	1.516	1447	5.360	1.657
CSAT_English (1-9)	134	3.836	1.664	350	3.869	1.442	1447	5.363	1.696

* *High school* encompasses individuals with some college experience, and *Academic college* includes those who have attained master's or doctoral degrees.

2.2.3 Descriptive statistics

Table 2.2 and Table 2.3 illustrate the distribution of CSAT grades in relation to students' initial college enrollment choices immediately following high school graduation (Table 2.2) and their final educational attainment within the sample period (Table 2.3). Table 2.2 reveals that, at the point of making initial college enrollment decisions, students who

[12] Some research highlights a substantial role of parental expectations and influence in children's educational choices (e.g., Attanasio and Kaufmann 2009; Foley et al. 2014).

[13] South Korea implemented a nine-grade system for reporting CSAT scores, which was designed to offer a relative evaluation of each student's performance in comparison to the entire cohort of test-takers. For example, the top-performing 4% of students in each subject received the highest grade, signifying their position of academic achievement on the national examination.

enrolled in vocational colleges had the lowest average CSAT grades. Interestingly, students who did not immediately enter any college after high school exhibited higher average CSAT grades than even those who entered academic colleges, potentially indicating that these students chose to defer enrollment to pursue higher qualifications. Examining the final educational achievement in Table 2.3, the average academic abilities of graduates from academic colleges show a slight increase compared to initial enrollment, while those who ultimately did not achieve a college degree show a significant decrease in average CSAT scores. The stronger correlation between academic ability measures (CSAT grades) and the final education outcome suggests that selection based on academic ability plays a relatively major role in determining the highest educational level attained.

Table 2.2: CSAT grades by initial enrollment choices

Initial enrollment	N	Mean	SD	Min	Max
high school	410	4.96	1.873	1	8
vocational college	537	1.452	1.460	1	8
non-selective academic college	944	4.34	1.583	1	9
selective academic college	480	6.21	1.727	1	9

Table 2.3: CSAT grades by final education outcomes

Final outcome	N	Mean	SD	Min	Max
high school	220	3.322	1.607	1	8
vocational college	521	3.418	1.460	1	8
non-selective academic college	1016	4.511	1.574	1	9
selective academic college	614	6.218	1.657	1	9

** Using college names in the data, we categorized academic colleges into two groups: selective and non-selective ones. The former contains colleges in Seoul and some public colleges outside Seoul. The public colleges are the flagship national universities and special-purpose colleges.*

Table 2.4 details the transitions observed between students' initial college enrollment choices and their final educational outcomes, presenting the frequency of individuals in each transition category. The CSAT English grade attained in high school is also presented in parentheses for each group, providing additional context to the transition patterns. The most frequent transition observed is from initially completing high school to ultimately achieving an academic college degree. Students who successfully made this transition show notably higher CSAT grades compared to their peers who remained at the high school graduation level. Conversely, students who transitioned downwards from either vocational or academic college enrollment to a final outcome of high school graduation exhibited lower average effort and CSAT grades, demonstrating that the transition would be also related to ability sorting.

Table 2.4: Transition from initial enrollment to final outcome

Initial enrollment	Final education outcome			
	high school	vocational college	Non-selective academic college	Selective academic college
high school	69 (3.26)	64 (3.64)	134 (5.00)	143 (6.32)
vocational college	58 (2.70)	417 (3.37)	59 (3.38)	3 (4.33)
Non-selective academic college	76 (3.40)	36 (3.55)	806 (4.44)	26 (4.81)
Selective academic college	17 (5.29)	4 (3.75)	17 (6.06)	442 (6.28)

Table 2.5 presents the frequencies of parental predictions regarding their child's eventual highest educational qualification against the child's realized final education outcome. This table specifically focuses on the high school cohort for whom parental predictions were collected. The data highlights both the instances where parental predictions align with actual outcomes and where discrepancies exist. A notable observation is that an academic college degree or higher is the predominantly predicted level of education by parents. However, the table also reveals that the predicted educational level is frequently higher than the qualification the child ultimately achieves, even though these predictions were made during the child's final year of high school when some information about their academic standing would have been available. Further analysis using regressions con-

firms a significant relationship between parental predictions and subsequent educational choice patterns, suggesting that factors beyond measured academic performance influence students' decisions in high school. This analysis also indicates that, when accounting for students' intentions and CSAT performance, the level of effort exerted is not found to be directly related to achieving a higher level of schooling.

Table 2.5: Frequencies across parental prediction and final education outcome

Parental prediction	Final education outcome			
	high school	vocational college	Non-selective academic college	Selective academic college
high school	6	5	5	3
vocational college	32	118	58	4
≥academic college	103	202	590	383
Total	141	325	653	390

** Survey question: "What is the level of education you predict your child will be able to attain?" Notice that we observe parental prediction only from the high school cohort.*

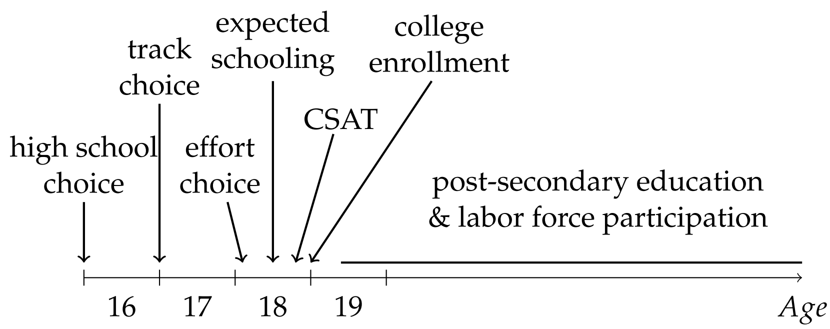
2.3 Model setup

We model each of the sequential and different educational decisions using a flexible discrete choice model. It features that the economic agents have uncertainty about their own latent state which represents individual heterogeneity. We drop the assumption of the standard random effects approach that they are perfectly aware of their own latent state endowed from the beginning, and instead, model the probabilistic beliefs over the state. We outline the sequence of choices and outcomes, and integrate the framework with the prediction made under the information friction about the latent state.

2.3.1 Overview

Figure 2.1 summarizes the timeline of the model. A middle school student enters either a vocational or general high school, and when the latter is chosen, she chooses her own high school track between science-focused and liberal arts-focused tracks at the end of the first year. She exerts effort with one of three levels, ranging from low to high, to prepare for the College Scholastic Ability Test (CSAT). It is after the effort choice but before the CSAT and

Figure 2.1: Timeline of the sequence



the realization of its results that the prediction of her future outcome of final education level is elicited. She takes the CSAT, and then, given her CSAT results and high school diploma, a post-secondary education choice is made among the alternatives: vocational college, non-selective academic college, and selective academic college.[14] We additionally distinguish the educational attainment from the initial enrollment choice.[15] Equipped with her own productivity enhanced during the endogenously determined education investment, she enters the labor market and yields wage returns.

We model agent i's optimization problems using a sequential choice model where the value function corresponding to the choice j from the available set J_s in stage s is denoted by $V_{sj}(Z_{is}, \theta_i, \epsilon_{isj})$. The value function depends on the state variables Z_{is}, the agent's latent state θ_i, and an idiosyncratic shock ϵ_{isj}. The latent state variable, drawn from a finite set of possible states of nature $(\theta_1, ..., \theta_K)$, represents permanent unobserved heterogeneity such as innate ability, preferences, and/or match quality specific to certain choices. θ_i, called his/her "type" hereafter. The importance of controlling for the unobservable has been highlighted in the literature to address the selection for causal inferences and counterfactual analysis (see, for example, Belzil and Hansen 2002; Beffy et al. 2012).

In contrast to this literature, we assume that agents possess imperfect information about their own true type. The heterogeneous belief of type k agent at the beginning of stage s is characterized by a probability distribution over the set of possible K types: $p_s^k =$

[14] Notice that our sample is restricted to the individuals completing education in high school and taking the CSAT. To model their higher education choices, we further classify the academic college in terms of college selectivity. Postgraduate education choice is not explicitly modeled.

[15] The idea behind the separation between enrollment right after high school graduation and final achievement is to consider a possible information revelation before reaching the final education outcome.

$(p_s^k(1), ..., p_s^k(K)) \in [0,1]^K$, with $p_s^k(r)$ the probability of belonging to type r in period and $\sum_{r=1}^K p_s^k(r) = 1$ in all stages s for type k. Additionally, it is assumed that the number of types, K, and how the unobserved heterogeneity types determine the choice-specific value functions and outcomes is common knowledge. Then, at stage s when the agent makes a choice after observing the choice-specific shocks ε_{ijs} realized, the agent maximizes the expected value function:

$$\max_{j \in J_s} \sum_{r=1}^{K} p_s^k(r) \; V_{sj}(Z_{is}, \theta_r, \varepsilon_{isj}) \, . \tag{2.1}$$

p_s^k is interpreted as the beliefs at the start of stage s conditional on the information up until the beginning of stage s. That is, p_s^k allows the prior to be stage-specific, embracing any belief process, for instance, updates in a Bayesian fashion. Note that the specification nests the perfect information assumption under which, if $\theta_i = \theta_k$, $p_s^k(k) = 1$ and $p_s^k(r) = 0$ for all s, $r \neq k$. When it comes to the other outcomes, CSAT grades and wages, which play a role of measurements and do not involve beliefs, they are modelled with an ordered logit and a model known as the Mincerian specification, respectively.

2.3.2 Econometric model

For computational simplicity, we approximate the choice-specific value function conditional on being unobserved heterogeneity type k as a linear function of the observable and unobservable components (e.g., Ashworth et al. 2021): $V_{sj}(Z_{is}, \theta_k, \varepsilon_{isj}) = v_{sj}(Z_{is}) + \alpha_{sj}^k + \varepsilon_{isj}$, where the unobservable to the econometrician are additively separable in the type-specific effect and idiosyncratic shock. Note that, instead of specifying functional form assumptions on flow utilities and solving a dynamic programming problem, approximating the choice-specific value function V_{sj} still enables economically relevant counterfactuals to be studied. In particular, it allows us to simulate choices under perfect information about own types.

The functional form, which is admittedly parsimonious, yields the following objective function of the maximization problem:

$$\sum_r p_s^k(r)\, V_{sj}(Z_{is}, \theta_r, \varepsilon_{isj}) = v_{sj}(Z_{is}) + \widetilde{\alpha}_{sj}^k + \varepsilon_{isj}\,, \tag{2.2}$$

where

$$\widetilde{\alpha}_{sj}^k = p_s^k(1) \cdot \alpha_{sj}^1 + \ldots + p_s^k(K) \cdot \alpha_{sj}^K\,. \tag{2.3}$$

The representation demonstrates our reinterpretation of the parameters resulting from uncertainty and the corresponding beliefs. For estimation, we further assume that $v_{sj}(Z_{is})$ is also linear in Z_{is} which will be divided into the initial and time-invariant individual characteristics denoted by X_i and the endogenous output in the previous stages. We put subscript s to the intercepts $\widetilde{\alpha}$ and coefficients β of X_i, which are parameters associated with the corresponding stage, and the effects of the previous choice or outcome are captured by δ.

Stage 1: high school track choice An individual chooses a combination of the type and track in high school, indexed by h, from the set of three options: $H = \{vocational(C),\ science(S),\ liberal\,arts($ The choice maximizes the expected value of the schooling option h that does not only includes the consumption value of studying the track but also captures its future values for nonpecuniary and pecuniary aspects:

$$d_{i1} = \arg\max_h\ X_i\beta_{1h} + \widetilde{\alpha}_{k1h} + \varepsilon_{i1h}\,.$$

Stage 2: effort choice Once an individual has decided her high school track $d_{i1} = h$, she invests effort in improving her preparation for the CSAT. The effort denotes average hours for self-study, being discretized into three levels e ($\in \{low\ (<10),\ middle\ (>=10, <20),\ high\ (>=30)\}$). The choice of the level of academic effort k is described as

$$d_{i2} = \arg\max_e\ \delta_{he} + X_i\beta_{2e} + \widetilde{\alpha}_{2e}^k + \varepsilon_{i2e}\,.$$

The additive shift of track-specific intercept δ_{he} implies that the values capturing both the current cost or the future returns associated with different effort level could be different across the high school tracks that students are attending.

Stage 3: CSAT Equipped with endogenously determined effort e in the track h, an individual takes and receives the discrete grade in a subject m of the test. The discrete measurement is generated by a latent random variable

$$\widetilde{Y}_{im} = \delta_{hm} + \delta_{em} + X_i\beta_{3m} + \alpha^k_{3m} + \varepsilon_{i3m} \, .$$

The discrete measurement Y_{im} takes the r ($\in \{1, 2, ..., 9\}$) iff $\eta_{m,r} < \widetilde{Y}_{im} <= \eta_{m,r+1}$, where $\{\eta_{m,1}, ..., \eta_{m,10}\}$ are thresholds with $\eta_{m,1} = 0$ and $\eta_{m,10} = \infty$ for each m.[16] It is notable that we treat the measurement as selection-free, meaning that Y_{im} is observed for every individual. The inclusion of δ_{hm} captures the effect of the type of schooling at the time of the test.[17] Effort also has direct effects on the test grades, and the observables X_i may also do.

Stage 4: Postsecondary education choice After receiving the CSAT grades and graduating from high school, the individual chooses one of alternatives for postsecondary education ℓ. The set of possible alternatives is composed of vocational college (V), non-selective academic college degree (NA), selective academic college degree (SA) and outside option (N). The outside option stands for either stopping completely continuing a study or delaying an entrance to a college. The high school graduate decides

$$d_{i4} = \arg\max_{\ell} \, \delta_{h\ell} + \delta_{y\ell} + X_i\beta_{4\ell} + \widetilde{\alpha}^k_{4\ell} + \varepsilon_{i4\ell} \, .$$

We assume that one's high school track and CSAT grade may directly influence the choice probabilities of the different highest education level, but effort has an impact only through

[16] Setting $\eta_{m,1} = 0$ is a normalization to pin down the overall level of the system. Normalizing one of the threshold, while estimating the intercept, provides the interpretable baseline mean that could potentially shift across the type.

[17] For example, students who take science-focused tracks in high school could perform better on the math subject due to having taken more and advanced math classes. More flexible specification that the coefficients depend on one's high school track choice(i.e., β_{3mh}) is available, but we restrict our specification.

the realized CSAT grade. In practice, we estimate the model by imposing a linear effect $\delta_{y\ell} = \delta_3 \times y_i$.

Stage 5: final education outcome Postsecondary education outcome of obtaining a degree ℓ' is classified into the same set of alternatives in stage 4: vocational college degree (V), non-selective academic college degree(NA), selective academic college degree(SA), and outside option (N) which denotes below graduation from a college. Comparing each (expected) value associated with ℓ' leads to the choice

$$d_{i5} = \arg\max_{\ell'} \delta_{h\ell'} + \delta_{y\ell'} + \delta^n_{\ell\ell'} I\{d_{i4} = N\} + \delta^d_{\ell\ell'} I\{\ell' < d_{i4}\} + \delta^u_{\ell\ell'} I\{\ell' > d_{i4}\} \\ + X_i \beta_{5\ell'} + \widetilde{\alpha}^k_{5\ell'} + \varepsilon_{i5\ell'} \, ,$$

where $\delta^n_{\ell\ell'}$ captures the cost of delayed enrollment. $\delta^d_{\ell\ell'}$ and $\delta^u_{\ell\ell'}$ capture the additional cost of switching downward and upward from a different type of college initially enrolled in after high school graduation. As specified in the value functions associated with the college enrollment choice in stage 4, $\delta_{h\ell'}$ and $\delta_{y\ell'}$ are the effects of one's high school track and CSAT grade, respectively.

Stage 6: Earnings Once reaching the endogenously determined level of education $d_{i5} = \ell'$, the individual enters the work force. No further educational choices are assumed to be made after the sample period in the data, which means the labor market is an absorbing state. Let W_{it} denote the potential wage that individual i would realize at t-th year after reaching the education level ℓ' if she were to work. Following what has become known as the Mincer model, we assume that the log wages w_{it} is determined by the linear function of individual's accumulated human capital as of the beginning of year t:

$$w_{it} = \delta_{\ell' w} + X_i \beta_{6\ell'} + \gamma \, exp_{it} + \alpha^k_{6\ell'} + \varepsilon_{i6t} \, .$$

exp_{it} denotes potential experience (i.e., number of years after reaching the education level ℓ').[18] Schooling level and accumulated work experience would be the main sources of the earnings differentials in wages, but also observable individual characteristics may be.

2.3.3 Stochastic assumptions

There are two components that are not observed to the econometrician: idiosyncratic shocks and unobserved heterogeneity. Idiosyncratic shocks in the different stages, $(\varepsilon_1, \varepsilon_2, \varepsilon_3, \varepsilon_4, \varepsilon_5, \varepsilon_6)$, are assumed to be independent from each other. We assume that the vector of random shocks ε_s follows *i.i.d.* type 1 extreme value distribution for $s = 1, 2, 4, 5$.[19] Concerns about the assumption of independence between the shocks associated with different alternatives in a stage can be lessened by introducing a generalized extreme value distribution, which can produce a more flexible substitution pattern. For example, we could impose a nested structure over the college alternatives. Similarly, the assumption on the identical scale of the shocks within a stage could be also relaxed by heteroskedastic extreme value as described in Bhat (1995). The scalar shock ε_{3m} follows the standard logistic distribution. With the ordered alternatives, it could be natural that one alternative is more similar to the one close to it and less similar to another one further away. The natural ordering and asymmetric substitution toward proximate neighbors can be allowed for.[20] We assume $\{\varepsilon_{6t}\}$ are *i.i.d.* over the potential experience t distributed normal $N(0, \sigma^2)$.[21]

[18] We do not model nonlinearities due to the small number of period we observed on the labor market. Note also that we do not model the labor choices (e.g., nonemployment, work full-time, work part-time) along with the wages.

[19] The scale parameters are normalized to one in order to identify the parameters of the value functions. Equal scales across stages are not for a comparison of the value functions in different stages

[20] Small (1987) proposed the ordered generalized extreme value (OGEV) model that is more suited for ordered discrete outcomes and consistent with the random utility model. The OGEV model relaxes the restriction of independence between the unobservable characteristics across different choices, but involves a small number of additional parameters to estimate. We apply the specification to model the effort choice, which introduces only one additional parameter ρ ($\in (0, 1]$) compared to the standard multinomial logit model. $1/\rho$ captures the correlation between the adjacent outcomes. (The exact probability associated with a choice is found in Appendix.)

[21] It could be concerned that the restrictive specification not allowing heterogeneity in wage determination depending on one's educational qualification can result in biased estimates. Hence, we also estimate the model taking into account potential heterogeneous wage growth (i.e., $\gamma_{\ell'}$) and heteroskedasticity (i.e., $\sigma^2_{\ell'}$). The estimates, reported in Appendix, are quantitatively not far from the ones obtained from the current restricted specification.

We take into account the unobserved heterogeneity which is a permanent latent state generating differences across individuals. As it remains constant throughout all stages, the presence of the factor creates potential correlations across the choices and outcomes. Self-selection implies that, since rational individuals make optimization decisions based on state variables unobserved by the econometrician, observed economic relationships should be viewed as endogenous rather than as exogenous causal relationships. We apply the finite mixture approach which became popular in the dynamic discrete choice literature since it was first used in Heckman and Singer (1984). As specified in our model, each individual is one of a fixed K number of time-invariant discrete types (i.e., $\theta_i \in \{\theta_1, , ..., \theta_K\}$) with the population share of the type k is denoted by $\pi_k \in (0, 1)$ satisfying $\sum_{k=1}^{K} \pi_k = 1$. The individual's time-invariant type is assumed to be independent of the initial observable characteristics X_i. Importantly, our model drops the standard assumption in the existing literature that agents perfectly know their own type from the beginning, even though the econometrician does not. Conditional on our specification that beliefs depend on one's type, the standard determinants to identify type probabilities remain valid.[22] In the next section, we illustrate the identification of beliefs over the types as well as type probabilities by leveraging panel data and an additional variable of predictions that reveal the beliefs held by the individuals.

2.3.4 Parent prediction

We use the predicted schooling level data in a similar way to the use of revealed choice data in estimation. We assume that a parent shares the same information set with one's child at the end of stage 2.[23] That is, the prediction elicited by the parent of student i endowed with type k is conditioning on the information set up until stage 2: $\Omega_{i2} = \{X_i, \{d_{is}\}_{s=1,2}, \{p_2^k(r)\}_{r=1,...,K}\}$. We stress two features relevant to the prediction data. One concerns the timing when the belief is elicited. It is before the CSAT grade is realized

[22] Kasahara and Shimotsu (2009) demonstrate three important determinants of identification, which are the time-dimension of panel data, the number of values the covariates can take, and the heterogeneity of the response of different types to changes in the covariates.

[23] To make the model simple, the prediction by the parent is assumed not to reflect the general valuation on education but to indicate the assessment of their child's own capabilities. If this is the former case, we would need to account for parental influences on a child's academic achievement to make it in accordance with the valuation, which is considered in Gallipoli and Makridis (2018).

in stage 3 that the prediction based on the belief is made. The other feature is that the two alternatives, selective and non-selective academic colleges, are not distinguished in the survey questionnaire. Let us denote ES_i as the variable representing the final schooling level predicted at the end of stage 2. Provided that the sequential model is specified correctly, they compute the future probabilities to end up with final education level $\ell' \in \{N,\ V,\ NA,\ SA\}$, conditional on the information set Ω_{i2}, by weighing the type-specific future probabilities with the corresponding belief about own type $\{p_3^k(r)\}_{r=1,\ldots,K}$ (i.e., Bayes' theorem):

$$\begin{aligned} P^{*k}_{i5\ell'} &= P(d_{i5} = \ell' | \Omega_{i2}) \\ &= \sum_{r=1}^{K} p_3^k(r)\, P(d_{i5} = \ell' | \theta_i = \theta_r, X_i, \{d_{is}\}_{s=1,2})\,, \end{aligned} \tag{2.4}$$

where $p_3^k(r) = P(\theta_i = \theta_r | X_i, \{d_{is}\}_{s=1,2})$. The future probability conditional on being type r can be computed by integrating over the future idiosyncratic components upto the final education stage:

$$\begin{aligned} &P(d_{i5} = \ell' | \theta_i = \theta_r, X_i, \{d_{is}\}_{s=1,2}) \\ &\quad = \int\int\int I\{d_{i5}(\varepsilon_{i5}) = \ell' | \theta_i = \theta_r, X_i, \{d_{is}\}_{s=1,2}, Y_i(\varepsilon_{i3}) = y, d_{i4}(\varepsilon_{i4}) = \ell\} \\ &\quad \times I\{d_{i4}(\varepsilon_{i4}) = \ell | \theta_i = \theta_r, X_i, \{d_{is}\}_{s=1,2}, Y_i(\varepsilon_{i3}) = y\} \\ &\quad \times I\{Y_i(\varepsilon_{i3}) = y | \theta_i = \theta_r, X_i, \{d_{is}\}_{s=1,2}\}\, dF_{\varepsilon_3} dF_{\varepsilon_4} dF_{\varepsilon_5}\,. \end{aligned} \tag{2.5}$$

Due to the absence of the distinction between academic colleges by their selectivity in the elicited expected schooling data, we denote by $ES_i = A$ either selective or non-selective academic colleges. Then, by defining $P^{*k}_{i5,A} = P^{*k}_{i5,NA} + P^{*k}_{i5,SA}$, the probability of the predicted education level is a step function:

$$\begin{aligned} P(ES_i = \bar{\ell}' | \Omega_{i2}) &= P(ES_i = \bar{\ell}' | P^{*}_{i5,N}, P^{*}_{i5,V}, P^{*}_{i5,A}) \\ &= \begin{cases} 1 & if\ \bar{\ell}' = arg\,max\,\{P^{*k}_{i5,N}, P^{*k}_{i5,V}, P^{*k}_{i5,A}\}, \\ 0 & if\ \ otherwise, \end{cases} \end{aligned} \tag{2.6}$$

for $\bar{\ell}' \in \{N,\ V,\ A\}$.

Inspired by Van der Klaauw (2012), we smooth the function by introducing a parameter λ which represents the degree of misreporting.[24] Individuals are assumed, after calculating the future choice probabilities, to report an alternative with probability

$$P(ES_i = \bar{\ell}'|\Omega_{i2}) = \frac{exp(P^{*k}_{i5,\bar{\ell}'}/\lambda)}{exp(P^{*k}_{i5,N}/\lambda) + exp(P^{*k}_{i5,V}/\lambda) + exp(P^{*k}_{i5,A}/\lambda)}. \tag{2.7}$$

We can see that $\lim_{\lambda \to 0} P(ES_i = \tilde{\ell}'|\Omega_{i2})$ is equivalent to the function without the parameter λ, and $\lim_{\lambda \to \infty} P(ES_i = \tilde{\ell}'|\Omega_{i2}) = 1/3$, which could be interpreted as negligence on the response. Implicit assumption on this specification is that the noise is attributed purely to measurement error, but not to misspecification of the structural model.

2.4 Estimation

In this section, we discuss our strategy to identify the type probabilities and type-specific components, relaxing the perfect information assumption: type k individuals being equipped with $p^k_s(k) = 1$ and $p^k_s(r) = 0$ for $r \neq k$ from the initial stage $s = 1$. We build on the standard normalization that the first alternative ($j = 1$) is set as the reference among the $|J_s| + 1$ number of alternatives in each stage of the discrete choice. The coefficients that determine the value corresponding to the reference are normalized accordingly to zero (i.e., $v_{s1}(Z_{is}) + \alpha_{ks1} = 0$ for all $Z \in \mathbb{Z}$) and interpreted relative to the reference. Our strategy for identifying is constructive. We first identify types and how they impact behavior. Since agents are allowed to be uncertain about their own type, it captures the joint effect of all type-specific value functions and the belief they have over types. Next, we establish the formal identification argument for the type-specific probabilistic beliefs, allowing us also to recover the type-specific value functions. Finally, we set up the likelihood function closely following the identification strategy and clarifying the identification power of the supplementary prediction data we exploit.

[24] Introduction of the structural "noise parameter" is also found in the earlier paper, for instance, Holt and Laury (2002) who used it in order to allow some errors from the perspective of the deterministic expected utility theory model. Their differences from Van der Klaauw (2012) are the way to make use of the noise into their model as well as consideration of the dynamics.

2.4.1 Identification: unobserved heterogeneity type

With the notation in the previous section intact and ignoring any data on beliefs, our model is observationally equivalent to a usual semistructural model that controls for the dynamic selection via a mixture over unobserved types observed by the individual but unobserved by the econometrician. Then, the product of type probabilities and type-specific intercepts are identified by the dynamics of the model and exclusion restrictions. In other words, we can determine π_k and $\widetilde{\alpha}_{ksj}(=\sum_{r=1}^{K} p_s^k(r)\,\alpha_{rsj})$ for each unobserved heterogeneity type k, where the latter is handled as a whole for each unobserved heterogeneity type k, leaving its components $\{p_s^k(r), \alpha_{rsj}\}_{r=1}^{K}$ to be disentangled afterward by leveraging the prediction. Specifically, we impose dynamic exclusion restrictions, so that the autocorrelation of the choices and outcomes helps identification, given that the unobserved components across the stages are correlated (only) through the unobserved types.[25]

Table 2.6: List of variables in each stage s

Variables list	1st stage high school track	2nd stage effort	3rd stage CSAT	4th stage college entry	5th stage final education	6th stage wages
Individual characteristics						
female	V	V	V	V	V	V
household income quartile	V	V	V	V	V	V
parent college education	V	V	V	V	V	V
high school in urban	V	V	V	V	V	V
cohort	V	V	V	V	V	V
Choices/outcomes						
high school track		V	V	V	V	
effort			V			
CSAT				V	V	
college entry					V	
final education						V
potential experience						V

* While it is also plausible to assume that parent's education could be associated with the child's employment probability but not with the wage itself, we rather let identification not build on the exclusion restriction of the initial observable characteristics.

[25]We rather do not rely on the exclusion restrictions from the initial characteristics of the individuals. For example, parental income is often considered to appear in the college admission and graduation but not in the wage equation. However, there could be potential mechanisms behind the scene to bring about the correlation between the initial characteristic and wage. Additionally, functional form helps to identify types as well, though we stress that identification of the types is achieved by more than functional form.

Table 2.6 summarizes the list of variables relevant to each stage and illustrates the dynamic exclusion restrictions imposed in our model. Since the effort choice made in the stage 2 affects the postsecondary choice in stage 4 only through the CSAT score in stage 3, the correlation between the effort and college choices helps identification. An intuitive example would be, conditional on low CSAT grade, the choice pattern of high effort and college graduation may identify "college-favor type". Similarly, wages depend on the postsecondary education choice but not directly on the CSAT grade. Then, conditional on academic college graduation, low CSAT grade and low wages can contribute to the identification of a "low ability type".[26] Lastly, it is worth mentioning that the share of unobserved heterogeneity types is identified without the additional prediction data.

2.4.2 Identification: belief about own type

We identify the heterogeneous belief distribution by leveraging the predicted education level based on the statistical specification described in the previous section. Van der Klaauw (2012) makes use of expected occupation data for better identification of the heterogeneity types so that he could improve the precision of estimated parameters. Different from the efficiency gains he gets, we use similar data associated with one's expectation in a way to drop the perfect information assumption while allowing for potential information friction and heterogeneity in beliefs. The formal identification argument is provided by formulating linear equations.

Note that our parsimonious specification of the type-specific component allowed only the intercepts of the value functions to vary across types. Given that type-specific intercepts $\{\tilde{\alpha}^k_{sj}\}$ is identified, our identification objective is to solve the linear system of equation 2.3 and to recover the probabilities $\{p^k_s(r)\}$ with which type k individual believes himself/herself to be type r and the intrinsic type-specific coefficient $\{\alpha^k_{sj}\}$ associated with alternative $j\ (\in J_s)$ in stage s. Except for the intercept of normalized alternative, $\tilde{\alpha}^k_{s1} = p^k_s(1) \cdot 0 + \ldots + p^k_s(K) \cdot 0$, we have $J_s \times K$ linear equations in each stage $s = 1, 2, 4, 5$. Since the probabilities have to sum over r to 1 (i.e., $p^k_s(1) + \ldots + p^k_s(K) = 1$) conditional on

[26] A possible concern would be that CSAT grades may play a role of proxy for college selectivity which enhances productivity appraised in the labor market, which might be partly relieved by distinguishing college depending on their selectivity.

k, the stage s involves $(K-1) \times K$ number of probabilistic beliefs and $|J_s| \times K$ number of intrinsic intercepts.[27] In total, we have $(|J_1| + |J_2| + |J_4| + |J_5|) \times K$ linear equations with pre-identified parameters on the left-hand side while $\{(|J_1| + |J_2| + |J_4| + |J_5|)K + 4 \cdot (K-1)K\}$ unknown parameters to identify on the right-hand side.

We try to pin down the probabilistic beliefs at the point of the prediction made, $p_3^k(r)$, by leveraging the prediction probabilities in equation 2.7. Note that taking logs of both sides and differencing for two probabilities to predict different alternatives, for example $ES_i = N$ and $ES_i = V$, yields

$$lnP(ES_i = N|\Omega_{i2}) - lnP(ES_i = V|\Omega_{i2}) = (P^{*k}_{i5,N} - P^{*k}_{i5,V})/\lambda \,, \tag{2.8}$$

for all k. Here, $P(ES_i = \bar{\ell}'|\Omega_{i2})$ is observable conditional on the types being identified in advance. $P^{*k}_{i5\bar{\ell}'}$ is, as illustrated in equation 2.5, a function of the parameters α^k_{s4} and α^k_{s5}, which are of our identification interest along with $p_3^k(r)$ as well as λ. Three pairs of alternatives associated with ES_i, $\{(N,V),(N,A),(V,A)\}$, for each type k provide $3K$ number of the differences in log prediction probabilities, whereas the set of linear equations involves $(|J_4| + |J_5|) \times K + K(K-1) + 1$ parameters that need to be identified.

In order to deal with $|J_4| + |J_5|$ number of intercepts $\{\alpha^k_{sj}\}_{s=4,5}$, aiming at determining the probabilistic beliefs $p_3^k(r)$, we characterize the belief process over the stages in a particular way. The process, which can be extended later, is assumed to be simple as follows:

> [A1.] Uncertainty disappears from the end of stage 3 when academic ability is measured by the high-stakes test:
>
> $$p_s^k(k) = 1 \text{ and } p_s^k(r) = 0 \text{ for } r \neq k \text{ when } s > 3\,.$$

This assumption on the belief process restricts $\alpha^k_{sj} = \tilde{\alpha}^k_{sj}$ for all $j \in J_s$ at $s > 3$, which in turn, reduces the number of parameters in the linear system associated with equation 2.8 to $K(K-1)+1$. Hence, the assumption makes $(K-1)K$ probabilistic beliefs and one noise parameter identifiable once the system of linear equations has a solution $\{p_3^k(r)\}_{k,r}$.

[27] As an illustration, assume that there are two types (i.e., $K = 2$). In stage 1, where there are two alternatives, we have 2 probabilistic belief parameters, $p_1^1(1)$ and $p_1^2(1)$, and two parameters associated with the choice equation, α^1_{12} and α^2_{12}. Note that $p_1^1(2) = 1 - p_1^1(1)$ and $p_1^2(2) = 1 - p_1^2(1)$.

We get the largest integer $K = 3$ that satisfies the necessary condition for a unique solution:

$$3K \geq (K-1)K + 1 \,. \tag{2.9}$$

The intuition behind it is that any systematic mistake in prediction beyond the pure measurement error is captured by $p_3^k(r)$, at least if we know about the different future behavior of different types represented by $\{P_{i5\tilde{l}'}^{*r}\}$. Importantly, the result does not prevent us from introducing more than three discrete heterogeneity types. Since the prediction depends on the information set Ω_{i2}, equation 2.8 has to hold for each value of X_i $(\in \Omega_{i2})$ whereas $p_3^k(r)$ is independent of X_i. In other words, the number of equations increases with the possible realization of the elements in X_i. Therefore, the number of types is not bounded by three but enlarged as the set of initial characteristics expands.

The last step is, by taking advantage of equation 2.3, to recover $(|J_1| + |J_2|) \times K$ number of intrinsic intercepts and $2 \cdot K(K-1)$ number of probabilistic beliefs from $(|J_1| + |J_2|) \times K$ linear equations. We additionally impose the assumption that characterizes the rest of the belief process as follows:

[A2.] Probabilistic belief stays constant over the first three stages:

$$p_1^k(r) = p_2^k(r) = p_3^k(r) \,,$$

Given that $p_3^k(r)$ is identified under Assumption [A1] with a reasonable integer K, Assumption [A2] reduces the unknown parameters, which enables the intrinsic intercepts to be identified by solving the system of linear equations.

Note that the assumptions A1 and A2 that are made here are easy to relax. First, if belief data are available in multiple periods, we could allow beliefs to be different in these periods too. Alternatively, we could allow for an intermediate period in which beliefs are updated with a specific functional form, e.g. a Bayesian update after observing their grade. As our counterfactual exercise provides students with the information they now have right before choosing college, we believe the current approach is sufficient to inform us on the relevant degree of uncertainty.

2.4.3 Likelihood function

We construct the individual likelihood function for the sequential model, closely following the identification strategy discussed in the previous subsection. Because individuals are one of K unobserved types, we account for this unobserved heterogeneity by constructing the type-specific likelihoods and integrating out the unobserved type in the standard way. For a particular individual i, conditional on her unobserved heterogeneity type being k, the contribution to the likelihood when she chooses high school track h and effort k, reports desired education level ES, gets the CSAT grade y, initially has higer education in l, reaches the educational level l', and receives a series of log wages w is

$$\begin{aligned}\mathcal{L}_i^k(\theta_k, p_1^k, p_2^k, \lambda) =& \mathcal{L}_{i1h}^k(\widetilde{\alpha}_1^k, \beta_1) \cdot \mathcal{L}_{i2k}^k(\widetilde{\alpha}_2^k, \beta_2, \delta_h) \cdot \mathcal{L}_{i3y}^k(\alpha_3^k, \beta_3, \delta_h, \delta_e, \eta) \\ & \cdot \mathcal{L}_{i4l}^k(\widetilde{\alpha}_4^k, \beta_4, \delta_h, \delta_y) \cdot \mathcal{L}_{i5l'}^k(\widetilde{\alpha}_5^k, \beta_5, \delta_h, \delta_y, \delta_\ell) \cdot \mathcal{L}_{i6w}^k(\alpha_6^k, \beta_6, \delta_{\ell'}, \gamma, \sigma_{\ell'}^2) \\ & \cdot \mathcal{L}_{iES}^k(p_3^k, \lambda, \alpha_3^k, \beta_3, \delta_h, \delta_e, \eta, \alpha_4^k, \beta_4, \delta_h, \delta_y, \alpha_5^k, \beta_5, \delta_h, \delta_y, \delta_\ell) \,. \end{aligned} \tag{2.10}$$

As described in the previous subsection of our identification strategy, we estimate the type-specific beliefs p_s^k based on the predicted education through the likelihood contribution $\mathcal{L}_{iES}^k$. Under Assumption A2, even if the probabilistic beliefs are involved in the likelihood functions in stage 1 and stage 2, $\mathcal{L}_{i1k}^k$ and $\mathcal{L}_{i2k}^k$, through the weights on the intrinsic intercepts $\widetilde{\alpha}_{ksj} = \sum_{r=1}^{K} p_s^k(r)\, \alpha_{rsj}$, we estimate the composite as a whole using the following likelihood function:

$$\mathcal{L} = \sum_i log \sum_k \pi_k \mathcal{L}_i^k \,, \tag{2.11}$$

where π_k denotes the population probabilities of being a particular type k. We use an expectation–maximization (EM) algorithm to obtain good starting values.

Subsequently, we solve the linear system of equation 2.3 so that the beliefs are disentangled from the intrinsic coefficients. This approach is straightforward, but its drawback is the lack of standard errors of the solution to carry out statistical inferences.

2.5 Empirical results

In our empirical analysis, we set the number of unobserved heterogeneity types to two (i.e., $K = 2$) and estimate the model using a subset of the Korean data. As the number of types is limited, we make the sample more homogeneous by dropping the individuals who do not graduate from any college.[28] The choice sets associated with the final qualification in stage 5 and the expected education level are adjusted accordingly. [29] We present the parameter estimates that fit the model to the observed data pattern quite well, and carry out the counterfactual policy analysis to study the effect of information provision to the individuals before they make their high school track decisions.

2.5.1 Estimated parameters

From the estimation result in Table 2.8, the key parameters of our main interest are the probabilistic beliefs and type probabilities. Before delving into the interpretations of them, we can define Type 1 individuals identified as "low type". They are less likely to choose science-focused track, put lower effort, having lower CSAT grade (conditional on effort), graduate from non-selective academic college, and receive lower returns to college. In the population, there are 41% low type individuals. Importantly, the beliefs indicate that that those who are truly low type believe they have a 60% chance to be low type. Since it is higher than the population probability, it shows that the agents are better informed about their type than the econometrician. However, since it is not 100%, it points to substantial uncertainty. Similarly, the high type individuals perceive themselves to be high type with a probability of 70%.

[28]When we did not do that, we obtained a college and non-college type which leads to less relevant counterfactual simulations. In future work, we will allow for more types instead of dropping observations to avoid the biases caused by dropping observations based on one of the outcomes of the model.

[29]There are three remaining choice options in stage 5: vocational college degree (V), non-selective academic college degree (NA), and selective academic college degree (SA). When the prediction is made, some expect not to have a college education (N), which is not consistent with the adjusted model. Then, we impute the prediction with vocational college (V) which is thought of as the closest alternative.

Table 2.7: Parameter Estimates

Parameters	Estimate
Type 1's belief, p_1^2	0.61093
	(0.092199)
Type 2's belief, P_1^2	0.29596
	(0.10943)
Type 1's share, π_1	0.41255
	(0.024641)
Measurement error, λ	0.23937
	(0.021025)

2.5.2 Model fit

We investigate the fit of the model by forward simulating each individual in the estimation sample based on the parameter estimates. Specifically, we begin by drawing an unobserved type for each individual from a distribution with the parameter value of estimated unobserved type probability. We then draw idiosyncratic shocks and compute the alternative-specific value terms to simulate one's choice from the first stage on. We perform this forward simulation 10 times for each individual.[30] Table 2.11 reports the observed and simulated choices. The column of *model fit* shows that the simulated frequencies based on our model estimates are quite close to the empirical ones, succeeding in quantitatively fitting the observed data patterns. The results support us to proceed with counterfactual policy analysis with the model estimates in the next subsection.

2.5.3 Counterfactual policy

To get a more precise view of the effect of information on the educational choices, we run counterfactual simulation that considers perfect information associated with one's latent state from the beginning. In Table 2.11, the column denoted by *counterfactual* represents the counterfactual educational choices and outcomes under the perfect information. On average, the impacts are small. We do see less people going for the liberal art track, and more for the vocational and science tracks. This suggest that more information implies sorting to more specialized tracks.

[30] Practically, we duplicate the individuals in the estimation sample and simulate their choices/outcomes at a time.

Table 2.8: Multinomial logit coefficients (stage 1,2,4,5)

	High school track		Effort		College enrollment			Final qualification	
Variables	Science track	Vocational track	Middle	High	Vocational	Non-selective	Selective	Non-selective	Selective
	(Base: Liberal-art track)		(Base: Low)		(Base: not enrolled in any college)			(Base: vocational college)	
female	−0.69186	−0.35301	0.16932	0.05502	0.92667	0.39367	0.30030	−0.32040	−1.01010
	(0.10017)	(0.13414)	(0.12931)	(0.10921)	(0.15651)	(0.13784)	(0.15392)	(0.18162)	(0.24502)
household income quartile	0.08182	−0.20142	−0.00439	0.12488	−0.35705	−0.24610	−0.32149	0.23192	0.30149
	(0.04694)	(0.06267)	(0.05928)	(0.05125)	(0.07333)	(0.06432)	(0.07280)	(0.08328)	(0.11390)
college educated parent	0.07796	−1.03480	0.20442	0.23979	−0.62920	−0.04851	0.11099	−0.12500	0.18625
	(0.11141)	(0.19270)	(0.14677)	(0.11949)	(0.19369)	(0.14967)	(0.16502)	(0.23368)	(0.28312)
urban residence	0.11908	0.10039	0.08626	0.38058	−0.22719	−0.56482	−0.51530	0.12562	0.17320
	(0.10020)	(0.13555)	(0.12854)	(0.10814)	(0.15223)	(0.13507)	(0.15220)	(0.17870)	(0.23870)
cohort	0.11813	0.92047	0.39632	−0.66861	−0.57717	−0.55477	−0.50625	0.52902	0.40827
	(0.10257)	(0.15759)	(0.14323)	(0.10871)	(0.17219)	(0.14599)	(0.16154)	(0.20860)	(0.26548)
CSAT grade					−0.62942	−0.32178	0.44759	0.35439	0.79314
					(0.05856)	(0.04733)	(0.05413)	(0.07167)	(0.09165)
College enrollment :									
not enrolled								−2.96220	−4.52370
								(0.29274)	(0.59312)
vocational college								−5.21910	−9.20720
								(0.29679)	(0.81749)
non-selective academic college									−4.70300
									(0.61232)
selective academic college								−1.73100	
								(0.62351)	
Type 1's Intercept	−1.7142	−0.6405	−1.7953	−1.6942	4.72600	3.72220	−0.96960	0.20443	−0.47006
	(-)	(-)	(-)	(-)	(0.37710)	(0.33333)	(0.38607)	(0.41857)	(0.72620)
Type 2's Intercept	0.3629	−1.5372	−1.6865	−0.3584	3.85610	3.81910	−1.32850	2.07480	0.98116
	(-)	(-)	(-)	(-)	(0.40758)	(0.34355)	(0.41250)	(0.47261)	(0.79203)

Table 2.9: Ordinal logit coefficients for CSAT grade (stage 3)

Variables	CSAT grade
female	0.28117
	(0.077537)
household income quartile	0.16870
	(0.036188)
college educated parent	0.73127
	(0.090624)
urban residence	0.17490
	(0.077080)
cohort	0.10932
	(0.082557)
High school: science track	0.071637
	(0.088367)
High school: vocational track	-2.2059
	(0.12988)
Effort: middle	0.74987
	(0.11141)
Effort: high	0.92933
	(0.09563)
Type 1's Intercept	2.8202
	(0.18995)
Type 2's Intercept	3.4218
	(0.19734)
Threshold 2	1.6773
	(0.11531)
Threshold 3	2.9587
	(0.13123)
Threshold 4	4.0112
	(0.13947)
Threshold 5	5.0830
	(0.14673)
Threshold 6	6.1050
	(0.15512)
Threshold 7	7.0406
	(0.16666)
Threshold 8	8.6062
	(0.21504)

Table 2.10: ln(wage) parameters by qualification group (stage 6)

Variables	Vocational college	Non-selective academic	Selective academic
female	-0.10302	-0.20874	-0.14857
	(0.019485)	(0.016446)	(0.024899)
household income quartile	-0.004691	0.038317	0.0045971
	(0.010837)	(0.007986)	(0.013249)
college educated parent	0.027264	0.086017	0.14962
	(0.029735)	(0.020072)	(0.025533)
urban residence	-0.0096783	-0.015753	0.021866
	(0.018519)	(0.015887)	(0.027964)
cohort	-0.052233	-0.054589	0.0077834
	(0.020373)	(0.019633)	(0.029848)
Type 1's Intercept	5.0899	4.7713	4.8299
	(0.04125)	(0.028498)	(0.051471)
Type 2's Intercept	4.7653	5.1152	5.2491
	(0.04125)	(0.028498)	(0.051471)
Years of experience		0.049681	
		(0.002703)	
Wage shock, σ		0.24387	
		(0.0028035)	

The overall effects combine the impact of making high-ability students more aware of their high ability, and the low-ability more aware of their low ability. It is therefore very likely that their effects cancel out on average, but providing information can still have large distributional consequences. Table 2.12 splits results by one's true type. We find quite large differences for each type. Type 1 low-ability students become more likely to choose a liberal-art or vocational track in high school, exerting lower effort and yielding lower CSAT grades. In contrast, the share of type 2 high-ability students attending the science-focused track increases, they obtain higher CSAT grades after exerting more effort.

To conclude, we find that providing perfect information leads to stronger sorting in high school. This has a long run impact as we see more low-ability students graduating from the vocational college, while the high-ability increase their chances to graduate from the selective academic college.

Table 2.11: Model Simulation

	Data		Model fit		Counterfactual	
	Count	Share	Count	Share	Count	Share
High school track						
Liberal-art track high school	1131	52.58%	11336	52.70%	10360	48.16%
Science track high school	716	33.29%	7186	33.41%	7995	37.17%
Vocational high school	304	14.13%	2988	13.89%	3155	14.67%
Effort						
Low effort	1253	58.25%	12460	57.93%	12086	56.19%
Middle effort	321	14.92%	3210	14.92%	3060	14.23%
High effort	577	26.82%	5840	27.15%	6364	29.59%
CSAT grade						
1	74	3.44%	698	3.25%	780	3.63%
2	205	9.53%	1968	9.15%	2011	9.35%
3	314	14.60%	3098	14.40%	3015	14.02%
4	379	17.62%	3801	17.67%	3703	17.22%
5	443	20.60%	4623	21.49%	4553	21.17%
6	351	16.32%	3476	16.16%	3462	16.09%
7	203	9.44%	2016	9.37%	2073	9.64%
8	139	6.46%	1407	6.54%	1462	6.80%
9	43	2.00%	423	1.97%	451	2.10%
College enrollment						
High school graduation	341	15.85%	3476	16.16%	3477	16.16%
Vocational college	479	22.27%	4656	21.65%	4710	21.90%
Non-selective academic college	868	40.35%	8674	40.33%	8511	39.57%
Selective academic college	463	21.52%	4704	21.87%	4812	22.37%
Final qualification						
Vocational college	521	24.22%	5005	23.27%	5101	23.71%
Non-selective academic college	1016	47.23%	10222	47.52%	10022	46.59%
Selective academic college	614	28.54%	6283	29.21%	6387	29.69%
Expected education level						
Vocational college graduation	386	14.11%	200	14.62%	260	19.01%
Academic college graduation	2350	85.89%	1168	85.38%	1108	80.99%

Table 2.12: Choices and Outcomes Distributions by Type

	Model fit		Counterfactual	
	Type 1	Type 2	Type 1	Type 2
High school track				
Liberal-art track high school	57.35%	49.44%	62.19%	38.33%
Science track high school	24.70%	39.51%	11.59%	55.10%
Vocational high school	17.95%	11.05%	26.22%	6.57%
Effort				
Low effort	62.79%	54.52%	71.22%	45.66%
Middle effort	15.33%	14.64%	15.38%	13.42%
High effort	21.88%	30.85%	13.40%	40.93%
CSAT grade				
1	4.74%	2.20%	6.34%	1.72%
2	12.25%	6.97%	14.92%	5.45%
3	17.74%	12.07%	18.72%	10.72%
4	19.52%	16.38%	18.97%	15.99%
5	20.02%	22.53%	18.41%	23.10%
6	13.21%	18.23%	11.94%	19.01%
7	7.50%	10.68%	6.52%	11.82%
8	3.88%	8.41%	3.17%	9.34%
9	1.14%	2.55%	1.02%	2.85%
College enrollment				
High school graduation	13.54%	18.00%	12.98%	18.40%
Vocational college	31.81%	14.53%	34.56%	13.02%
Non-selective academic college	35.10%	43.99%	35.18%	42.64%
Selective academic college	19.55%	23.49%	17.29%	25.94%
Final qualification				
Vocational college	38.38%	12.67%	41.60%	11.18%
Non-selective academic college	36.32%	55.37%	35.81%	54.15%
Selective academic college	25.30%	31.95%	22.59%	34.67%
Expected education level				
Vocational college graduation	21.36%	9.72%	38.99%	4.63%
Academic college graduation	78.64%	90.28%	61.01%	95.37%

2.6 Conclusion

This research examines the sequential process of educational investment decisions, particularly in the context of South Korea, where high educational attainment coexists with significant academic pressure and concerns about potential misallocation of talent. Our study departs from the conventional assumption of perfect information regarding students' own unobserved heterogeneity, such as abilities and preferences, instead exploring how uncertainty and heterogeneous beliefs about it influence educational decisions and outcomes. We build a dynamic discrete choice model designed to capture the sequential nature of educational choices, from high school track selection to eventual educational attainment. The model assumes that one's unobserved heterogeneity, which is typically unobserved by the econometrician, is also unobserved to the economic agent. A key methodological innovation involves leveraging readily available prediction data, specifically expectations of the final educational qualification, to identify and characterize individuals' probabilistic beliefs about their own unobserved heterogeneity. This approach allows us to disentangle the impact of the beliefs from intrinsic type-specific preferences in shaping educational outcomes.

The empirical analysis, utilizing the Korean Education and Employment Panel (KEEP) data, provides compelling evidence that uncertainty is a nontrivial factor underlying educational decisions in Korea. Our findings suggest that students may not be perfectly informed about their own unobserved heterogeneity and could potentially face uncertainty stemming from the latent state when making critical educational choices early in their academic careers. This can lead to investment decisions that are not optimally aligned with their comparative advantages, potentially resulting in systematically lower ex-post returns than initially anticipated. Furthermore, our counterfactual simulations highlight the potential impact of information provision. By simulating a scenario of perfect information regarding students' latent types, we demonstrate that reducing information friction could lead to a better sorting of students into educational tracks that are more suited to what they are endowed with, ultimately improving overall efficiency in the allocation of students.

This paper contributes to the literature by providing a novel framework for identifying heterogeneous beliefs nesting the perfect information assumption which is standard in the dynamic discrete choice literature, particularly when only limited belief data (such as point predictions) are available. By relaxing the common assumption of perfect information, we offer a more nuanced understanding of the factors driving educational decisions and the potential consequences of misperceptions. However, the study has certain limitations that warrant consideration for future research. These include the challenges of obtaining selection-free measurements for outcomes like CSAT grades and wages. Even though they are only observed for individuals who take the test or enter the workforce, respectively, we abstract from the selection issue. The assumption that students become fully informed about their types after receiving CSAT grades, with beliefs remaining static during other transitional periods, is another simplification. Addressing these limitations in future work would further enhance the understanding of the complex interplay between beliefs, educational investment, and human capital formation in general.

Chapter 3.

The Role of Learning in Occupational Decisions and Wage Dynamics

Gyung-Mo Kim

Abstract: If workers are uncertain about their labor market productivity, this uncertainty could have crucial implications for occupational sorting and wage dynamics. This paper starts by documenting the key features in the US labor market that can be explained by the presence of uncertainty about workers' own skills and the learning process that unfolds depending on specific experiences. A structural model is built on these empirical patterns and characterizes that workers in their initial occupations decide each year whether to stay or quit in response to the wage outcomes while accumulating and learning about their skills. The dynamic discrete choice model, extending conventional discrete-time duration analysis, is estimated using the Kalman filter and a conditional choice probability estimator. I find supporting evidence that workers experience significant uncertainty about their skills. The results depict that the difference in the dynamics of wage distribution across occupations can be attributed to workers' sorting as well as human capital accumulation in different occupations. Counterfactual analysis shows that, in occupations with non-trivial probabilities of mismatch due to the choices based not on the true skills but on the beliefs, an information provision policy has a sorting effect. By reducing the share of the mismatched, the policy influences employment duration in the initial occupations, resulting in higher wage growth and lower wage dispersion.

3.1 Introduction

The human capital framework over the life cycle (e.g., Ben-Porath 1967; Mincer 1974) has been the basis for modeling earnings and provided good explanations consistent with observed wage patterns. The human capital investment decision problem is inherently characterized by uncertainty over future returns, similar in kind to any risky investment decision. Uncertainty, particularly stemming from individual-specific endowment such as one's own abilities, can lead to suboptimal investment decisions that may be viewed as mistakes. For instance, a worker might switch occupations upon discovering that their performance is rewarded differently than expected.[1] This raises several key questions: Do individuals experience information friction and learn about their own endowment through experience? To what extent does imperfect information serve as a barrier to exploiting their comparative advantages? And what impact would providing information to individuals have on wage distribution and its dynamics?

This paper addresses these research questions by focusing on occupational mobility in the labor market, which is frequent among young workers and has significant implications for wage levels, growth, and dispersion. Occupational decisions may not only allow individuals to seek better matches for their skills but also play a crucial role in skill development and, consequently, earnings potential. Recent research suggests that workers' learning is crucial (e.g., Papageorgiou 2014; Groes et al. 2015) to understand occupational decisions and mobility. They highlight that uncertainty about workers' time-invariant abilities and learning processes can explain workers' sorting patterns across occupations. The pattern, illustrated in this paper, that male workers with longer tenure in their first full-time occupations are less likely to quit suggests that learning about one's productivity from the experience could play a role in occupational decisions. I present an additional empirical pattern that the decision to stay in or quit their initial occupations is associated with new information obtained from wages in the previous period. This supports that learning about one's earning potential based on the observed wages could be a relevant mechanism for the occupational decision.

[1] Other examples relevant to uncertainty about one's attribute may include dropping out of high school track or college and working in occupations that typically do not require a college degree while having graduated from college.

To separate the potential learning process from different mechanisms that may influence occupational mobility, I consider a structural model that jointly treats the wages and duration of staying in an initial occupation in learning environments. The model accounts for workers' selective attrition arising due to Bayesian learning about their own productivity, which generalizes conventional discrete-time duration models by incorporating richer time-series properties for unobservables. Specifically, this describes how workers could learn about their human capital (hereafter simply "skill") based on individual-specific signals and subsequently make the occupational decision of either staying in or quitting their current occupation. Workers face uncertainty about initial skill in the occupations they are exogenously assigned to, and compute the expected lifetime gains of staying working by forming expectations with respect to their current skill and its accumulation in the future. I assume that the flow utility of working in the current period is the constant relative risk aversion (CRRA) function of the wage with a relative risk aversion parameter set to one, where the wages are closely tied to workers' output production. Then, the observed wage at the end of each period signals the worker's skill, and the belief is updated accordingly in a Bayesian manner.[2] The skill accumulates stochastically through learning by doing. The econometric specification allows for explicitly testing the hypothesis of Bayesian learning against initially known ability.

Identification of the parametric model in this paper is a special case of the non-parametric identification results in the literature. Bunting et al. (2024) provide identification results for potential outcome equations and ability distributions in a general class of learning models. Their identification results can be extended and applied in the presence of selective attrition. Then, under an identification assumption that workers do not possess private information conditional on what the econometricians can observe, standard identification results in the dynamic discrete choice literature hold in the learning model of this paper. The identification assumption relies on the main virtue of the NLSY79 which contains selection-free measurements that allow workers' skills before entering the labor market to be controlled for. Once identified, the model is estimated using the Kalman filter

[2] In each occupation, the more output depends on the unobserved ability and the less volatile productivity shock is, the more information about the abilities would be revealed on the job. This may generate experimentation incentives for workers who value information. However, for identification, the experimentation incentive through risk aversion is rather fixed, while correlated learning about their multi-dimensional occupation-specific skills is abstracted.

and conditional choice probability (CCP) estimation method. My approach enables the unobserved skills to be addressed, while the burden of solving the dynamic programming problem is avoided by exploiting the terminal state nature inherent in this optimal stopping problem.

Estimates from the US labor market provide statistically significant results rejecting the null hypothesis that workers initially know about their own skill associated with their initial occupation. Hence, learning about own uncertain skill affects the decisions of the workers to quit their occupations that equip different wage determination and skill accumulation processes, rendering different learning environments. The differences influence the dynamics of the mean wage and within-occupation wage dispersion across occupations. To separate the effect of selective attrition through learning that brings about workers' sorting, I run a counterfactual policy simulation. Given that a nontrivial share of workers is found to be mismatched in the status quo, a policy, such as counseling, that can remove the uncertainty from the start of the workers' careers is considered. With the counterfactual policy in effect, workers in their initial occupations make choices based not on beliefs but on true skills. I find that perfectly informing workers mitigates occupational mismatch and affects the duration of the first employment relationship, which in turn, enhances wage growth and reduces wage dispersion through the sorting effect that low-skill workers leave their initial occupations early.

This paper contributes to two strands of literature that are central to understanding labor market outcomes: occupational decisions in learning environments and wage dynamics. The first strand of literature deals with uncertainty particularly about individuals' abilities and learning processes. Recent research using data that directly elicits individuals' beliefs has confirmed that they have uncertainty about their own personal traits and form biased beliefs (e.g., Stinebrickner and Stinebrickner 2014; Hoffman and Burks 2020; Bobba and Frisancho 2022), which significantly influences decision-making. However, the idea of imperfect information and learning is not new to understanding occupational choice and career progression. Jovanovic (1979) pioneered the learning model that captures the job transition patterns, and Miller (1984) generalizes the model by introducing different job types that provide different expected per-period benefits and information about future returns. Antonovics and Golan (2012) extend the model by incorporating correlations between occupation-specific productivity determined by the characteristics of the occupations.

There are several views to understanding the observed patterns of occupational mobility; skill accumulation (Jovanovic and Nyarko 1997; Sanders 2014), a search process for a better match (Jovanovic 1979; Neal 1999), and a process of learning about one's own ability (James 2012; Gorry et al. 2019). This paper focuses on learning about own skill that accumulates stochastically, keeping other mechanisms implicit and abstracting from general equilibrium effects. By building an easily comprehensible model that links the learning to occupational choice in the early career, I provide direct evidence that supports the presence of uncertainty and learning about own skill. Two papers, James (2012) and Arcidiacono et al. (2025), are the closest to this work, but I carry out a more detailed investigation into wage dynamics within and across different occupations through workers' sorting based on learning and human capital accumulation.

The second strand of literature is about the sources concerning wages distribution and its dynamics. Rubinstein and Weiss (2006) provide an extensive review of the studies about various economic explanations of wage growth over the life cycle: human capital investment, search and learning. Whereas not everyone accepts the human capital framework as the basis for modeling wages, the approach is surprisingly robust compared to other models in explaining wage patterns (Polachek et al. 2015). Building on Ben-Porath (1967) who established a framework that links wage growth to human capital accumulation, Magnac et al. (2018) further take into account individual heterogeneity influencing human capital investment decisions. This extension links individual heterogeneity in the human capital process to wage dispersion as well as wage growth, offering an explanation of the decrease in wage dispersion in workers' early careers followed by an increase. Taber and Vejlin (2020) quantify the wage variations attributed to several different sources incorporated in one model and find that heterogeneous pre-market skill is the most important source of wage variation. This paper considers individual heterogeneity not only in pre-market skill but also in its evolution through learning by doing along with learning about the uncertain skill. I find that wage variation explained by initially unknown skill is comparable to the variation attributed to idiosyncratic shocks. The decrease in wage dispersion in some occupations along the workers' early career could come from the sorting of workers. In other words, the workers who remain working in occupations involving less volatile human capital accumulation and providing more precise information have relatively similar skill levels and so the less dispersed wages as they choose to stay in their early careers.

The remainder of the paper is organized as follows. Section 2 presents the data and provides descriptive evidence suggestive of learning. Section 3 describes a dynamic model of occupational decisions, where workers have imperfect information about their labor market productivity, and update their beliefs through the wage observations while accumulating skills through learning by doing. The model identification and relevant econometrics issues are also discussed. of the model, Section 4 details the estimation procedure and presents the estimation results. Section 5 studies the role of informational frictions on educational and labor market outcomes. Finally, Section 6 concludes.

3.2 Data and descriptive statistics

In this paper, I focus on male workers in the US labor market and document the patterns of earnings dynamics and occupational choices, specifically whether they stay in or quit their occupations. The primary data source for this analysis is the NLSY79, which tracks a nationally representative sample of individuals starting in 1979 and provides detailed employment information for each worker. I begin by briefly describing the data sources and presenting key sample statistics, with additional details regarding the sample selection criteria provided in the Appendix. In the following subsections, I present descriptive statistics that highlight the empirical features motivating and shaping the theoretical framework of this paper. These statistics illustrate the diverse patterns of earnings dynamics across occupations and suggest that learning about one's own skill could be crucial in understanding workers' occupational choices. Additionally, I emphasize that wages, based on workers' responses to realized wage shocks, may serve as a key source of information for this learning process.

3.2.1 Data: 1979 National Longitudinal Survey of Youth (NLSY79)

I utilize the Work History Data File of the NLSY79 to construct annual panels that track individuals aged 14 to 22 as of January 1, 1979. This single cohort with a large sample size allows researchers to study individuals who have faced relatively similar economic conditions over their life cycles.[3] The dataset provides information on the usual hours worked per week and the number of weeks worked per job within a year, enabling the calculation of total hours worked for each job annually. I define the primary job for each individual in a given year as the one with the highest number of hours worked. This primary job is supplemented with detailed employment information, including occupation title and earnings. To ensure consistency in occupational classifications across years, all occupational codes are converted into the Census 1990 Three-Digit Occupation Codes. Worker wages are measured by the rate of pay for the primary job, including tips, overtime, and bonuses before deductions, and are converted to a weekly rate. All wages are deflated by the Personal Consumption Expenditures Price Index and expressed in 2000 dollars.

Based on responses to employment questions, each individual is assigned to one of seven mutually exclusive activities each year: nonemployment or full-time employment in one of six occupational categories. These categories are defined by aggregating occupations that encompass the entirety of US employment. I follow Autor and Dorn (2013) and others to classify the occupational titles coded in three digits into broader groups.[4] Two groups, Executive/Managerial and Professional/Technical occupations, represent highly educated and well-paid jobs. In 1979, between one-quarter and two-thirds of workers in these occupations held at least a four-year college degree. Sales/Administrative Support represents a middle-skilled, white-collar occupational group predominantly held by women with a high school diploma or some college education. Additionally, there are groups of middle-

[3] A common challenge in this context is to distinguish between cohort, time, and age effects. Because of the relatively narrow age range of participants initially surveyed by the NLSY79, it would be appropriate to view this data as capturing outcomes for one specific cohort of individuals over their lifecycle. Conditional on the single cohort, the model in the next section will incorporate time effect in the form of variation in the output price across calendar years. Then, the age effect only through occupational tenure can be identified by using the variation in different ages of entry into the labor market.

[4] For example, Autor and Dorn (2013) divide the occupations into the groups of Managers, professionals, technicians, finance, public safety; Production and craft; Transportation, construction, mechanics, mining, farm; Machine operators, assemblers; Clerical, retail sales; Service. The authors show that the distinct groups require different abstract routine, manual, task inputs.

and low-skilled blue-collar occupations, typically held by men with a high school degree or less education, including Production/Operators and Transportation/Construction. Lastly, the Low-skill Service category encompasses occupations in protective services, food preparation, cleaning, and personal care. Most workers in service occupations have no post-secondary education, although employment in this group has grown rapidly over the past several decades when post-secondary education has expanded fast.

A main virtue of the NLSY79 is that it includes measures that allow workers' skills before entering the labor market to be potentially controlled for. All respondents took the Armed Services Vocational Aptitude Battery (ASVAB) test at the start of the survey, which measures various cognitive skills. I construct each measure of verbal, math, and craftsmanship skills using the following two test components, respectively: Word Knowledge, Paragraph Comprehension, Arithmetic Reasoning, Mathematics Knowledge, Mechanical Comprehension, and Electronics Information. In addition, respondents were surveyed in 1979 and 1980 regarding their non-cognitive attitudes. Following literature emphasizing the importance of non-cognitive skills on earnings, I use the Rotter Locus of Control Scale and the Rosenberg Self-Esteem Scale to measure social skills. To construct scalar values for these different skill dimensions, I first adjust the mean and variance of each score across ages to account for systematic age effects, following the method of Altonji et al. (2012b). I then apply Principal Component Analysis (PCA), normalizing the score associated with the first principal component to have a mean of 0 and a variance of 1.

I restrict the sample to males working in their first occupation after completing their education in order to examine occupational decision—stay in or quit— and wage patterns for a relatively homogeneous group over a period of time. This allows me to control for social differences (e.g., fertility) and not to delve into the relationship between wages and hours worked. Following Guvenen et al. (2020), I consider only workers "attached" to the labor force, restricting the sample to full-time workers with more than 1,200 imputed working hours in a given year. Additionally, to avoid issues with left truncation and to construct complete work histories, I exclude individuals who were already working at the start of the sample period. Further details on the sample selection process are provided in Appendix A. My final sample includes 2755 individuals and 17,036 individual-year observations.

Table 3.1: Summary Statistics by Occupation

	Executive/ Managerial	**Professional/ Technical**	**Sales/ Admin support**	**Production/ Operators**	**Transportation/ Construction**	**Low-skill Service**
# Obs	2939	3372	1662	2081	5568	1414
# Individuals	313	358	340	378	991	375
Average age at entry	23.39	24.788	22.374	21.087	20.936	21.413
Education						
$\leq$ High School	.16	.084	.412	.786	.781	.667
Some college	.243	.17	.279	.138	.172	.24
$\geq$ 4-year college	.597	.746	.309	.077	.047	.093
Race						
Hispanic	.15	.151	.185	.18	.172	.136
Black	.153	.154	.238	.257	.233	.355
Non-Hispanic/Non-Black	.696	.696	.576	.563	.595	.509
Skill measures						
Average verbal skill	.647	.735	.283	-.322	-.261	-.276
Average math skill	.706	.905	.251	-.322	-.323	-.297
Average social skill	.331	.276	.117	-.172	-.108	-.14
Duration (year)						
Average duration	9.39	9.419	4.888	5.505	5.619	3.771
Median duration	6	6	3	3	3	2

Note: Annual records could stop being observed in the sample, which yields incomplete duration. They are treated as missing at random.

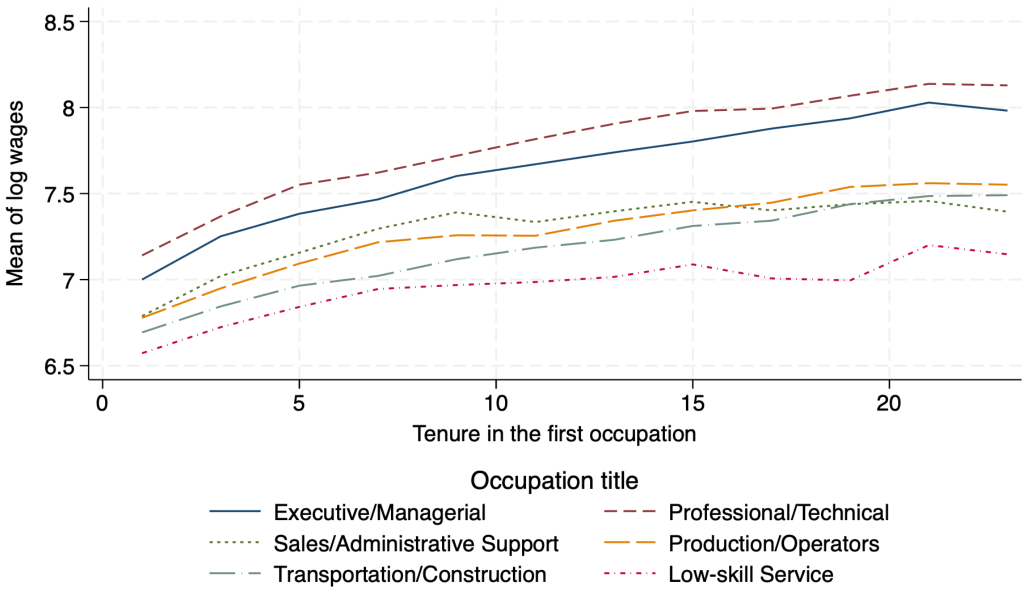
3.2.2 Descriptive statistics

I present the key features of the data. Numerous researchers have documented the properties of earnings and wages over the life cycle (see, for example, Huggett et al. (2006) for the U.S. or Magnac and Roux (2021) for France). However, the analysis in this paper differs in that I focus specifically on the wage dynamics of workers in their initial full-time occupations and their decision to sort themselves out of the occupations. This focus, rather than labor market outcomes over the entire lifecycle, naturally directs the analysis towards the wage dynamics during the initial employment relationship, the duration of its spell, and the reasons behind the decision to quit the initial occupations. In this subsection, I depict the wage dynamics that vary across occupations, which motivates the classification of the more detailed occupations than simply distinguishing white-collar and blue-collar occupations.

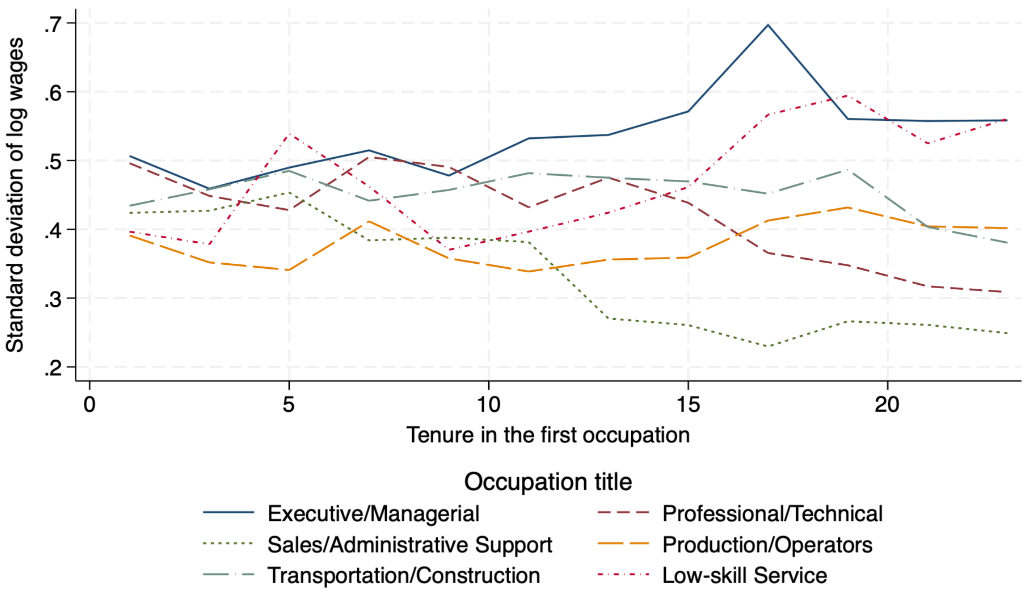
Wage dynamics across different occupations

Figure 3.1: Wage dynamics

(a) Mean of log wages across different occupations



(b) Standard deviation of log wages across different occupations



The upper panel in Figure 3.1 displays the mean log wages over tenure across different occupations. A common pattern observed is that wages increase rapidly during the first 10 years of tenure, reflecting the period of fast skill accumulation. This early-stage wage growth is consistent across all occupational categories and is followed by a flattening pattern after this rapid increase. Despite this general pattern, there are apparent differences in both wage levels and their growth rates across occupations. Executive/Managerial and Professional/Technical occupations exhibit both the highest initial wage levels and the steepest wage growth over time, reflecting greater returns to experience in these high-skill occupations. In contrast, Production/Operators and Transportation/Construction occupations show lower starting wages and more modest wage growth over time, and workers in Low-skill Service occupations experience the lowest overall wages and the slowest wage growth, emphasizing the limited potential for human capital accumulation in this occupation.

The lower panel of Figure 3.1 illustrates the standard deviation of log wages over tenure, offering insights into wage dispersion across occupations. The patterns of wage dispersion differ notably between occupations with no distinctive common feature. Even across the occupations classified equally into white-collar occupations, the wage dispersion within Executive/Managerial occupation increases over tenure, while it decreases in Professional/Technical and Sales/Administrative Support occupations. Among the blue-collar occupations, the wage dispersion remains relatively stable over tenure in Production/Operators and Transportation/Construction, whereas Low-skill Service occupation display more significant fluctuations and slightly increasing patterns. Those different patterns stress the need to adopt the refined occupational categories rather than just the distinction between blue- and white-collar occupations to better understand the wage dynamics.

Two main forces could generate the observed wage dynamics: human capital accumulation and workers' sorting. As human capital accumulates at different rates across occupations, the mean wage increases while yielding different trajectories. At the same time, workers are sorted, which may lower the wage dispersion if workers who decide to stay in their occupations get homogeneous. The increase in dispersion over tenure can still be understood within this framework through the stochastic components potentially embedded in human capital accumulation and wage processes, both of which add heterogeneity to the wages of the workers who remain working in the initial occupations. In the next section, I build the structural model on the two forces that can explain the different patterns of wage dynamics.

Occupational decision: Stay/Quit

Figure 3.2: Hazard rate

Share of workers leaving the occupations conditional on working until the previous period

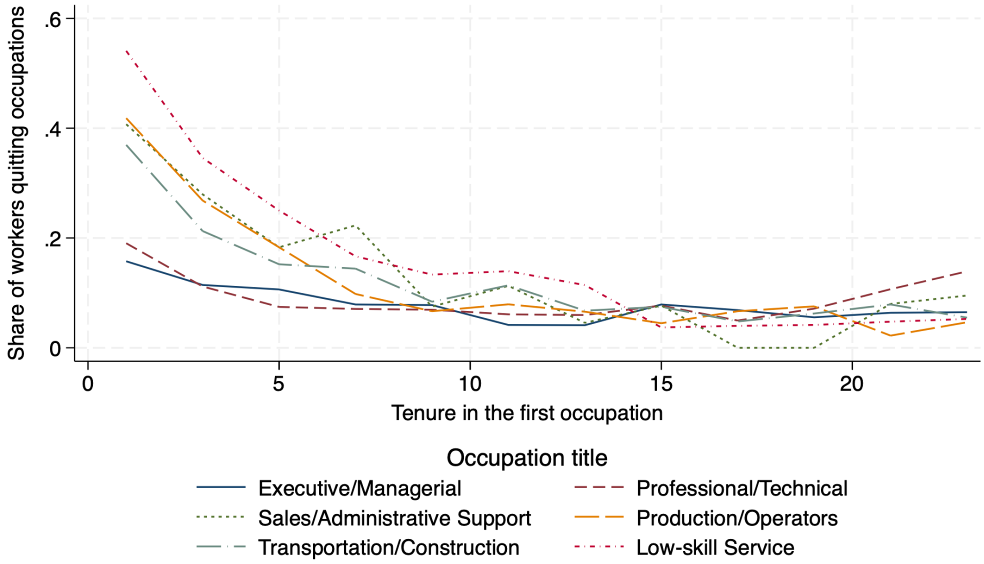


Figure 3.2 illustrates a decline, as tenure increases, in the share of workers who decide to quit among those who remain working up until the tenure in their first occupations. This pattern supports the hypothesis that workers face uncertainty about their skills, which gets resolved gradually through their experience in their early careers. As they gain a better understanding of their skill in the current occupation, the workers who settle into occupations may tend to those who believe themselves productive enough, resulting in decreased occupational mobility over tenure.[5]

[5]Although the decreasing mobility with age or tenure could be evidence of learning (Neal 1999), the feature can be also explained by a search model that does not even account for learning about match-specific productivity.

Then, the higher rate of occupational switching in the early years could be explained by several factors. For instance, the speed of learning could be fast in the initial phase when the workers face great uncertainty about their own skill, which, in turn, helps them to decide whether to quit soon after the start of working in their first occupation. Another explanation could be that workers in their early careers have more attractive outside offers, especially in some occupations that serve as stepping stones. For instance, occupations with lower wage growth, such as Low-skill Service and Sales/Administrative Support, may offer opportunities for workers to develop skills and enhance productivity that can be exploited in the other occupations as well, making them attractive stepping stones.

3.2.3 Suggestive evidence of selection and learning

This paper focuses on the spell of the workers in their initial occupations, which would be the period when uncertainty about their abilities is likely to be most prevalent, so the learning process could play a significant role. Given that the sample consists of an unbalanced panel by construction, I first examine whether the duration of being in the sample (i.e., working in the initially assigned occupations) is endogenously determined. After exploring the selective attrition from the sample, I provide suggestive evidence that learning, particularly through wage observations, could serve as a relevant mechanism driving the non-random sample.

Presence of selective attrition

To investigate the presence of selective attrition, I conduct a series of simple tests applying the variable addition test proposed by Verbeek and Nijman (1992). Specifically, I regress log wages on proxies for human capital (e.g., skill measures, occupational tenure, education) and, importantly, include a dummy variable indicating whether the worker leaves his initial occupation in the next period. A significant coefficient on this attrition dummy implies the presence of selectivity bias, suggesting that the sample could be non-randomly selected through attrition. Table 3.2 presents the estimation results and supports the presence of selective attrition underlying the wage process through the workers' decisions to stay in their initial occupations. More details regarding the practice are described in Appendix B.

Not just rejecting the hypothesis of no selective attrition in the sample, the negative estimates of the attrition dummy indicate that workers who leave the sample tend to have lower wages compared to those who remain. This alludes to the potential mechanism of selective attrition, meaning that workers having lower wages than average or expected are more likely to exit the sample. The estimates controlling for occupation-specific effects through the mean, tenure, and other individual characteristics in different specifications further confirm that this interpretation holds even after accounting for these factors. In the next section, I build a structural model that accounts for the selective attrition mechanism explicitly in conjunction with the wage process.

Table 3.2: Test for Selective Attrition

	(1)	(2)	(3)	(4)
Attrition next period	-0.131***	-0.111***	-0.119***	-0.116***
	(0.0122)	(0.0122)	(0.0122)	(0.0121)
Controls				
Individual characteristics	o	o	o	o
Occupation-specific mean	x	o	o	o
Occupation-specific tenure effect	x	x	o	o
Occupation-specific effects of individual characteristics	x	x	x	o
# of obs	14100	14100	14100	14100
Adjusted R-squared	0.420	0.430	0.434	0.447

Note: Standard errors are in parentheses. *** denotes significance at the 1% level.

Possible selection mechanism: Learning from wage observations

With the pieces of evidence in the previous figures and table where it is found that the wage process could be subject to selective attrition and the attrition can be viewed as an outcome of learning, the following question arises: What is the source of the information that workers use in learning and subsequent decision whether to quit their initial occupations? A likely candidate is the wage, which would be closely tied to a worker's productivity on the job. If workers gain new information from their own realized wages, their occupational decisions may respond to it, which is the hypothesis studied in the following.

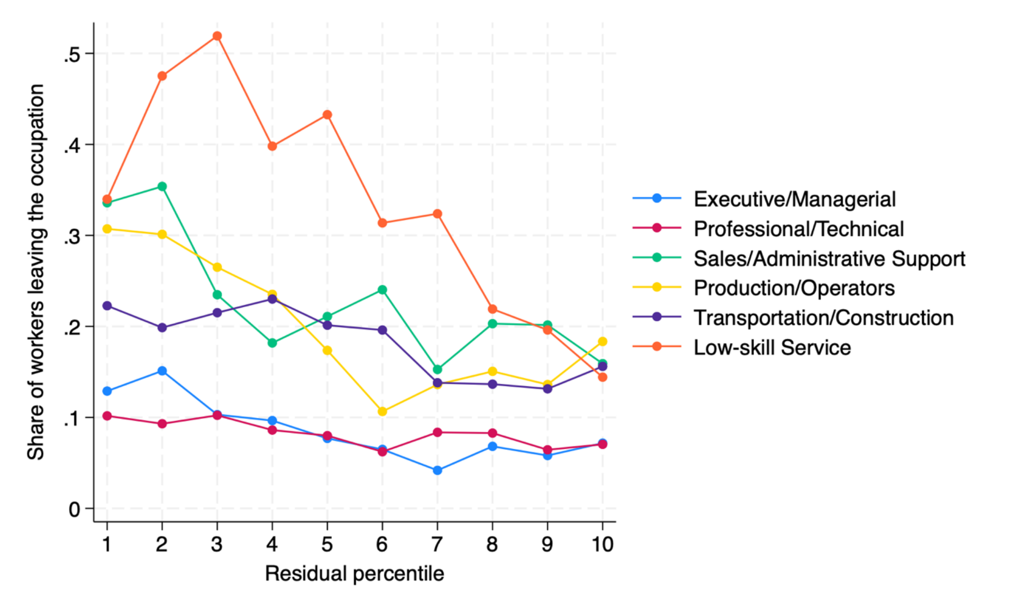
To investigate if wages provide information that influences the workers' decision to quit their occupations, the first task would be to capture the "new information" the workers obtain from the realized wages. Given that the analysis focuses on workers in their initial occupation after the transition from school to the labor market, it could be expected that the private information that workers hold is relatively homogeneous at the outset. I regress log wages on individual characteristics, as done in the specification (4) in Table 3.2, but this time without the dummy variable indicating the attrition next period.[6] Then, the new information would be contained in the wage residuals computed by comparing the actual wage and the predicted wage based on observable characteristics.[7]

Figure 3.3 displays the relationship between the wage residuals and the probability of quitting the occupations in the next period. The figure shows that the fraction of workers quitting their initial occupations is higher for those experiencing lower residuals in percentile which are negative values indeed, and it decreases as they face higher residuals. This trend indicates that workers whose productivity, assumed to be reflected in their wages, is below the expected one based on their observable characteristics are more likely to quit. Conversely, workers with positive residuals—those producing above expectations—are less likely to leave their occupations. The pattern suggests that if wages serve as an accurate reflection of productivity, workers whose performance falls short of expectations are more inclined to leave the occupations for better matches.

[6] Note that individual fixed effects are not included, which is to get the residuals reflecting individual deviations from the general pattern rather than the deviations around the individual mean. individual fixed effects impose that all residual profiles sum to zero for each person, which is an undesirable feature in this practice.

[7] Considering heterogeneous effects over tenure, additional features found in the analysis are, as having higher tenure, (i) the effects of observed initial individual characteristics except for the ASVAB test scores on the log wage fall, and (ii) the occupational decision gets less responsive to the wage residual.

Figure 3.3: Share of workers leaving the occupations in response to the wage shock



3.3 Structural model

Given the descriptive statistics, I build a structural model that enables separation of the potential channels bringing about the observed patterns of occupational choice and wage dynamics. Disentangling the mechanisms is essential to understanding if uncertainty about one's own ability plays a role in the labor market and quantifying its effects in different occupations. I consider a model of employment dynamics that features uncertainty in a learning environment. The model provides intuition for how information about ability is produced and how the learning process works, which offers a perspective on the interaction between beliefs, wages, and occupational sorting. Focusing on the worker's first occupation spell, in which uncertainty could prevail the most, the model aims to investigate whether learning about one's own ability in the occupation actually plays a role in workers' decisions to stay working or quit. The worker's optimization problem is viewed as an optimal stopping decision formulated in a familiar dynamic programming problem where a finite dependence property (Hotz and Miller 1993; Arcidiacono and Miller 2011) holds. After the model is introduced, I discuss the identification and econometric considerations of the model under the assumptions imposed in the structural model.

3.3.1 Discrete time duration model

A basic building block for the analysis is a dynamic discrete choice model that extends conventional discrete-time duration analysis. Individual workers are assumed to start spells in an exogenously given occupation, which abstracts from the initial conditions problem conditional on the observables. With t denoting the number of years since worker i started working in occupation j he is assigned, let $d_{ijt} = 1$ if the worker stays at time t and $d_{ijt} = 0$ if he quits. The exogenously specified initial condition $d_{ij1} = 1$ means that every worker starts with zero experience in the occupations they are assigned. At the end of each period, he receives a wage based on which he updates his belief about his own skill (or called productivity, human capital stock) which is initially determined by one's ability. The skill accumulates based on his occupation through learning by doing. Initially unknown ability, in conjunction with a permanent shock in the skill accumulation, makes the worker not exactly know his skill level, so he forms and revises the belief about his own skill and gains associated with staying in or quitting the occupation. The discrete choice, modeled starting from period t=2, has a terminal period $T_i + 1$ (≥ 2) at the beginning of which the worker decides to quit. Then, the choice $d_{iT_i+1} = 0$ signifies that work i has remained working up to period T_i.

Human capital and Output production

Following the human capital models, worker i's potential productivity in occupation j in period t is directly related to skill level H_ijt. The worker initially assigned to the occupation is uncertain about his skill which he learns through a sequence of noisy signals from the output production.[8] I assume that potential output production may be represented by

$$Y_{ijt} = H_{ijt}\exp\{\sigma_p(j)\varepsilon^p_{it}\} \tag{3.1}$$

where $\varepsilon^p_{it} \sim N(0,1)$ represents transitory variation in the productivity at work. Along with its variability, $\sigma_p(j)$, varying across the occupations, the realized idiosyncratic productivity shock is denoted by ε^p_{ijt}.

[8]Uncertainty in human capital investments could arise from various sources. For example, the future price of different skills in the production function would be another source because the market demand and the production technology progress stochastically over time. This paper focuses only on the uncertainty rooted in one's attribute, specifically ability.

The initial skill and its accumulation via the experience in the occupation is specified as follows:

$$logH_{ij1} = X_i\gamma_j + \kappa_{ij} + A_{ij} \tag{3.2}$$

$$logH_{ijt+1} = (1-\delta_j)logH_{ijt} + \alpha_j + \sigma_h(j)\varepsilon^h_{it+1} \tag{3.3}$$

where X_i depicts one's observed characteristics before entering the labor market with their contributions to the initial skill varying across occupations by γ_j.[9] There is an individual heterogeneity component, κ_{ij}, known to the workers. Given the quality measures regarding the worker's observed skills, I assume $\kappa_{ij} = \eta_i\rho_j$ where η_i is the observed skill endowments vector. A_{ij} reflects, using Heckman et al. (1998) terminology, the ability to "earn" in occupation j and embodies the contribution of the initially unknown endowment to subsequent earnings. It is unknown to the worker and learned throughout the tenure in the corresponding occupation. Log of the skill accumulates by $\alpha_j > 0$ and depreciates by its rate $\delta_j \in (0,1)$, both of which are specific to current occupations but common across the workers within the same occupation. Skill accumulation shock, $\varepsilon^h_{it} \sim N(0,1)$, has individual-specific and persistent effects on workers' productivity through the newly produced skill.

Learning-by-doing type of model in this paper shares similar features with the seminar Ben-Porath model (1967) and its extensions (e.g., Heckman et al. 1998; Magnac et al. 2018), even though in their framework, differently from this paper, workers choose the amount of investment in skill accumulation; worker's productivity is directly proportional to skill level, and each period worker's skill is augmented by the amount of new human capital created along with its depreciation by the amount proportional to the current human capital. The key feature of the Ben-Porath model that skill rises over the life cycle at a diminishing rate can be captured in this learning-by-doing model, on top of the constant depreciation rate, either by human capital depreciation rates varying over age or/and more complicated functional forms of skill accumulation. However, in this paper, which focuses on the early career when wage growth mainly happens, I restrain the model from

[9] To be clear, X_i includes an intercept, which captures the initial human capital common to the workers.

those specifications that are aimed to replicate the concave earnings profile in the latter lifecycle.

Information structure and earnings determination

I assume that workers and employers have symmetric information about workers' ability to earn, A_{ij}, which is not directly observed. At the beginning of the employment relationship, the agents have a prior about A_{ij}, which is consistent with its population distribution: $N(\mu_A(j), \sigma_A(j)^2)$.[10] They know with certainty the parameters of output production and human capital production functions, and workers' wages are determined by a one-period contract that firms offer by specifying it as a function of the output realized in the period. As in Gibbons and Murphy (1992) and more recently in Camargo et al. (2022), a linear contract along with competition among firms (e.g., $W_{ijt} = (1 - b_{ijt})E[Y_{ijt}|I_{it}] + b_{jt}Y_{ijt}$) is a popular specification that allows complex issues (e.g., productivity-enhancing contract, risk sharing) to be studied by yielding systematic differences between wages and output production.[11] However, in the absence of workers' amount of output production or compensation policy in data, it is usually challenging to identify the piece rate. This paper focuses on the specific case of $b_{jt} = 1$, which links observed compensation and unobserved workers' output production.[12]

Workers are assumed to be rewarded completely on the basis of output production. Namely, the piece rate in the linear contract set to one:

$$W_{ijt} = P_{jm(t)}Y_{ijt} \tag{3.4}$$

where $P_{jm(t)}$ is the occupation-specific output price in calendar year m at time t which denotes duration in the first occupation. Its variation over time by aggregate shock is

[10] Ability is occupation-specific, leaving the covariance between the components unrestricted. Assuming one-dimensional ability implies that the best workers are better at all occupations. Even though many empirical articles in the literature (Gibbons et al. 2005; Antonovics and Golan 2012; Groes et al. 2015) pay attention to a one-dimensional model of ability, the empirical results in the literature favor the model of comparative advantage (see, for example, Papageorgiou 2014; Arcidiacono et al. 2025). This paper does not add complexity by focusing on the ability associated with the initial occupation.

[11] The specification nests particular specifications found in empirical studies. Gibbons et al. (2005) assume that wages are determined by the expected output (i.e., $b_{jt} = 0$), and in Arcidiacono et al. (2025), wages depend only on the realized output (i.e., $b_{jt} = 1$).

[12] This is in line with the suggestive evidence that wages contain information about workers' abilities.

assumed to be perfectly foreseen by the workers for simplicity.[13] This yields, by replacing the skill accumulation process recursively, the following equation for log wages:

$$\begin{aligned} w_{ijt} &= p_{jm(t)} + h_{ijt} + \varepsilon^p_{ijt} \\ &= p_{jm(t)} + (1-\delta_j)^{t-1}(X_i\gamma_j + \eta_i\rho_j) + \frac{1-(1-\delta_j)^{t-1}}{1-(1-\delta_j)}\alpha_j + \{(1-\delta_j)^{t-1}A_{ij} + \xi_{ijt}\} \end{aligned} \tag{3.5}$$

where $w_{ijt} \equiv logW_{ijt}$, $p_{jm} \equiv logP_{jm}$, and $h_{ijt} \equiv logH_{ijt}$. The last bracket stands for the ex-ante unpredicted part of the log wage. Ability A_ij underlies the uncertainty in the initial skill that lasts to the subsequent periods while its effect decreases over time. On the other hand, the variation attributed to the skill production shocks $\xi_{ijt} \equiv \sum_{\ell=1}^{t}(1-\delta_j)^{\ell}\varepsilon^h_{ij\ell} + \varepsilon^p_{ij\ell}$ accumulates over time.

Bayesian learning and belief update

Working in a certain occupation provides the accumulation of occupation-specific skill and information to learn how good one is in the occupation. Workers form beliefs about their skill levels and update them based on the wage observed at the end of the period. That is, at the start of period t, given one's prior belief about h_{ijt} characterized $N(\mu^*_{ijt}, \sigma^{2*}_{ijt})$, the worker observes a signal $s_{ijt} = w_{ijt} - p_{jm}(= h_{ijt} + \varepsilon^p_{ijt})$.[14] Since the human capital stock cannot be separated from the idiosyncratic productivity shock in the signal, the worker forms posterior in the Bayesian fashion based on the signal. Considering the skill that is supposed to be accumulated through learning by doing, the belief is updated. Given the normality assumptions, the posterior (or the prior in the next period) is characterized by its mean and variance:

$$\mu^*_{ijt+1} = (1-\delta_j)\Big[\frac{\sigma^2_p(j)}{\sigma^{2*}_{ijt} + \sigma^2_p(j)}\mu^*_{ijt} + \{1 - \frac{\sigma^2_p(j)}{\sigma^{2*}_{ijt} + \sigma^2_p(j)}\}s_{ijt}\Big] + \alpha_j \tag{3.6}$$

$$\sigma^{2*}_{ijt+1} = (1-\delta_j)^2\frac{\sigma^2_p(j)}{\sigma^{2*}_{ijt} + \sigma^2_p(j)}\sigma^{2*}_{ijt} + \sigma^2_h(j) \tag{3.7}$$

[13] The process of the output prices could be modeled using an AR(1) process, which would be aligned with the literature that assumes aggregate shocks in the labor market follow an AR(1) process. For simplicity, I adopt the perfect foresight assumption.

[14] Keeping track on workers' beliefs about skill level rather than about ability itself makes the notation more simple, still yielding the same learning process; for example, the signal about "ability" at time t is $A_{ij} + \frac{\varepsilon_{ijt}}{(1-\delta_j)^t} = w_{ijt} - \{p_{jm(t)} + (1-\delta_j)^{t-1}(X_i\gamma_j + \eta_i\rho_j) + \frac{1-(1-\delta_j)^{t-1}}{1-(1-\delta_j)}\alpha_j\}/(1-\delta_j)^{t-1}$.

with their initial values: $\mu^*_{ij1} = X_i\gamma_j + \rho_j\eta_i + \mu_A(j)$ and $\sigma^{2*}_{ij1} = \sigma^2_A(j)$.

Note that the magnitude of the noise in the signal, $\sigma^2_p(j)$, is constant over time whilst it varies across the occupations. Intuitively, workers in different occupations learn about their skill levels at different speeds of learning. Namely, new information about own ability in the signal is more reliable when the noise is small (i.e., the signal-to-noise ratio is high), so less weight on the prior and more weight on the informative signal are placed. Therefore, learning is not independent of occupation j to which the worker is initially assigned. Note that the precision of the posterior beliefs does not necessarily monotonically decrease over tenure due to the shock to skill production. Instead, the variance converges to a value determined by the variability of the idiosyncratic shocks: $\sigma^2_p(j)$ and $\sigma^2_h(j)$.

Preference and dynamic programming

Let the vector of state variables be denoted by Ω_{it}, where $\Omega_{it} \equiv (\mu^*_{ijt}, \sigma^*_{ijt}, m(t))$ contains the variables characterizing the belief about h_{it} and the calendar year. At the beginning of each period, the worker faces an outside option whose mean value is $v_0(\Omega_{it})$ but with an idiosyncratic component $\varepsilon^u_{0,ijt}$. He decides, maximizing lifetime expected utility, to stay working if the value of continued employment is greater than the outside option and quits otherwise. The worker's per-period utility is separable in the expected log wage $\mathbb{E}_{h,\varepsilon^p}[w_{ijt}(h_{ijt}, \varepsilon^p_{ijt})]$ with the productivity shock ε^p_{ijt} being realized at the end of the period while the idiosyncratic utility shock $\varepsilon^u_{1,ijt}$ is observed before making the choice. I further assume that $\varepsilon^u_{ijt} = (\varepsilon^u_{0,ijt}, \varepsilon^u_{1,ijt})$ is distributed i.i.d. Type 1 Extreme Value with mean zero, and variance equal to $(\pi/\sqrt{6})\sigma_u(j)$[15] This shock, interpreted as a shock at the worker level or idiosyncratic worker characteristics, captures search frictions as they are treated in the form of exogenous shocks in this model.

While the choice to stay working yields the current wage gain and the continuation value associated with the optimal decision next period based on the posterior, the choice process ends when the worker chooses to quit. This allows the worker's problem to be framed by an optimal stopping problem, which can be written in the form of a Bellman equation

[15]This is equivalent to normalizing the location parameter to $-\sigma_u(j)$ multiplied by the Euler's constant, where $\sigma_u(j)$ is the scale parameter.

with the value function:

$$V_t(\Omega_{it}, \varepsilon^u_{ijt}) = \max\{v_{0t}(\Omega_{it}) + \varepsilon^u_{0,ijt}, \mathbb{E}_{h,\varepsilon^p}[w_{ijt}(h_{ijt}, m(t), \varepsilon^p_{ijt})] + \varepsilon^u_{1,ijt}$$
$$+ \beta \int_{\varepsilon^u} \int_{\varepsilon^h} \int_{\varepsilon^p} V_{t+1}(\Omega_{it+1}, \varepsilon^u_{ijt+1}) dF(\Omega_{it+1}|\Omega_{it}, \varepsilon^p_{ijt}, \varepsilon^h_{ijt+1}) dF(\varepsilon^p_{ijt}) dF(\varepsilon^h_{ijt+1}) dF(\varepsilon^u_{ijt+1})\} \tag{3.8}$$

Define $v_{1t}(\Omega_{it})$ as the value function summarizing the lifetime expected utility the worker would expect to receive from choosing a stay, net of the idiosyncratic component of the current period. Then, his optimization problem is viewed to choose an alternative by comparing $v_{0t}(\Omega_{it}) + \varepsilon^u_{0,ijt}$ and $v_{1t}(\Omega_{it}) + \varepsilon^u_{1,ijt}$

3.3.2 Identification

The parameters associated with each occupation are identified separately within the workers having different series of choices and wage outcomes in the same occupation. Therefore, for ease of notation, I abstract from the j subscript denoting the occupation from here on. I state the following standard normalization:
(a) Population distribution of A_i follows $N(\mu_A, \sigma^2_A)$ where μ_A is normalized to 0
(b) A_i is independent of X_i and η_i
(c) Idiosyncratic shocks are independent of each other and over time
(d) Discount factor $\beta = 0.95$
Additionally, since skill level is an unobserved object that does not have natural measures, observed high wages can be rationalized by either a large human capital stock or high returns to human capital. I normalize the wage return to human capital stock to one, which is already imposed in the model. Due to scale indeterminacy inherent in the utility specification, the scale is normalized by setting the coefficient of the log wage to one. Given the normalization and the parametric assumptions in the structural model, this subsection is devoted to explaining the crucial identification assumption and the source of variations to discuss the identification of the distributions of ability, skill accumulation shock, and productivity shock.

The identification problem is, from the selected population distribution of choices and wages ($f_{(w_1,...,w_T),(d_2,...,d_{T+1})}$, where $T = \max_i\{T_1, ..., T_i\}$)) to recover the conditional distributions of the choice probabilities ($f_{d_t|A}$) and potential wages ($f_{w_t|A}$). Identification of the model in this paper is based on Bunting et al. (2024), who provide identification results for the distribution of potential outcomes in a general class of learning models that accounts for the presence of selection issues through learning. The key idea is to characterize the relationship using Bayes' rule:

$$f_{w_1,...,w_T} = f_{w_1,...,w_T|d_2,...,d_{T+1}} \frac{f_{d_2,...,d_{T+1}}}{f_{d_2,...,d_{T+1}|w_1,...,w_T}} \tag{3.9}$$

where the conditional density $f_{w_1,...,w_T|d_2,...,d_{T+1}}$ is directly identified from the data and weighted by a selection adjustment term to recover $f_{w_1,...,w_T}$. Once the selection weights are identified relying on the learning framework, the parameters determining the potential wage distribution, including the distributions of ability and the other unobservables, can be identified.[16]

The identification strategy in Bunting et al. (2024) can be extended to the case of attrition, where no outcome is observed in the sample once an individual chooses to quit and fall into the absorbing state.[17] Additionally, I assumed that each worker with initially unknown ability does not possess persistent private information that affects their decision, which implies that the choice depends on beliefs about ability updated only through previously observed wages and covariates. Namely, the learning model can be seen as a model of selection on observables, where the conditional choice probability function does not depend on any latent variable and is thus identified directly from the data. Then, the wage parameters and the ability distribution can be identified, which allows standard identification arguments from the dynamic discrete choice literature (see, e.g., Hotz and

[16] The residual denoted by ξ_{it} in equation (5) is a sum of the skill accumulation shocks and idiosyncratic productivity shock, of which variance varies over tenure. Under the assumptions imposing normality, independence between them, and identical distribution over time, the distributions of the shocks can be separately identified once the variance of the residual distributions are identified in different tenures.

[17] For example, Assumption L4 in their paper needs to be relaxed when the selection issues are posed by attrition.

Miller 1993; Magnac and Thesmar 2002) to be applicable so that the rest of the primitives on which conditional value functions of the choice model build can be identified.[18]

Lastly, I add the intuition that the panel dimension of the choice process provides additional information to disentangle the variances of the two unobservable in the wage process: the skill accumulation shock and idiosyncratic productivity shock. Without the skill accumulation shock, the variance of workers' beliefs about their own skills converges to zero. This means that any occupational choice after some periods is barely attributed to learning and responses to the realized wages. Hence, if the skill accumulation shock is significant, the variations of occupational choices in the latter periods in the panel can help the variance of the skill accumulation shock to be identified. The choices in response to wage observations enable the variance of the productivity shock to be identified. For example, consider worker A and worker B, whose skills and beliefs are the same, and there are two signals, say, a good signal and a bad signal, for simplicity. Worker A receive receives a good signal first, and a bad signal next, and worker B receives it in the other order around. Abstracting from the duration dependence through the different outside option values in the two periods, the probability of quitting after having the information through the signals would not be equal in the learning environment. Hence, when they learn about their ability, one would be more likely to quit than the other, which provides information about the weight placed on the prior which is a function of the variance of the idiosyncratic productivity shock.

3.3.3 Econometric Issues

Selective attrition and potential bias

Central to the structural model in the paper is that, in the presence of learning about ability on the job, luck (i.e., realizations of ε^P and ε^h) leads to different career trajectories. That is, the duration of working in the occupation provides information about the realized luck in the preceding periods. Consequently, the decision to stay up to a certain period is not random and is dependent on past skill accumulation and productivity shocks through

[18]Note that, in conjunction with the distributional assumptions and conventional Bayesian learning, the tractable sequence of beliefs characterized by only two parameters makes the model possess a first-order Markov structure.

the learning process, even under independence between the unobservables in the wage and choice equations.[19] Estimating wage parameters raises the econometric issue of correcting the selection bias resulting from selective attrition. The approach in this paper is to estimate the wage equation jointly with the explicit selection process specified. The usual panel data approaches, both fixed effects and random effects models, applied to the data generated by such a model yield biased estimates of the wage parameters. To make the specification comparable to the usual Mincer wage equation, let's first assume no skill depreciation $\delta = 0$ and no skill accumulation shock $\varepsilon_{it}^{h} = 0$ (i.e., deterministic human capital accumulation). Once I define the deviation of the observed vector of time-varying variable x_{it} from its individual mean by $\tilde{x}_{it} \equiv x_{it} - \sum_{\ell=1}^{T_i} x_{i\ell}/T_i$, and analogously for the other variables, it is straightforward to show that the condition for consistency of the fixed effects estimator (i.e., within estimator), $\hat{\theta}_{FE} = \sum_i \sum_{\ell}^{T_i} [\tilde{x}'_{i\ell}\tilde{x}_{i\ell}d_{i\ell}]^{-1}[\tilde{x}'_{i\ell}\tilde{w}_{i\ell}d_{i\ell}]$,is

$$\mathbb{E}[\tilde{\varepsilon}_{it}^{p}|d_{i1} = \cdots = d_{iT_i} = 1, d_{iT_i+1} = 0] = 0. \tag{3.10}$$

The sufficient condition would not generally hold because d_{it} is a function of belief which is determined by $\{\varepsilon_{il}^{p}\}_{l=1}^{t-1}$ and $\{\varepsilon_{il}^{h}\}_{l=2}^{t-1}$ through the dynamics of learning.[20] Intuitively, the condition for staying for T_i periods is not independent of the signals so the realized shocks; For example, suppose that a worker who believes himself to have high productivity as a result of positive signals he has received is more likely to stay in his initial occupation (i.e., $v_{1t}(\Omega_{it}) - v_{0t}(\Omega_{it})$ increases in μ_{ijt} for all t). Then, the worker staying for T_i period may not have experienced too negative wage signal in the periods before T_i. Namely, the condition for staying in up until T_i would imply a truncation of the unconditional distribution of $\{\varepsilon_{il}^{p}\}_{l=1}^{T_i-1}$. Therefore, selective attrition due to the learning process would yield a bias in the within estimator.[21] More discussion and a simulation result is found

[19] Nonrandom selection and attrition problems coming from a nonzero correlation between the unobservable in the wage and the one in utility is well understood in the literature. In this paper, I focus on the bias generated by learning about ability while maintaining the assumption that the idiosyncratic productivity shocks, skill accumulation shocks, and utility shocks are independent.

[20] A case in which the consistency condition holds with the sufficient condition being not necessarily met is the situation where observations are missing deterministically given the explanatory variables.

[21] The instrument proposed in Gibbons et al. (2005) exploits the property of a learning model that Bayesian beliefs are a martingale. However, the IV estimator could still fail to correct the attrition bias. To see this, rewrite the current wage using the previous wage linked by the human capital production. It leads unobservables in the equation to be a composite of skill accumulation and idiosyncratic productivity shocks:

$$w_{it} = p_{m(t)} - (1-\delta)p_{m(t-1)} + (1-\delta)w_{it-1} + \alpha + [\varepsilon_{it}^{h} + \{\varepsilon_{it}^{p} + (1-\delta)\varepsilon_{it-1}^{p}\}]. \tag{3.11}$$

in Appendix D. Since the condition for consistency of the random effects estimator is stronger than the one for the within estimator, I do not discuss the bias in the random effects estimator here.

Selection at entry and unobserved heterogeneity

Even though econometricians can control for some individual characteristics (e.g., years of education before entering the labor market) usually observed in data, there could be initial unobserved heterogeneity that is known to the economic agents (i.e., workers and potential employers) but unknown to the econometricians. Namely, it could be unlikely that initial human capital stock is exogenous conditional on the observed variables at the time of labor force participation. The examples include that initially observed schooling is an outcome of a stochastic process containing measurement error or that observed proxies for the academic/cognitive abilities may omit to capture the productivity in the workplace, in which cases the exogeneity conditioning only on observable would be problematic.

This problem with selection at entry can be addressed by allowing for unobserved heterogeneity types in the spirit of Heckman and Singer (1984) and Keane and Wolpin (1997). Specifically, workers are defined as one of K types with type-specific components that capture permanent characteristics. The finite mixture approach assumes that econometricians know that there are K types. The type-specific unobserved heterogeneity known to the agents and the unobserved abilities initially unknown to the agents are independent. Then, the fact that discrete choice depends on one's known type as well as beliefs about unknown skill levels prevents the distribution of the factors from being easily identified. However, exploiting a main virtue of the NLSY79 that contains auxiliary selection-free measurements of various skills, along with the panel dimension of the data, will allow the mixture model to be identified.

Specificity of skills and value of quitting

The suggested instruments of the lagged wages, $\{w_{i\ell}\}_{\ell=1,2,\cdots,t-1}$, are correlated with w_{it-1} but uncorrelated with the unobservable component in the equation. However, the IVs are still susceptible to selective attrition; the condition to have valid IVs is to stay longer than three periods, which yields the restricted estimation sample conditioning on the observability. This implies that the moment conditions for the IVs could still result in biased estimates.

Skill, determined by initially unknown ability and stochastic skill accumulation process, could be general in the sense that productive workers in one occupation could be more likely to be productive in some other occupations. This can potentially have the value of an outside option dependent on either one's skill or ability in the current occupation.[22] In this case, beliefs about one's own skill in the initial occupation affect not only the value of staying in but also the value of quitting. Exclusion restrictions may need to be imposed to identify, without relying on or further invoking parametric assumptions, the value of the outside option separately from the effect associated with the variables in the expected wage.

One of the possible restrictions would be occupation-specific regressors that generate variations only in the utility of outside options; for example, occupation-specific demand for labor. This is a common strategy found, for example, in Heckman and Sedlacek (1985) and d'Haultfoeuille and Maurel (2013). More specifically in the present setting, occupation-specific vacancies in occupation j' $(\neq j)$ are excluded from the wage equation for occupation j, and the vacancies in j' are uncorrelated with unobserved components of wages in j conditional on vacancies in j.[23] For simplicity, throughout the paper, I simplify the model by assuming that skills are occupation-specific and uncorrelated to each other, which rationalizes the specification that the value of quit is independent of one's own skill in the current occupation.

3.4 Estimation

In this section, I detail the maximum likelihood estimation procedure for the dynamic discrete choice model, which incorporates learning about one's own skill based on observed wages along with its stochastic evolution. Given unobserved state variables of skills, the likelihood is computed using the Kalman filter that recursively estimates the unobserved skill from observed signals of wages. It is worth noting that, given a set of parameter values, the estimate of the skill each period indeed corresponds to the belief formed by

[22] For simplicity, I assume that the value of quitting is deterministic and independent of one's belief about own skill, which plays the role of an exclusion restriction.

[23] Once general equilibrium effects through labor supply have an effect within a period, the condition could be violated, but the feedback effect may have a time lag.

the workers. Then, this algorithm allows me to calculate the conditional distribution of skills, wages, and occupational choices sequentially from the first period to the last period. When the likelihood contribution of the choices is computed, I use the CCP estimation exploiting the terminal state nature. Parameter estimates and the model fit are discussed in the following subsections.

The idea of using the Kalman filter is that the model can be framed within the context of a state-space model with hidden state variables and observable measurements. Additionally, it is a dynamic linear model in which the system evolves over time according to linear stochastic equations, with all latent variables having normal distributions. Optimal estimates of the hidden states $\{h_{it}\}$ based on noisy observations $\{w_{it}\}$ are provided by the recursive algorithm that consists of two main phases: Prediction step and Update phases. Prediction phase is to predict the state at the start of each period before making a decision and observing a new measurement:

$$\begin{aligned}\text{State prediction}: \hat{h}_{it} &= E[h_{it}|I_{it-1}] \\ &= (1-\delta)\hat{h}_{it|t-1} + \alpha \\ \text{Prediction uncertainty}: s_{it} &= Var[h_{it}|I_{it-1}] \\ &= (1-\delta)^2 s_{it|t-1} + \sigma_h^2\end{aligned}$$

where $I_{it-1} = \{X_i, d_{i1} = \cdots = d_{it-1} = 1, w_{i1}, \ldots, w_{it-1}\}$ summarizes all the information up to period $t-1$. Update phase incorporates the information in the measurement based on the current priori prediction and the wage observation, which refines the estimate of the skill level:

$$\text{State update}: \hat{h}_{it|t-1} = \hat{h}_{it-1} + K_{it}\nu_{it}$$
$$\text{Update uncertainty}: s_{it|t-1} = (1 - K_{it})s_{it-1}$$

where $\nu_{it} = w_{it} - \{p_{m(t)} + \hat{h}_{it-1}\}$ is innovation (or prediction error) coming from the realization of measurement equation. The Kalman gain, $K_{it} = s_{it-1}/(s_{it-1} + \sigma_p^2)$, determines the weight given to the new measurement relative to the predicted state.

3.4.1 Maximum Likelihood Estimation

The likelihood function for the model involves the joint probability of the observed wages and occupational decisions, considering the unobserved skills. The Kalman filter facilitates the computation of the likelihood by providing the necessary components to evaluate the probability of the observed wage and the probability of the observed choice made on the belief about own skill. Note that the choice is assumed to depend on the workers' beliefs about their skills updated through the observed sequence of wages, implying that, conditional on the model parameters, each worker's belief matches the estimate of the skill updated through the Kalman filter. Suppose the likelihood contribution of individual i who makes the choice to stay working until period T_i and quit in period $T_i + 1$. Given the sequence of T_i number of binary decisions and T_i number of wage outcomes, his contributions to the likelihood can be written by:[24]

$$\begin{aligned} L_i(d_{i2} = \cdots = d_{it-1} = 1, w_{i1}, \ldots, w_{it-1} | X_i, d_{i1} = 1) \\ = L_i(w_{i1}|X_i)\, L_i(d_{i2} = 1|I_{i1}) L_i(w_{i2}|I_{i1}, d_{i2} = 1) \\ \cdots L_i(d_{iT_i} = 1|I_{iT_i-1}) L_i(w_{iT_i}|I_{iT_i-1}, d_{iT_i} = 1)\, L_i(d_{iT_i+1} = 0|I_{iT_i}) \end{aligned} \tag{3.12}$$

Likelihood of the wages

Kalman filter facilitates the construction of the likelihood function of worker i by calculating the conditional distribution of human capital and wages sequentially from the first period $t = 1$ to the last period $t = T_i$. I perform the prediction and update steps recursively. where the log wage at t is a linear function of normal random variables given information up to $t - 1$. Thus, the likelihood of observing log wage w_{it} is given by the normal density:

$$L_i(w_{it}|I_{it-1}, d_{it} = 1) = (2\pi s_{it-1})^{-1/2} exp\Big[- \frac{\{w_{it} - (p_{m(t)} + \hat{h}_{it-1})\}^2}{2s^2_{it-1}} \Big] \tag{3.13}$$

Likelihood of the choices

[24] In period 1, all workers are assigned to their initial occupations, which is not modeled in this paper, so the discrete choices of either staying or quitting from the second period on contribute to the likelihood.

The worker's expected values associated with staying in or quitting the current occupation are constructed by exploiting the insights from Hotz and Miller (1993) and, more generally, Arcidiacono and Miller (2011). Taking advantage of the terminal state nature of the dynamic discrete choice problem allows the future value conditional on staying in the current period to be expressed as follows:

$$\begin{aligned} v_{1t}(\hat{h}_{it}, s_{it}) &= \mathbb{E}_{h,\varepsilon^p}[w_{it}(\hat{h}_{it}, \varepsilon^p_{it})] + \beta \int_{\varepsilon^h}\int_{\varepsilon^p} \sigma_u \log\Big[\frac{exp\{v_{0t+1}\}}{\sigma_u} + \frac{exp\{v_{1t+1}(\hat{h}_{it+1}, s_{it+1})\}}{\sigma_u}\Big] \\ &\qquad dF(\hat{h}_{it+1}, s_{it+1}|\hat{h}_{it}, s_{it}, \varepsilon^p_{it}, \varepsilon^h_{it+1})dF(\varepsilon^p_{it})dF(\varepsilon^h_{it+1}) \\ &= p_{m(t)} + \hat{h}_{it} + \beta\Big[v_{ot+1} + \int_{\varepsilon^h}\int_{\varepsilon^p}\{-\sigma_u \log P_{0t+1}(\hat{h}_{it+1}, s_{it+1})\} \\ &\qquad dF(\hat{h}_{it+1}, s_{it+1}|\hat{h}_{it}, s_{it}, \varepsilon^p_{it}, \varepsilon^h_{it+1})dF(\varepsilon^p_{it})dF(\varepsilon^h_{it+1})\Big]. \end{aligned}$$

where the state variables other than those characterizing the beliefs are abstracted away in the expression. In the first equality, the ex-ante value of making the optimal decision next period is expressed as the sum of the conditional values associated with each choice using the property of the i.i.d. type 1 extreme value utility shocks. $P_{0,t+1}(\hat{h}_{it+1}, s_{it+1}) \equiv Prob(d_{it+1} = 0|\hat{h}_{it+1}, s_{it+1})$ in the second equality denotes the probability to make the quitting choice next period, conditional on the next period's belief that is determined by the realization of ε^p_{it} and ε^h_{it}. The ex-ante value is expressed by the conditional value from quitting and a function of conditional choice probabilities (CCPs). The latter can be interpreted as a non-negative adjustment term that adjusts for the possibility that the quitting decision could not be the optimal choice. This interpretation can be illustrated by the relation that as the probability of quitting goes to zero, the adjustment term grows infinitely.

The difference in value functions, $v_{1t} - v_{0t}$, matters for the observed choice probability: $P_{1t} = \exp(v_{1t} - v_{0t})/\{1 + \exp(v_{1t} - v_{0t})\}$. By substituting v_{1t} with the expression above, the probability can be expressed only by the utility parameters and the one-period ahead choice probability. This CCP estimation method provides a way to estimate the utility parameters not solving the dynamic programming problem, which consists of the following two steps. The first step involves estimating and predicting the CCPs. The second step then takes the first-step CCPs as data and estimates the remaining structural parameters in the multinomial logit with an offset term. The CCP estimation that estimates a model in stages does not affect the consistency of the estimates, but does reduce efficiency.[25] Then, the likelihood of observing choice d_{it} is given by

$$L_i(d_{it}|I_{it-1}) = \left[\frac{\exp\{(v_{1t} - v_{0t})/\sigma_u\}}{1 + \exp\{(v_{1t} - v_{0t})/\sigma_u\}}\right]^{d_{it}} \left[1 - \frac{\exp\{(v_{1t} - v_{0t})/\sigma_u\}}{1 + \exp\{(v_{1t} - v_{0t})/\sigma_u\}}\right]^{1-d_{it}} \tag{3.14}$$

where

$$v_{1t} - v_{0t} = p_{m(t)} + \hat{h}_{it} + \beta v_{ot+1} - v_{ot} + \beta \int_{\varepsilon^h} \int_{\varepsilon^p} \{ -\sigma_u \log P_{0t+1}(\hat{h}_{it+1}, s_{it+1})\} \\ dF(\hat{h}_{it+1}, s_{it+1}|\hat{h}_{it}, s_{it}, \varepsilon^p_{it}, \varepsilon^h_{it+1}) dF(\varepsilon^p_{it}) dF(\varepsilon^h_{it+1}).$$

3.4.2 Estimation Results

In this subsection, I examine the estimation results in Table 3.3. The distribution of abilities across occupations reflects the initial beliefs of workers as they enter the labor market. The findings indicate that the standard deviations of ability distributions (σ_A) are statistically significant across all occupations, confirming that workers face uncertainty regarding their abilities at the onset of their careers. An increase in ability by one standard deviation leads to a wage increase of approximately 38-40%, except in Professional/Technical and Low-skilled Service occupations. In the Professional/Technical occupation, abilities are initially more dispersed, resulting in an approximate 55% wage increase with each standard deviation rise in ability. It is noteworthy that the standard deviation in the ability component is higher than that in the productivity shock (σ_p), suggesting that unexplained wage variation is initially more explained by ability than by idiosyncratic productivity shocks. However, comparing the relative importance gets complicated, as the effect of ability on wages is compounded by skill shocks impacting newly accumulated human capital and workers' sorting based on beliefs.[26]

[25]To calculate the valid standard errors that account for the multiple stage estimation procedure, a bootstrap procedure can be used. CCPs in the first stage can be estimated in any flexible way (e.g., neural network) as long as it provides consistent estimates.

[26]The magnitude of the idiosyncratic productivity shock also reflects the degree of noise, and its relative magnitude with respect to the uncertainty in skill determines the speed of learning. However, in the presence of skill accumulation shocks that add uncertainty and wage dispersion., the small variability of productivity shock does not necessarily guarantee a faster learning environment.

The growth of log wages is determined by the rate of skill accumulation (α) and its depreciation (δ). In other words, wage growth can be substantial due either to significant gains from experience and/or minimal losses over tenure. Two high-skill occupations, Executive/Managerial and Professional/Technical, exhibit distinct mechanisms in this respect. In Executive/Managerial occupation, skill accumulation occurs at a relatively low rate but experiences less depreciation, making initial human capital more critical in determining future wage potential. In Professional/Technical occupation where the variance of skill accumulation shock (σ_h^2) is large, on top of the high wage growth, the risk-neutral workers are induced to stay because of the expectation that a large skill accumulation shock could be realized in the future. This incentive is larger for younger workers since skill accumulation shock permanently affects their skill levels.

The last row indicates the scale of the utility shock, modeled as a type I extreme value distribution. A high utility shock value would suggest that worker sorting could occur significantly through factors outside the scope of this model. Specifically, sorting through learning is less prevalent in blue-collar occupations, such as Production/Operators and Transportation/Construction. This implies that other labor theories, such as job search and the stepping-stone model, may be relevant to these occupations. The estimates for other wage parameters align with prior literature. For instance, even where some estimates are not statistically significant, math skills tend to be more highly rewarded in high-skill occupations like Executive/Managerial and Professional/Technical, whereas verbal skills are more advantageous in low-skill occupations.

3.4.3 Model Fit

The point estimates show the existence and relevance of the learning process with occupational decision. Before delving into more quantitative practice using the structural estimates, I check for the model fit. Using the estimated parameters of the model in the previous subsection, I simulate a data set from the explanatory variables in the original data and repeat the simulation ten times. In each simulated data set, individuals enter at the calendar time of their actual entry in the occupation, and their true ability and idiosyncratic shocks are drawn from the corresponding distributions characterized by the

Table 3.3: Parameter Estimates by Occupation

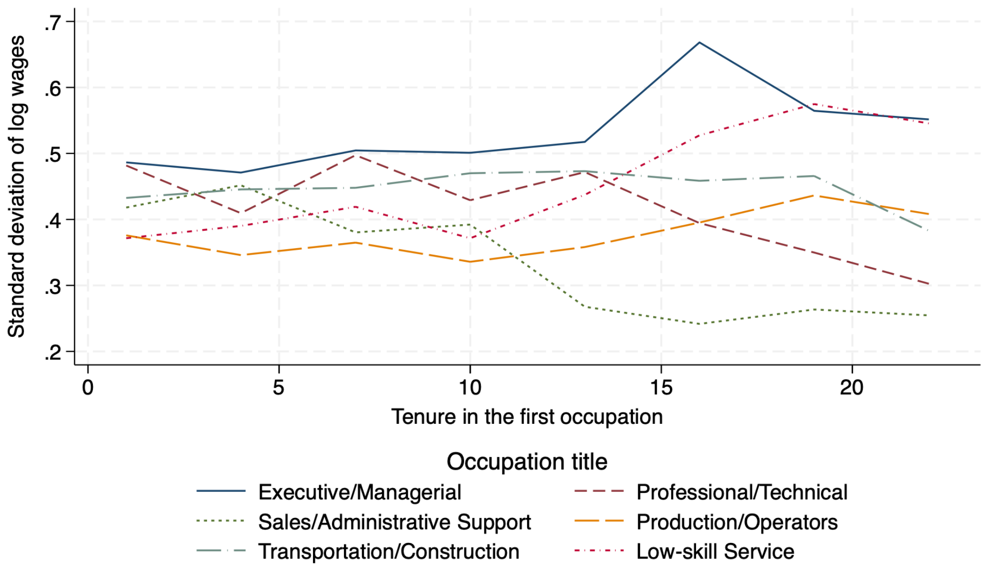
	Executive/ Managerial	Professional/ Technical	Sales/ Admin Support	Production/ Operators	Transportation/ Construction	Low-skill Service
Ability distribution						
σ_A, SD of ability	0.3984	0.5480	0.3923	0.3868	0.3806	0.3366
	(0.0357)	(0.0474)	(0.0280)	(0.0234)	(0.0146)	(0.0198)
Human capital production and productivity shock						
α, New human capital production	0.4207	1.5266	0.4584	0.4083	0.5091	0.0231
	(0.0474)	(0.0730)	(0.0465)	(0.0331)	(0.0213)	(0.0121)
δ, Depreciation of log human capital	0.0486	0.1945	0.0571	0.0525	0.0672	0.0210
	(0.0373)	(0.0514)	(0.0435)	(0.0385)	(0.0241)	(0.0000)
σ_h, SD of Skill shock	0.1800	0.2997	0.1532	0.0939	0.1833	0.1032
	(0.0000)	(0.0391)	(0.0000)	(0.0432)	(0.0400)	(0.0395)
σ_p, SD of Productivity shock	0.2680	0.1802	0.3145	0.1986	0.2174	0.2661
	(0.0000)	(0.0593)	(0.2558)	(0.0983)	(0.0588)	(0.0493)
Initial human capital						
Constant	6.7944	6.8899	6.7127	6.7794	6.6619	6.5443
	(0.0000)	(0.0001)	(0.0000)	(0.0041)	(0.0000)	(0.0017)
Observed Verbal Skill	-0.0333	0.0398	0.0207	0.0491	0.0613	0.0471
	(0.0262)	(0.0516)	(0.0310)	(0.0766)	(0.0419)	(0.0577)
Observed Math Skill	0.2694	0.1355	0.1137	0.0983	-0.0005	0.0321
	(0.0644)	(0.0728)	(0.0893)	(0.0253)	(0.1475)	(0.0338)
Value of quitting (outside option)						
Constant	29.9826	29.9722	29.9556	29.9688	29.9638	29.8170
	(0.1351)	(0.1674)	(0.1018)	(0.5746)	(0.2520)	(0.1709)
Observed Verbal Skill	1.7521	-1.3253	-0.3935	1.9502	-0.0418	-2.0419
	(0.0000)	(0.3969)	(0.4368)	(0.4303)	(0.2963)	(Inf)
Observed Math Skill	-0.2384	0.3129	-0.0305	-0.7557	-0.4844	0.9639
	(0.0255)	(0.1422)	(0.0389)	(0.0299)	(0.0284)	(Inf)
Tenure (duration dependence)	0.6533	-2.2945	0.8804	1.7653	2.0439	2.5653
	(0.0534)	(0.0424)	(0.1721)	(0.1019)	(0.0488)	(0.0905)
σ_u, Scale of utility shock	0.2905	0.1523	0.2469	0.3372	0.3428	0.3011
	(0.0413)	(0.0460)	(0.0603)	(0.0468)	(0.0302)	(0.0281)

* Standard errors in parentheses.

estimated model parameters. Then, the decision to stay in or quit the first occupation is endogenously determined along with the simulated wages.

Figure 3.4 presents the standard deviation of the log wages in each occupation. Note that the moments associated with the choice probability and the mean of log wages are matched in the maximum likelihood procedure. I simulate and draw the figure to see whether the estimated structural model can generate the observed pattern that is not directly taken into account in the estimation. Figure 3.4 illustrates that, in terms of wage dispersion, the estimated model can replicate the patterns of trajectories over tenure in different occupations presented in Figure 3.1 (b). With a ten times larger number of workers in the simulated data, the trajectories are smoother than the observed one in each occupation, while the general trend fits well.

Figure 3.4: Model fit: Hazard rates



3.5 Counterfactual Analysis

In this section, I take advantage of the structural estimates to simulate workers' true ability, skill accumulation, and beliefs, which are not observed in the data. Then, it is available to define and measure mismatch to study the implications of learning on workers' sorting. Next, I consider a counterfactual policy that removes uncertainty as soon as they start the employment relationship. The comparison between data and counterfactuals enables the effect of learning to be studied, accounting for its dynamic effect on the duration of the spell working in the initial occupations and wage dynamics through sorting as well as human capital accumulation. In order to do the practice, additional assumptions are needed to fully solve the model and simulate data. I assume that all workers quit their initial occupation at most one year after the maximum tenure observed in the data, and the terminal period is known with certainty. The corresponding terminal values associated with staying and quitting are parametrically extrapolated. Then, I solve the model backward from the terminal period and get the solution (i.e., policy function). Given the initial characteristics and the calendar year of workers' entry into the occupations in the original data, I simulate data sets by drawing workers' true ability and idiosyncratic shocks from the corresponding distributions characterized by the parameter estimates.

Then, the duration of the spell in the initial occupations through the choice of either staying or quitting, along with wage outcomes, is endogenously determined.

3.5.1 Types of Mismatch

The decision to quit is based not on the true skill but on the belief that has been updated through the series of wage observations, so inevitably pure luck (i.e., realization of idiosyncratic shocks) plays a role. In this learning environment, there are two types of mistakes that the workers could make, which are defined here as mismatch. The first kind of mismatch is to stay because of the belief, even though the actual skill justifies a decision to quit. Namely, this type I mismatch is to decide to stay due to beliefs when the actual skill favors separation. The second kind, type II mismatch, occurs when the current occupation is mistakenly discarded: a false positive. Within the structural model defined in Section 3, the probabilities of the mismatch can be defined as follows:[27]

Type I mismatch (false rejection of the outside option) :

$$P(v_{0t} + \varepsilon^u_{0,it} < v_{1t}(\mu^*_{it}, \sigma^{2*}_{it}) + \varepsilon^u_{1,it} \mid v_{0t} + \varepsilon^u_{0,it} > v_{1t}(h_{it}) + \varepsilon^u_{1,it})$$

Type II mismatch (false acceptance of the outside option) :

$$P(v_{0t} + \varepsilon^u_{0,it} > v_{1t}(\mu^*_{it}, \sigma^{2*}_{it}) + \varepsilon^u_{1,it} \mid v_{0t} + \varepsilon^u_{0,it} < v_{1t}(h_{it}) + \varepsilon^u_{1,it})$$

Figure 3.5 shows the share of mismatched workers in the two occupations in which initially unknown abilities are distributed with very similar dispersion: Executive/Managerial and Sales/Admin Support. I find that, in the early tenure years, both types of mismatch are non-trivial and more frequent in Sales/Admin Support, which is attributable to its features in the learning process and outside options. In this occupation, the wage signal is relatively noisier, implying that workers have higher chances of experiencing extreme luck and the learning process is slow. In conjunction with this learning environment, the value of quitting increases faster. These features lead the Sales/Admin Support workers with initially high ability to wrongly quit the occupation, and workers with initially low

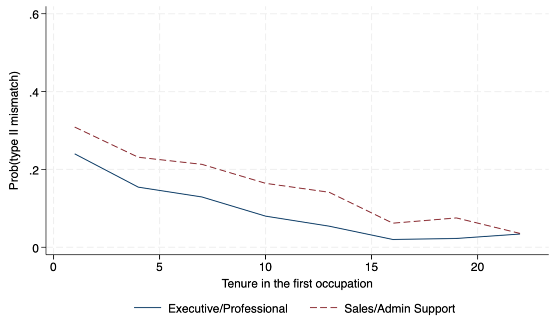
[27]The probabilities consider the fact that the current decision could be driven by the realization of the utility shocks, so the mistakes are ex-ante not deterministic. Also, note that the incentive to stay working while anticipating a potential large skill gain through the skill accumulation shock in the future is incorporated into the continuation value of staying.

ability to stay driven by optimistic beliefs. The mismatch may limit the workers' chance to accumulate the skill in the occupation that could be potentially aligned with their true ability.[28] Once the beliefs are refined as the workers stay longer, the mismatch displays convergence, but the probabilities are not zero. This is because the underlying uncertainty never disappears when the skill accumulation is stochastic.

Figure 3.5: Dynamics of mismatch

(a) Type I mismatch

(b) Type II mismatch



3.5.2 Counterfactual Policy: Information provision

I consider a counterfactual policy to have workers better informed about their ability during the first year of the employment relationship. A special case would be that workers have perfect information from the second period on, which can help them make better choices without the information friction. In practice, this could be an intervention to introduce a period and a chance for workers to inspect their skill level more seriously on the job.[29]

I simulate the occupational decisions and wage dynamics of the workers who are initially assigned to Sales/Admin Support occupation where Type 1 mismatch is relatively more frequent than Type 2 mismatch. Influenced by the counterfactual policy, neither type of mismatch occurs. Partly because the effect of preventing workers from falsely rejecting the

[28]Both types of mismatch also have implications for the firm-side which is not accounted for in the structural model. For example, a high type I mismatch results in a higher probability of losing highly productive workers. while high type II mismatch leads to longer duration to stay working for low productivity workers, which can potentially incur (opportunity) costs to the firms.

[29]This policy can cost additional effort to the workers and/or resources to the firms. For simplicity, I abstract from the additional changes, potentially induced by the policy, that alter the economic agents' behavior outside of the structural model.

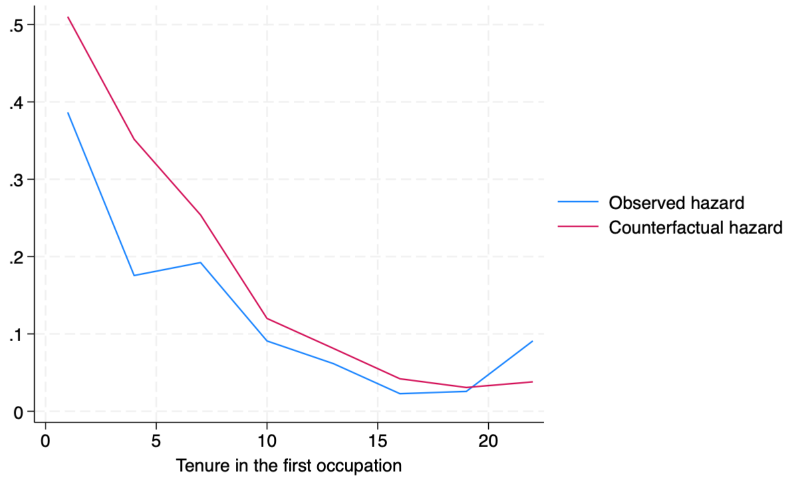
outside option is large, the counterfactual policy leads to a higher hazard rate as illustrated in the first panel of Figure 3.6. That is, once workers can make the occupational decision by correctly comparing the benefits of staying in and quiting the initial occupation with perfect information, they would become more likely to quit the occupation. As a result of the sorting that the workers whose skill is high enough to anticipate enough wage potential decide to stay each year, actual wage dynamics become different. The impact on wage distributions is described in the second and third panel of Figure 3.6. Through the sorting effect, the counterfactual policy yields a faster wage growth while the skill levels of the stayers get less dispersed.

3.6 Conclusion

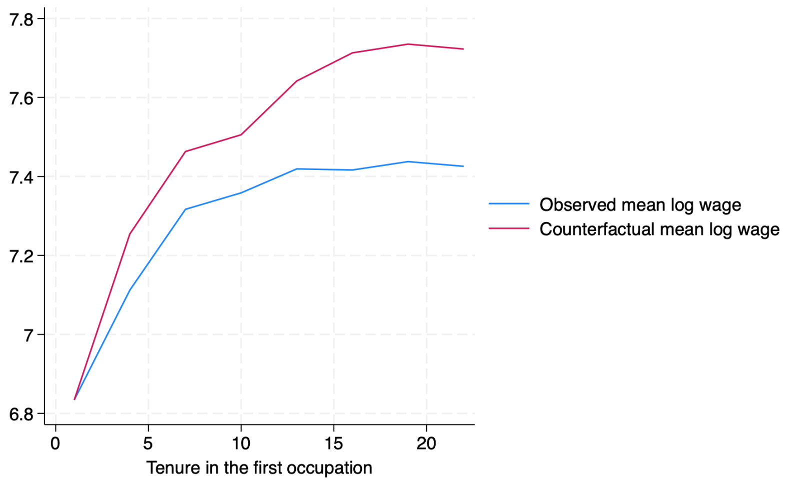
In this paper, I study the occupational decision and wage dynamics based on the human capital framework. I first present empirical motivation that different occupations show different patterns of wage dynamics. Along with the differences across the occupations, it is shown that the decision to stay in or quit their first occupations is associated with new information obtained from wages realized in the previous period. Given the hypothesis that workers' learning about their own skill could explain their response to the new information while influencing the wage dynamics, I built a structural model. The model incorporating the workers' learning in a Bayesian fashion in the labor market is estimated using the Kalman filter and the CCP estimation method. The estimates can mimic the data patterns fairly well. The estimation results are in line with the recent research suggesting the importance of workers' learning about their own abilities. First of all, I provide supporting evidence that the workers in their first occupation actually experience uncertainty about their own abilities. It is found to have significant implications on occupational decisions and workers' sorting, which, in turn, has an impact on wage dynamics. More specifically, I show that a non-trivial share of workers make choices that could lead to a type of mismatch attributed to their beliefs and would not have been made if the actual skill levels were informed to the workers. The counterfactual policy of interest, information provision about workers' skills from the start, is found to have a sorting effect on the workers by influencing the duration of working in the initial occupation, particularly Sales/Admin Support occupation in my counterfactual analysis, while increasing the mean and lowering the dispersion of the log wage distribution among the workers staying in the first occupations.

Figure 3.6: Counterfactual labor market outcomes: Sales/Admin Support occupation

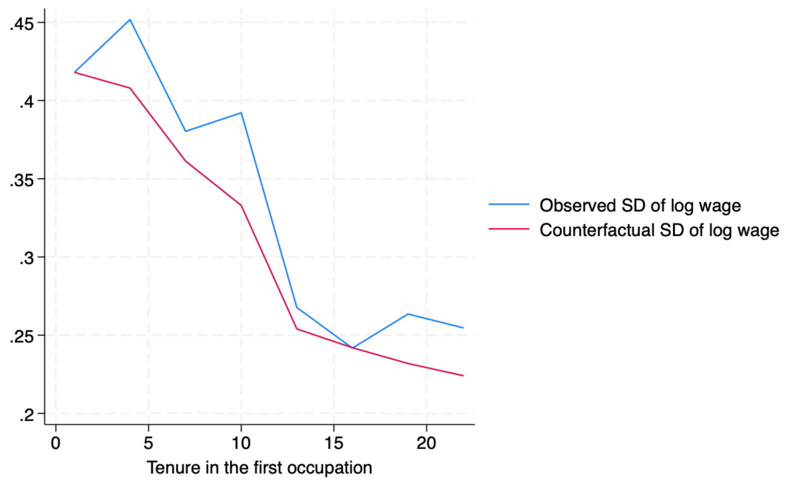
(a) Counterfactual hazard rate



(b) Counterfactual mean of log wage



(c) Counterfactual standard deviation of log wages



Bibliography

Abbring, Jaap H, "Identification of dynamic discrete choice models," *Annu. Rev. Econ.*, 2010, *2* (1), 367–394.

Agopsowicz, Andrew, Chris Robinson, Ralph Stinebrickner, and Todd Stinebrickner, "Careers and mismatch for college graduates: College and noncollege jobs," *Journal of Human Resources*, 2020, *55* (4), 1194–1221.

Albrecht, James and Susan Vroman, "A Matching Model with Endogenous Skill Requirements.," *International Economic Review*, 2002, *43* (1).

Allen, Jim and Rolf Van der Velden, "Educational mismatches versus skill mismatches: effects on wages, job satisfaction, and on-the-job search," *Oxford economic papers*, 2001, *53* (3), 434–452.

Altonji, Joseph G, Erica Blom, and Costas Meghir, "Heterogeneity in human capital investments: High school curriculum, college major, and careers," *Annu. Rev. Econ.*, 2012, *4* (1), 185–223.

_ , Prashant Bharadwaj, and Fabian Lange, "Changes in the characteristics of American youth: Implications for adult outcomes," *Journal of Labor Economics*, 2012, *30* (4), 783–828.

Antonovics, Kate and Limor Golan, "Experimentation and job choice," *Journal of Labor Economics*, 2012, *30* (2), 333–366.

Arcidiacono, Peter, "Affirmative action in higher education: How do admission and financial aid rules affect future earnings?," *Econometrica*, 2005, *73* (5), 1477–1524.

_ and Paul B Ellickson, "Practical methods for estimation of dynamic discrete choice models," *Annu. Rev. Econ.*, 2011, *3* (1), 363–394.

_ and Robert A Miller, "Conditional choice probability estimation of dynamic discrete choice models with unobserved heterogeneity," *Econometrica*, 2011, *79* (6), 1823–1867.

_ and _ , "Identifying dynamic discrete choice models off short panels," *Journal of Econometrics*, 2020, *215* (2), 473–485.

_ , Esteban Aucejo, Arnaud Maurel, and Tyler Ransom, "College attrition and the dynamics of information revelation," *Journal of Political Economy*, 2025, *133* (1), 53–110.

___ , **V Joseph Hotz, and Songman Kang**, "Modeling college major choices using elicited measures of expectations and counterfactuals," *Journal of Econometrics*, 2012, *166* (1), 3–16.

Ashworth, Jared, V Joseph Hotz, Arnaud Maurel, and Tyler Ransom, "Changes across cohorts in wage returns to schooling and early work experiences," *Journal of labor economics*, 2021, *39* (4), 931–964.

Attanasio, Orazio and Katja Kaufmann, "Educational choices, subjective expectations, and credit constraints," Technical Report, National Bureau of Economic Research 2009.

Attanasio, Orazio P and Katja M Kaufmann, "Education choices and returns to schooling: Mothers' and youths' subjective expectations and their role by gender," *Journal of Development Economics*, 2014, *109*, 203–216.

Autor, David H and David Dorn, "The growth of low-skill service jobs and the polarization of the US labor market," *American economic review*, 2013, *103* (5), 1553–1597.

___ **and Michael J Handel**, "Putting tasks to the test: Human capital, job tasks, and wages," *Journal of labor Economics*, 2013, *31* (S1), S59–S96.

___ , **Lawrence F Katz, and Melissa S Kearney**, "Trends in US wage inequality: Revising the revisionists," *The Review of economics and statistics*, 2008, *90* (2), 300–323.

Baert, Stijn, Bart Cockx, and Dieter Verhaest, "Overeducation at the start of the career: stepping stone or trap?," *Labour Economics*, 2013, *25*, 123–140.

Barnichon, Regis and Yanos Zylberberg, "Underemployment and the trickle-down of unemployment," *American Economic Journal: Macroeconomics*, 2019, *11* (2), 40–78.

Beffy, Magali, Denis Fougere, and Arnaud Maurel, "Choosing the field of study in postsecondary education: Do expected earnings matter?," *Review of Economics and Statistics*, 2012, *94* (1), 334–347.

Belzil, Christian and Jörgen Hansen, "Unobserved ability and the return to schooling," *Econometrica*, 2002, *70* (5), 2075–2091.

Ben-Porath, Yoram, "The production of human capital and the life cycle of earnings," *Journal of political economy*, 1967, *75* (4, Part 1), 352–365.

Bhat, Chandra R, "A heteroscedastic extreme value model of intercity travel mode choice," *Transportation Research Part B: Methodological*, 1995, *29* (6), 471–483.

Blundell, Richard, "What have we learned from structural models?," *American Economic Review*, 2017, *107* (5), 287–292.

Bobba, Matteo and Veronica Frisancho, "Self-perceptions about academic achievement: Evidence from Mexico City," *Journal of Econometrics*, 2022, *231* (1), 58–73.

Bunting, Jackson, Paul Diegert, and Arnaud Maurel, "Heterogeneity, Uncertainty and Learning: Semiparametric Identification and Estimation," Technical Report, National Bureau of Economic Research 2024.

Burdett, Kenneth, "A theory of employee job search and quit rates," *The American Economic Review*, 1978, *68* (1), 212–220.

Camargo, Braz, Fabian Lange, and Elena Pastorino, "On the Role of Learning, Human Capital, and Performance Incentives for Wages," Technical Report, National Bureau of Economic Research 2022.

Cameron, Stephen V and James J Heckman, "Life cycle schooling and dynamic selection bias: Models and evidence for five cohorts of American males," *Journal of Political economy*, 1998, *106* (2), 262–333.

__ **and** __ , "The dynamics of educational attainment for black, hispanic, and white males," *Journal of political Economy*, 2001, *109* (3), 455–499.

Carneiro, Pedro, James J Heckman, and Edward J Vytlacil, "Estimating marginal returns to education," *American Economic Review*, 2011, *101* (6), 2754–2781.

__ **, Karsten T Hansen, and James J Heckman**, "2001 Lawrence R. Klein lecture estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice," *International Economic Review*, 2003, *44* (2), 361–422.

Carrillo-Tudela, Carlos and Ludo Visschers, "Unemployment and endogenous reallocation over the business cycle," *Econometrica*, 2023, *91* (3), 1119–1153.

Chiang, Chun-Fang and Brian Knight, "Media bias and influence: Evidence from newspaper endorsements," *The Review of economic studies*, 2011, *78* (3), 795–820.

Cho, Sung Jin, "Overeducation in the Korean labor market: Evidence from young workers," *Journal of Economic Research*, 2014, *19* (1), 19–59.

Clark, Brian, Clément Joubert, and Arnaud Maurel, "The career prospects of overeducated Americans," *IZA Journal of Labor Economics*, 2017, *6*, 1–29.

Cunha, Flavio, James Heckman, and Salvador Navarro, "Separating uncertainty from heterogeneity in life cycle earnings," *oxford Economic papers*, 2005, *57* (2), 191–261.

Declercq, Koen and Frank Verboven, "Enrollment and degree completion in higher education without admission standards," *Economics of Education Review*, 2018, *66*, 223–244.

Delavande, Adeline and Basit Zafar, "University choice: The role of expected earnings, nonpecuniary outcomes, and financial constraints," *Journal of Political Economy*, 2019, *127* (5), 2343–2393.

_ **and Hans-Peter Kohler**, "HIV/AIDS-related expectations and risky sexual behaviour in Malawi," *The Review of Economic Studies*, 2016, *83* (1), 118–164.

DellaVigna, Stefano and Ethan Kaplan, "The Fox News effect: Media bias and voting," *The Quarterly Journal of Economics*, 2007, *122* (3), 1187–1234.

der Klaauw, Wilbert Van, "On the use of expectations data in estimating structural dynamic choice models," *Journal of Labor Economics*, 2012, *30* (3), 521–554.

d'Haultfoeuille, Xavier, Christophe Gaillac, and Arnaud Maurel, "Rationalizing rational expectations: Characterizations and tests," *Quantitative Economics*, 2021, *12* (3), 817–842.

d'Haultfoeuille, Xavier and Arnaud Maurel, "Inference on an extended Roy model, with an application to schooling decisions in France," *Journal of Econometrics*, 2013, *174* (2), 95–106.

Foley, Kelly, Giovanni Gallipoli, and David A Green, "Ability, parental valuation of education, and the high school dropout decision," *Journal of Human Resources*, 2014, *49* (4), 906–944.

Gallipoli, Giovanni and Christos A Makridis, "Structural transformation and the rise of information technology," *Journal of monetary economics*, 2018, *97*, 91–110.

Gautier, Pieter A, "Unemployment and search externalities in a model with heterogeneous jobs and workers," *Economica*, 2002, *69* (273), 21–40.

Gibbons, Robert and Kevin J Murphy, "Optimal incentive contracts in the presence of career concerns: Theory and evidence," *Journal of political Economy*, 1992, *100* (3), 468–505.

_ , Lawrence F Katz, Thomas Lemieux, and Daniel Parent, "Comparative advantage, learning, and sectoral wage determination," *Journal of labor economics*, 2005, *23* (4), 681–724.

Gong, Y, "Signal-based learning models without the rational expectations assumption: Identification and counterfactuals," *Working Paper*, 2019.

Gorry, Aspen, Devon Gorry, and Nicholas Trachter, "Learning and life cycle patterns of occupational transitions," *International Economic Review*, 2019, *60* (2), 905–937.

Gottschalk, Peter and Michael Hansen, "Is the proportion of college workers in noncollege jobs increasing?," *Journal of Labor Economics*, 2003, *21* (2), 449–471.

Groes, Fane, Philipp Kircher, and Iourii Manovskii, "The U-shapes of occupational mobility," *The Review of Economic Studies*, 2015, *82* (2), 659–692.

Groote, Olivier De and Frank Verboven, "Subsidies and time discounting in new technology adoption: Evidence from solar photovoltaic systems," *American Economic Review*, 2019, *109* (6), 2137–2172.

Guvenen, Fatih, Burhan Kuruscu, Satoshi Tanaka, and David Wiczer, "Multidimensional skill mismatch," *American Economic Journal: Macroeconomics*, 2020, *12* (1), 210–244.

Hartog, Joop, "Over-education and earnings: where are we, where should we go?," *Economics of education review*, 2000, *19* (2), 131–147.

Heckman, James J, "The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models," in "Annals of economic and social measurement, volume 5, number 4," NBER, 1976, pp. 475–492.

—, "Sample selection bias as a specification error," *Econometrica: Journal of the econometric society*, 1979, pp. 153–161.

— **and Burton Singer**, "Econometric duration analysis," *Journal of econometrics*, 1984, *24* (1-2), 63–132.

— **and Guilherme Sedlacek**, "Heterogeneity, aggregation, and market wage functions: an empirical model of self-selection in the labor market," *Journal of political Economy*, 1985, *93* (6), 1077–1125.

—**, John Eric Humphries, and Gregory Veramendi**, "Returns to education: The causal effects of education on earnings, health, and smoking," *Journal of Political Economy*, 2018, *126* (S1), S197–S246.

Hoffman, Mitchell and Stephen V Burks, "Worker overconfidence: Field evidence and implications for employee turnover and firm profits," *Quantitative Economics*, 2020, *11* (1), 315–348.

Holt, Charles A and Susan K Laury, "Risk aversion and incentive effects," *American economic review*, 2002, *92* (5), 1644–1655.

Hotz, V Joseph and Robert A Miller, "Conditional choice probabilities and the estimation of dynamic models," *The Review of Economic Studies*, 1993, *60* (3), 497–529.

Hudomiet, Peter, Gábor Kézdi, and Robert J Willis, "Stock market crash and expectations of American households," *Journal of Applied Econometrics*, 2011, *26* (3), 393–415.

Huggett, Mark, Gustavo Ventura, and Amir Yaron, "Human capital and earnings distribution dynamics," *Journal of Monetary Economics*, 2006, *53* (2), 265–290.

Jacob, Brian, Brian McCall, and Kevin Stange, "College as country club: Do colleges cater to students' preferences for consumption?," *Journal of Labor Economics*, 2018, *36* (2), 309–348.

James, Jonathan, "Ability matching and occupational choice," 2011.

— , "Learning and occupational sorting," Technical Report 2012.

Jovanovic, Boyan, "Job matching and the theory of turnover," *Journal of political economy*, 1979, *87* (5, Part 1), 972–990.

— **and Yaw Nyarko**, "Stepping-stone mobility," in "Carnegie-Rochester Conference Series on Public Policy," Vol. 46 Elsevier 1997, pp. 289–325.

Kalouptsidi, Myrto, Paul T Scott, and Eduardo Souza-Rodrigues, "Identification of counterfactuals in dynamic discrete choice models," *Quantitative Economics*, 2021, *12* (2), 351–403.

Kaplan, Greg and Sam Schulhofer-Wohl, "The changing (dis-) utility of work," *Journal of Economic Perspectives*, 2018, *32* (3), 239–258.

Kapor, Adam J, Christopher A Neilson, and Seth D Zimmerman, "Heterogeneous beliefs and school choice mechanisms," *American Economic Review*, 2020, *110* (5), 1274–1315.

Kasahara, Hiroyuki and Katsumi Shimotsu, "Nonparametric identification of finite mixture models of dynamic discrete choices," *Econometrica*, 2009, *77* (1), 135–175.

Katz, Lawrence F and Kevin M Murphy, "Changes in relative wages, 1963–1987: supply and demand factors," *The quarterly journal of economics*, 1992, *107* (1), 35–78.

— **et al.**, "Changes in the wage structure and earnings inequality," in "Handbook of labor economics," Vol. 3, Elsevier, 1999, pp. 1463–1555.

Keane, Michael P and Kenneth I Wolpin, "The career decisions of young men," *Journal of political Economy*, 1997, *105* (3), 473–522.

— **, Petra E Todd, and Kenneth I Wolpin**, "The structural estimation of behavioral models: Discrete choice dynamic programming methods and applications," in "Handbook of labor economics," Vol. 4, Elsevier, 2011, pp. 331–461.

Kinsler, Josh and Ronni Pavan, "The specificity of general human capital: Evidence from college major choice," *Journal of Labor Economics*, 2015, *33* (4), 933–972.

Korpi, Tomas and Michael Tåhlin, "Educational mismatch, wages, and wage growth: Overeducation in Sweden, 1974–2000," *Labour economics*, 2009, *16* (2), 183–193.

Koşar, Gizem and Cormac O'Dea, "Expectations data in structural microeconomic models," in "Handbook of Economic Expectations," Elsevier, 2023, pp. 647–675.

Krueger, Alan B and Mikael Lindahl, "Education for growth: Why and for whom?," *Journal of economic literature*, 2001, *39* (4), 1101–1136.

Lazear, Edward, "Education: consumption or production?," *Journal of Political Economy*, 1977, *85* (3), 569–597.

Lee, Hanol, Jong-Wha Lee, and Eunbi Song, "Effects of educational mistmatch on wages in the Korean labor market," *Asian Economic Journal*, 2016, *30* (4), 375–400.

Leuven, Edwin and Hessel Oosterbeek, "Overeducation and mismatch in the labor market," *Handbook of the Economics of Education*, 2011, *4*, 283–326.

Magnac, Thierry and David Thesmar, "Identifying dynamic discrete decision processes," *Econometrica*, 2002, *70* (2), 801–816.

_ **and Sébastien Roux**, "Heterogeneity and wage inequalities over the life cycle," *European Economic Review*, 2021, *134*, 103715.

_ **, Nicolas Pistolesi, and Sébastien Roux**, "Post-schooling human capital investments and the life cycle of earnings," *Journal of Political Economy*, 2018, *126* (3), 1219–1249.

Manski, Charles F, "Nonparametric bounds on treatment effects," *The American Economic Review*, 1990, *80* (2), 319–323.

_ , "Measuring expectations," *Econometrica*, 2004, *72* (5), 1329–1376.

Mas, Alexandre and Amanda Pallais, "Valuing alternative work arrangements," *American Economic Review*, 2017, *107* (12), 3722–3759.

McGuinness, Seamus, "Overeducation in the labour market," *Journal of economic surveys*, 2006, *20* (3), 387–418.

Miller, Robert A, "Job matching and occupational choice," *Journal of Political economy*, 1984, *92* (6), 1086–1120.

Mincer, Jacob, "Schooling, experience, and earnings. Human behavior & social institutions no. 2.," 1974.

Moore, Don A and Paul J Healy, "The trouble with overconfidence.," *Psychological review*, 2008, *115* (2), 502.

Navarini, Lorenzo and Dieter Verhaest, "Educational Attainment, Overeducation and Wages: Evidence from a Dynamic Model," in "Conference of the European Association of Labour Economists, Location: Prague" 2023.

Neal, Derek, "The complexity of job mobility among young men," *Journal of labor Economics*, 1999, *17* (2), 237–261.

Nybom, Martin, "The distribution of lifetime earnings returns to college," *Journal of Labor Economics*, 2017, *35* (4), 903–952.

Oliver, Kang Buseong Christian, "S Korea faces problem of 'over-education'," 2010.

Oosterbeek, Hessel and Hans Van Ophem, "Schooling choices: Preferences, discount rates, and rates of return," *Empirical Economics*, 2000, *25*, 15–34.

Papageorgiou, Theodore, "Learning your comparative advantages," *Review of Economic Studies*, 2014, *81* (3), 1263–1295.

Pavan, Ronni, "Career choice and wage growth," *Journal of Labor Economics*, 2011, *29* (3), 549–587.

Polachek, Solomon W, Tirthatanmoy Das, and Rewat Thamma-Apiroam, "Micro-and macroeconomic implications of heterogeneity in the production of human capital," *Journal of Political Economy*, 2015, *123* (6), 1410–1455.

Rodríguez, Jorge, Sergio Urzúa, and Loreto Reyes, "Heterogeneous economic returns to post-secondary degrees: Evidence from Chile," *Journal of Human Resources*, 2016, *51* (2), 416–460.

Roy, Andrew Donald, "Some thoughts on the distribution of earnings," *Oxford economic papers*, 1951, *3* (2), 135–146.

Rubinstein, Yona and Yoram Weiss, "Post schooling wage growth: Investment, search and learning," *Handbook of the Economics of Education*, 2006, *1*, 1–67.

Rust, John, "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, 1987, pp. 999–1033.

_ , "Structural estimation of Markov decision processes," *Handbook of econometrics*, 1994, *4*, 3081–3143.

Sanders, Carl, "Skill accumulation, skill uncertainty, and occupational choice," *Working Paper*, 2014.

Sicherman, Nachum, "" Overeducation" in the labor market," *Journal of labor Economics*, 1991, *9* (2), 101–122.

_ **and Oded Galor**, "A theory of career mobility," *Journal of political economy*, 1990, *98* (1), 169–192.

Small, Kenneth A, "A discrete choice model for ordered alternatives," *Econometrica: Journal of the Econometric Society*, 1987, pp. 409–424.

Stange, Kevin M, "An empirical investigation of the option value of college enrollment," *American Economic Journal: Applied Economics*, 2012, *4* (1), 49–84.

Stinebrickner, Ralph and Todd R Stinebrickner, "A major in science? Initial beliefs and final outcomes for college major and dropout," *Review of Economic Studies*, 2014, *81* (1), 426–472.

Taber, Christopher and Rune Vejlin, "Estimation of a roy/search/compensating differential model of the labor market," *Econometrica*, 2020, *88* (3), 1031–1069.

Trachter, Nicholas, "Stepping stone and option value in a model of postsecondary education," *Quantitative Economics*, 2015, *6* (1), 223–256.

Verbeek, Marno and Theo Nijman, "Testing for selectivity bias in panel data models," *International Economic Review*, 1992, pp. 681–703.

Willis, Robert J and Sherwin Rosen, "Education and self-selection," *Journal of political Economy*, 1979, *87* (5, Part 2), S7–S36.

Wiswall, Matthew and Basit Zafar, "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, 2015, *82* (2), 791–824.

__ **and** __ , "Preference for the workplace, investment in human capital, and gender," *The Quarterly Journal of Economics*, 2018, *133* (1), 457–507.

Yamaguchi, Shintaro, "Tasks and heterogeneous human capital," *Journal of Labor Economics*, 2012, *30* (1), 1–53.

Zafar, Basit, "How do college students form expectations?," *Journal of Labor Economics*, 2011, *29* (2), 301–348.

Appendix

Appendix for Chapter 1

A. Korean education and employment panel structure

Figure 7: Panel Structure

<table>
<tr><th>Year</th><th>1st wave, 2004</th><th>…</th><th>4th wave, 2007</th><th>…</th><th>12th wave, 2015</th></tr>
<tr><td rowspan="2">Middle School Cohort</td><td>3rd grade of middle school students (2000)</td><td colspan="4">Tracking</td></tr>
<tr><td colspan="2"></td><td>3rd grade of general /vocational high school students (600 / 1000)</td><td colspan="2">Tracking</td></tr>
<tr><td rowspan="2">High School Cohort</td><td>3rd grade of general high school students (2000)</td><td colspan="4">Tracking</td></tr>
<tr><td>3rd grade of vocational high school students (2000)</td><td colspan="4">Tracking</td></tr>
</table>

Table 4: Number of respondents

Wave	Middle school	High school	Middle school	High school
1	3879		3879	
2	3523		3442	
3	3080		2902	
4	2500	2840	2500	2569
5	2773	2741	2204	2319
6	2652	2711	1962	2121
7	2543	2513	1803	1929
8	2429	2430	1642	1767
9	2466	2522	1546	1683
10	2485	2560	1504	1635
11	2507	2546	1457	1579
12	2498	2554	1423	1537

* The left table shows the number of respondents in each survey. In contrast, the right table shows the number when the respondents not surveyed in a certain year are removed from that year on. Note that the numbers still include individuals reporting invalid/missing values of the key variables. In the 4th wave of the survey, the middle school cohort students turn to the senior year in high school which will be the initial period in the dynamic discrete choice model, so the observations in the previous waves are ruled out.

B. Data construction rules and overeducation measure

1. I exclude individuals who do not graduate from high school and do graduate before the senior year (i.e., early graduation). The initial period in the dynamic model is the senior year in high school, so the choice made in the period represents the high school type students graduate from.

2. Any individual is classified as attending college for the year if he/she reports being enrolled in at least one of the two semesters which runs from March to June and from September to December. Colleges are classified into two groups, vocational college (VC) and academic college (AC). 1-year education in a polytechnic is not treated as a college education but as home production.

3. Employment choices are coded based on the labor force status when respondents are interviewed. This period overlaps to a great extent with the second semester of an academic year. An individual is considered as being employed if he reports working but not studying in the year. That is, working while in school (including one semester for work and one semester for school) is not considered. I only consider the primary job if an individual has more than one job, and wages are recorded as the average monthly compensation of the primary job.

4. A worker whose earning (expressed in 2015 wons) is below the one earned by receiving minimum hourly wage for 10 hours per week is classified as not working. (This practice is often found in the literature on dynamic discrete choice models of human capital accumulation. Arcidiacono et al. (2025) is an example.) Unpaid family workers are also not considered working.

5. Based on the self-assessed overeducation, workers having a college degree, either associate or bachelor's degree, are classified into one of the two categories: overeducation or matched employment. That is, he/she is considered overeducated in the labor market if she "agrees" or "strongly agrees" that her education level is higher than the required level of his job. I further assume that overeducated workers work in the non-college labor sector (i.e., jobs that do not require college-level skill).

6. Those who are not in education and employment are classified as, so called, home production (i.e., not in education or employment).

C. CSAT grades prediction

I assume that the CSAT grades reflect "innate" ability. By taking the average of the CSAT Korean grade and CSAT English grade, when observed, a scalar measure of ability is constructed. The measure runs from 1 to 9 with 0.5 as the unit of increase. To address missing values, regression equations of the grades for each subject are estimated and then used to predict the missing values. One of the regressors is the high school types that individuals attend. It makes sense to include the variable because there might be ability sorting across different high school types. Based on the same reasoning, the type of college that they enroll in within the two years after high school graduation is included as a regressor.

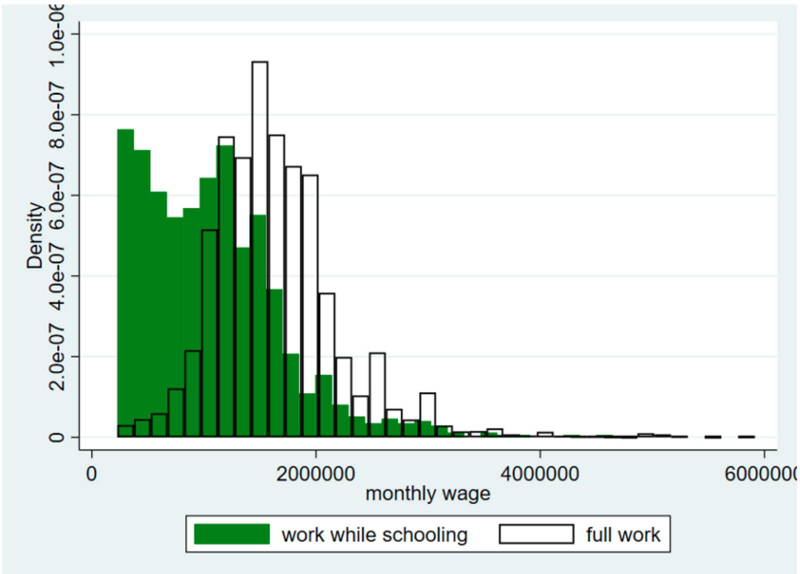
Table C presents the full list of explanatory variables and reports the estimation results. In addition to the estimates of the ordered logit model, the OLS regression result is reported to have easier interpretation. As expected, it is found that those with higher intended education level and higher effort level (i.e., hours for self-study and private tutoring coded into a categorical variable) are more likely to get significantly higher test grades. I replace the missing grades with the predicted ones based on the ordered logit specification, and then they are used to build the ability measure.

* CSAT grade is originally symmetric about 5. Therefore, after constructing the ability measure ranging from 1 to 9, it is subtracted by 5 to make its range [–4, 4].

D. Measures of overeducation (overeducation)

There are broadly three classes of overeducation measures, based on how required schooling is measured, that compare individuals' education attainment with educational requirement of jobs or occupations. As pointed out by Leuven and Oosterbeek (2011), there are various ways to determine the skill requirement of a job/occupation, but no approach is

Figure 8: Predicted (green) and observed (white) ability measure



immune to classification errors and a-priori assumptions. Here, I describe the measures and their advantages/disadvantages.

Self-assessment

Measures based on workers' self-assessment rely on questions that ask workers about the schooling requirements of their job. Instead, it can be asked to respondents directly whether they are overeducated, undereducated, or rightly educated for their job. The advantage is that this measure is based on all the relevant information. The downside of this subjective measure would be that respondents could have a tendency to over- or under-state their employment status even systematically differently across some attributes.

Realized match

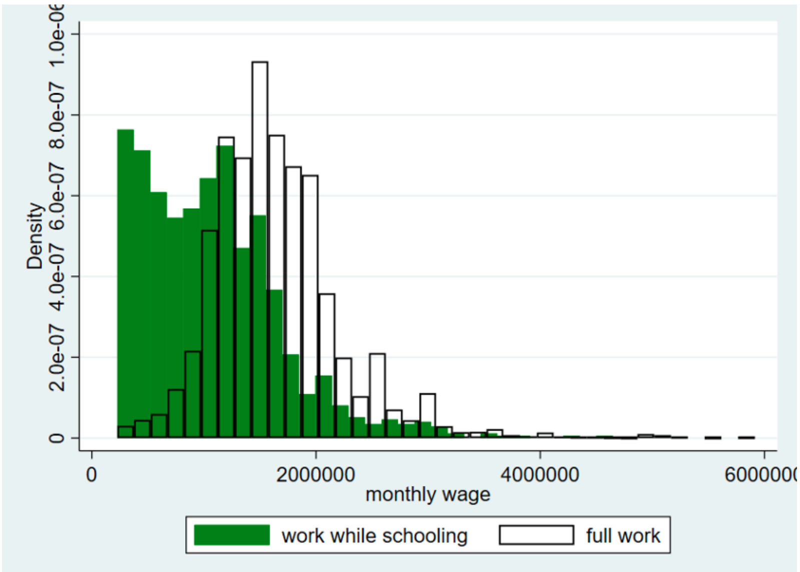
The required amount of schooling for a worker is inferred from the mean (or mode) of completed schooling of all workers holding the same occupation. Workers are defined to be overeducated if their completed level of schooling deviates at least one standard deviation from the mean in their occupation, which is somewhat arbitrary. Gottschalk and Hansen (2003) suggested a way to define a college job; Occupations with a large (i.e., above a threshold) college premium estimated from occupation-specific log wage regressions is classified as college jobs because the skills college workers have are valued by employers in that occupation. However, realized match does not measure the requirements of the occupations in the sense that it is the result of demand and supply in the labor market. Also, variation in required schooling across jobs within an occupation is ignored.

Job analysis

The measurement of required schooling level via job analysis based on information contained in occupational classification (e.g., Dictionary of Occupational Titles). It is based on the technology of the job, but a problem is that the evaluation of educational requirement is infrequent. More importantly, it is often assumed that there is a fixed required schooling level within an occupation, disregarding different requirement across jobs.

E. Wages distribution when working while attending college

Figure 9: Wages distributions: work while attending college (green) and not (white)



F. Finite dependence and difference in conditional value functions

Along the sequence of choices $\{d_t = j, d_{t+1} = 0, d_{t+2} = 0\}$, the conditional value function can be expressed in terms of two-period ahead flow payoffs, conditional choice probabilities, and the expected value function:

$$
\begin{aligned}
v(Z_t, d_t = j) = u(Z_t, d_t = j) - \beta \int \ln p(d_{t+1} = 0|Z_{t+1}) f(Z_{t+1}|Z_t, d_t = j) dZ_{t+1} \\
+ \beta \int v(Z_{t+1}, d_{t+1} = 0) f(Z_{t+1}|Z_t, d_t = j) dZ_{t+1} + \beta\gamma \\
= u(Z_t, d_t = j) + \beta \int u(Z_{t+1}, d_{t+1} = 0) f(Z_{t+1}|Z_t, d_t = j) dZ_{t+1} \\
- \beta \int \ln p(d_{t+1} = 0|Z_{t+1}) f(Z_{t+1}|Z_t, d_t = j) dZ_{t+1} \\
+ \beta^2 \int \int u(Z_{t+2}, d_{t+2} = 0) f(Z_{t+2}|Z_{t+1}, d_{t+1} = 0) f(Z_{t+1}|Z_t, d_t = j) dZ_{t+2} dZ_{t+1} \\
- \beta^2 \int \int \ln p(d_{t+2} = 0|Z_{t+2}) f(Z_{t+2}|Z_{t+1}, d_{t+1} = 0) f(Z_{t+1}|Z_t, d_t = j) dZ_{t+2} dZ_{t+1} \\
+ \beta^3 \int \int \int V(Z_{t+3}) f(Z_{t+3}|Z_{t+2}, d_{t+2} = 0) f(Z_{t+2}|Z_{t+1}, d_{t+1} = 0) f(Z_{t+1}|Z_t, d_t = j) dZ_{t+3} dZ_t \\
+ \beta^2\gamma + \beta\gamma
\end{aligned}
$$

The last equality uses the definition of:

$$
v(Z_{t+2}, d_{t+2} = 0) = u(Z_{t+2}, d_{t+2} = 0) + \beta \int V(Z_{t+3}) f(Z_{t+3}|Z_{t+2}, d_{t+2} = 0) dZ_{t+3}
$$

Similarly, $v(Z_t, d_t = 0)$ can be expressed along the sequence of choices $\{d_t = 0, d_{t+1} = j, d_{t+2} = 0\}$. The expected value functions cancel out when subtracting $v(Z_t, d_t = 0)$ from $v(Z_t, d_t = j)$, allowing the differenced conditional value function to be written only in terms of two-period ahead flow payoffs and estimated CCPs. The expression simplifies further when $u(Z_t, d_t = 0)$ is normalized to 0.

Appendix for Chapter 2

A. Data construction: general high school students

The initial number of general high school students in the dataset is 3190. The rules to restrict the sample is described below, and the numbers inside parentheses correspond to the number of valid observations left in our sample.

1. Graduate from a general high school, either science- or liberal arts-focused track (3179),
2. Report predetermined individual characteristics (household income, household education level, and school region) in the senior high school year (2802),
3. Take CSAT and report the test grades (i.e., drop those not taking the test or not reporting grades) (2418),
4. Report self-study hours/category in the senior high school year (2417),
5. Respond, at least once, to the survey from the year when the individual turns to age 25. Had reported her college attendance at least once before graduation if she identifies herself as an academic college graduate (1931).

How we construct the variables of observed choices: we focus on the final outcomes, in turn disregarding some details in the data. For example, changes in high school track or college dropouts are observed but not considered in the model.

B. College Scholastic Ability Test (CSAT)

It is worth mentioning several features of the CSAT because they would be relevant to establishing a measurement system: (a) students are evaluated after they have chosen and been exposed to different education because the test is taken by the end of high school while they are in different high school types and tracks, (b) students select the subjects to take examinations in. The subjects are largely divided into mathematics, Korean, English, social sciences/natural sciences/vocational education, and second foreign languages. Mathematics exam is additionally divided into two types (i.e., comprehensive/advanced) depending on the material addressed. (c) Lastly, standardized test scores were generally provided to the test takers, but in 2007, the scores were re-scaled on standardized 9-grades scale and provided to the test takers because of the coarse grading policy implemented but abolished right next year. The year 2007 corresponds to the year when middle school cohort students in KEEP took the exam in their last year of high school.

C. Ordered Generalised Extreme Value (OGEV)

Small (1987) proposed the OGEV model from the GEV class that is more suited for ordered discrete outcomes and consistent with RUM. While maintaining the flexibility of allowing the explanatory variables to have different coefficients and significance levels for the utilities attached to different choices, the OGEV model also relaxes the restriction of independence between the unobservable characteristics across different choices. Specifically, the OGEV model allows for correlations between the disturbances of outcomes that are "close to" each other in the ordering. The further the two outcomes j and k are located from one another, the smaller the correlation between the two disturbances ε_j and ε_k $(j,k = 1, ..., J)$ is. When $|j-k|$ is greater than a "pre-"selected integer M, the correlation is set to zero. It is possible to allow the window width M across nearby correlated outcomes to be arbitrarily large. To reduce the complexity of the model and therefore the associated computing demand in the maximisation procedures for estimation, we consider a standard OGEV model with only the adjacent outcomes correlated (i.e., $M = 1$). It involves, compared to the standard multinomial logit model, one additional parameter ρ of which inverse captures the correlation between the adjacent outcomes. The associated probabilities have the form

$$P_{ij} = exp(\rho^{-1}V_{ij}) \frac{[\{exp(\rho^{-1}V_{ij-1}) + exp(\rho^{-1}V_{ij})\}^{\rho-1} + \{exp(\rho^{-1}V_{ij}) + exp(\rho^{-1}V_{ij+1})\}^{\rho-1}]}{\sum_{j'=1}^{J+1}\{exp(\rho^{-1}V_{ij'-1}) + exp(\rho^{-1}V_{ij'})\}^{\rho}},$$

where, for notational convenience, $exp(\rho^{-1}V_{i0}) = exp(\rho^{-1}V_{iJ+1}) = 0$. $0 < \rho \leq 1$, and the probability converge to the multinomial logit one as $\rho \to 1$.

D. More general wage specification

The following tables report the estimation result of the model equipped with more general specification of the wage determination: potentially heterogeneous wage growth ($\gamma_{\ell'}$) and scale of the idiosyncratic wage shock ($\sigma_{\ell'}$) across the educational qualification ℓ'.

Table 5: Parameter Estimates

Parameters	Estimate
Type 1's belief, p_1^2	0.59897
	(0.096212)
Type 2's belief, P_1^2	0.28484
	(0.11234)
Type 1's share, π_1	0.39876
	(0.026162)
Measurement error, λ	0.2404
	(0.021127)

Table 6: Multinomial logit coefficients (stage 1,2,4,5)

Variables	High school track (Base: Liberal-art track)		Effort (Base: Low)		College enrollment (Base: not enrolled in any college)			Final qualification (Base: vocational college)	
	Science track	Vocational track	Middle	High	Vocational	Non-selective	Selective	Non-selective	Selective
female	−0.69259	−0.35227	0.16905	0.05449	0.92614	0.39328	0.30114	−0.31999	−1.0094
	(0.10028)	(0.13408)	(0.1293)	(0.1092)	(0.1564)	(0.13782)	(0.15385)	(0.18071)	(0.24484)
household income quartile	0.08359	−0.20158	−0.00442	0.12564	−0.35796	−0.24747	−0.32214	0.2364	0.30665
	(0.04701)	(0.06271)	(0.05930)	(0.05126)	(0.07343)	(0.06437)	(0.07284)	(0.0829)	(0.11403)
college educated parent	0.07568	−1.0325	0.20418	0.23841	−0.62462	−0.05193	0.11397	−0.13525	0.17973
	(0.11147)	(0.1926)	(0.14675)	(0.11946)	(0.19319)	(0.14955)	(0.16494)	(0.23077)	(0.28196)
urban residence	0.11809	0.10115	0.08609	0.3802	−0.22722	−0.56508	−0.5145	0.12428	0.17475
	(0.10031)	(0.13549)	(0.12853)	(0.10814)	(0.15218)	(0.13504)	(0.15213)	(0.17786)	(0.23852)
cohort	0.11806	0.91955	0.39662	−0.66829	−0.57925	−0.5541	−0.50605	0.5301	0.41061
	(0.10268)	(0.15752)	(0.14323)	(0.10871)	(0.17202)	(0.14596)	(0.16146)	(0.20739)	(0.26514)
High school: science track					−0.30431	−0.23455	0.44886	0.18132	0.35497
					(0.18107)	(0.1478)	(0.1637)	(0.21953)	(0.27286)
High school: vocational track					0.02986	−0.35963	0.85995	−0.19573	0.44012
					(0.26557)	(0.25663)	(0.32796)	(0.26735)	(0.4479)
CSAT grade					−0.63003	−0.31921	0.44592	0.3556	0.79142
					(0.05844)	(0.04730)	(0.05441)	(0.07119)	(0.09156)
College enrollment: not enrolled								−2.9705	−4.5247
								(0.29182)	(0.59397)
vocational college								−5.2227	−9.2025
								(0.2938)	(0.81766)
non-selective academic college									−4.691
									(0.61327)
selective academic college								−1.7468	
								(0.62384)	
Type 1′s Intercept	−0.93345	−0.9894	−1.7497	−1.1821	4.74	3.7443	−0.96842	0.20036	−0.5102
	(0.19456)	(0.2374)	(0.23631)	(0.2042)	(0.37954)	(0.33556)	(0.39016)	(0.41775)	(0.73496)
Type 2′s Intercept	−0.25464	−1.2718	−1.722	−0.75892	3.8852	3.79	−1.3033	2.0048	0.93921
	(0.16139)	(0.2351)	(0.2263)	(0.18428)	(0.3996)	(0.34222)	(0.41788)	(0.46084)	(0.78481)

Table 7: Ordinal logit coefficients for CSAT grade (stage 3)

Variables	CSAT grade
female	0.28076
	(0.077567)
household income quartile	0.17009
	(0.036287)
college educated parent	0.72973
	(0.090611)
urban residence	0.17422
	(0.077103)
cohort	0.10935
	(0.082598)
High school: science track	0.068228
	(0.089048)
High school: vocational track	-2.2101
	(0.13001)
Effort: middle	0.75074
	(0.11148)
Effort: high	0.92942
	(0.095737)
Type 1's Intercept	2.8069
	(0.19267)
Type 2's Intercept	3.4172
	(0.19735)
Threshold 2	1.6776
	(0.11536)
Threshold 3	2.9591
	(0.13136)
Threshold 4	4.012
	(0.13971)
Threshold 5	5.0841
	(0.14715)
Threshold 6	6.1063
	(0.15569)
Threshold 7	7.0422
	(0.16732)
Threshold 8	8.6081
	(0.21564)

Table 8: ln(wage) parameters by qualification group (stage 6)

Variables	Vocational college	Non-selective academic	Selective academic
female	-0.10036	-0.21203	-0.14927
	(0.01908)	(0.016766)	(0.026267)
household income quartile	-0.0082239	0.038125	0.0066505
	(0.0098157)	(0.0080439)	(0.014092)
college educated parent	0.033315	0.08584	0.14937
	(0.026345)	(0.020119)	(0.027089)
urban residence	-0.0075094	-0.015137	0.020181
	(0.017666)	(0.01597)	(0.029637)
cohort	-0.046867	-0.060041	0.0064547
	(0.020347)	(0.020153)	(0.031173)
Type 1's Intercept	0.046885	0.054273	0.049692
	(0.0028077)	(0.0038383)	(0.0063562)
Type 2's Intercept	5.1079	4.7618	4.8166
	(0.033986)	(0.032997)	(0.067118)
Years of experience	4.7847	5.1436	5.2367
	(0.037334)	(0.029107)	(0.05449)
Sigma	0.23498	0.24571	0.25537
	(0.0043446)	(0.004344)	(0.0064432)

Appendix for Chapter 3

A. Data construction

To construct annual panel data for our main analysis, we use NLSY79's Work History Data File, which records individuals' employment histories up to five jobs on a weekly basis from 1978 to 2010. Following Pavan (2011), total hours worked for each job within a year is calculated, based on which I define primary jobs for each year as the one for which an individual spent the most hours worked within the year. We construct panel data with annual frequency (from 1978 to 2010) from a series of observations of primary jobs.

Measurement error in occupational switching has received particular attention (e.g., Carrillo-Tudela and Visschers 2023). To avoid miscoded occupational switches, I address it by dropping transitions that immediately revert (which often indicates incorrect coding in the middle year). Arguably, conditioning occupation switches on simultaneous employer switches would reduce the measurement error, but I keep the possibility to change one's occupation under the same employer open while categorize the occupations broadly so that a minor measurement error could be address in this way.

Worker's wages are measured by the usual rate of pay for the primary job at the time of interview. These wages include tips, overtime and bonuses before deductions and are converted to an hourly rate using usual hours worked when not reported as such. All the wages are deflated by the price index for personal consumption expenditures into real term in the 2000 dollars. We drop the observations if their wages are missing.

I limit my sample to the individuals who make their initial long-term transition from school to labor markets during the survey period. In practice, those who work more than 1,200 hours in the initial year of the survey is dropped, and the year when they work more than 1,200 hours least for two following years during the survey period is set as his initial period considered in the sample. Those who are in the military service more than two years during the period are also eliminated from our sample. For the individuals who go back to school from the labor force during the survey period, I assume they start their career from the point they reenter labor markets, and drop the observations before that time. Being different from the wage growth literature excluding workers who are weakly attached to the labor force, I include them in the analysis. Finally, those with a valid occupation code and wage, demographics and ASVAB scores information above age 16 are kept.

B. Test for selection: variable addition

I discuss a simple way to test the presence of sample selection, attrition in particular, which motivates a model to control for and delve into the selection mechanism. Consider regressing the log wage received by worker i assigned to his initial occupation j having worked t years on observed characteristics. For simplicity, following the Mincer wage equation, consider the usual case where years of schooling, experience, and other observed ability measurements are in X_{it} as proxies for human capital, but time-invariant unobserved heterogeneity not captured by the observables is in the error term:

$$log(W_{ijt}) = X_{it}\gamma_j + \{A_{ij} + \varepsilon^p_{ijt}\} \tag{15}$$

Selection or attrition bias could occur because, with $d_i = (d_1, d_2, \ldots, d_{T_i}, d_{T_i+1})$ denoting the observability of the wages, the conditional expectation of the error term could be not equal zero: $E[A_{ij} + \varepsilon^p_{ijt}|d_i]$. If this conditional expectation is known, one could add it as an extra regressor so that the parameters in the extended model can be estimated consistently. This is the essence of the well-known two-step estimation procedure in the cross-sectional sample selection model proposed by Heckman (1976, 1979). However, the conditional expectation is not known before the selection process is known or jointly modeled. Building on Verbeek and Nijman (1992) who propose a simple test for selective nonresponse in panel data, the conditional expectation is approximated in a simple way: $E[A_{ij} + \varepsilon^p_{ijt}|d_i] = \kappa_{jt} d_{it+1}$.

The statistical result that says d_{it+1} significantly affects the log wage (i.e., κ_{jt} is significantly different from zero) rejects the hypothesis of no selection of sample. To make the test simple, I further assume $\kappa_{jt} = \kappa$. I start with regressing the log wage on observed individual characteristics such as race, education, the Armed Services Vocational Aptitude Battery (ASVAB) scores before labor market entry, and years of experience. The estimation results based on different specifications in this paper say d_{it+1} has an effect, which leads to the conclusion that the issues with selective attrition and the resulting non-random sample could underlie the wage process through the workers' decisions to stay in or quit their initial occupations.[30]

C. Linear wage contract and identification issues

Suppose a linear wage contract $W_{it} = a_{it} + b_t Y_{it}$, where Y_{it} is the produced output. Related identification issues are discussed in the following. Equipped with competitive market assumption ensuring zero ex-ante expected profit, fixed pay a_{it} can be expressed by a function of the expected output: $a_{it} = (1 - b_t)E[Y_{it}|\Omega_{it}]$, where $\Omega_{it} = \{\mu^*_{it}, \sigma^{2*}_{it}, X_{it}, \eta_i\}$ is information set at the start of period t. This yields the following wage equation:

$$\begin{aligned} W_{it} &= (1 - b_t)E[Y_{it}|\Omega_{it}] + b_t Y_{it} \\ &= (1 - b_t)\{\mu^*_{it} + \frac{\sigma^{2*}_{it}}{2} + \frac{\sigma^2_p}{2}\} + b_t \exp(h_{it} + \varepsilon^p_{it}) \end{aligned} \tag{16}$$

Piece rate Note that the (raw) wages based on the linear contract are composed of two components: fixed pay $(1 - b_t)E[Y_{it}|\Omega_{it}]$ and variable pay $b_t Y_{it}$. When one of the two components is observed along with wages W_{it}, the piece rate can be directly identified. To see this, I use the following relationship $E[W_{it}(\Omega_{it}, \varepsilon^p_{it})] = (1 - b_t)E[E[Y_{it}|\Omega_{it}]] + bE[Y_{it}]$, which yields $E[W_{it}] = E[Y_{it}]$. Hence, by denoting the observed fixed pay and variable pay by W^F and W^V, respectively, the following ratios identify the piece rate:

$$\frac{E[W^V]}{E[W_{it}]} = 1 - b_t, \quad \frac{E[W^F]}{E[W_{it}]} = b_t \tag{17}$$

Given the absence of information about the share of fixed and variable pays, without an exogenous variation that enables the piece rate to be identified, a parametric specification needs to be imposed not to invoke an identification problem; piece rates b_t are fixed to a certain value.

[30]The result is robust with respect to the specification including either fixed or random effects with the same set of explanatory variables.

D. Simulation practice: attrition bias

Individual heterogeneity and endogenous quitting decisions can posit a serious econometric challenge. In the absence of learning and stochastic skill accumulation, a popular approach to address the endogeneity through the time-invariant unobserved heterogeneity would be to introduce individual fixed effects and estimate the observational equation on the subsample of workers whose wage outcomes are observed repeatedly.

However, an issue of sample selection can lead the estimated coefficients of the time-varying variables using the fixed effects approach (i.e., within or first-difference estimators) to be biased. An example would be the presence of learning as in the model in this paper. In a learning environment, Gibbons et al. (2005) proposed instruments, in conjunction with the fixed effect approach, that exploit the key property of Bayesian learning models in which beliefs are a martingale.[31] The instruments, wages realized two or more periods ago, are based on the idea that new information obtained in the current period is orthogonal to the prior belief.

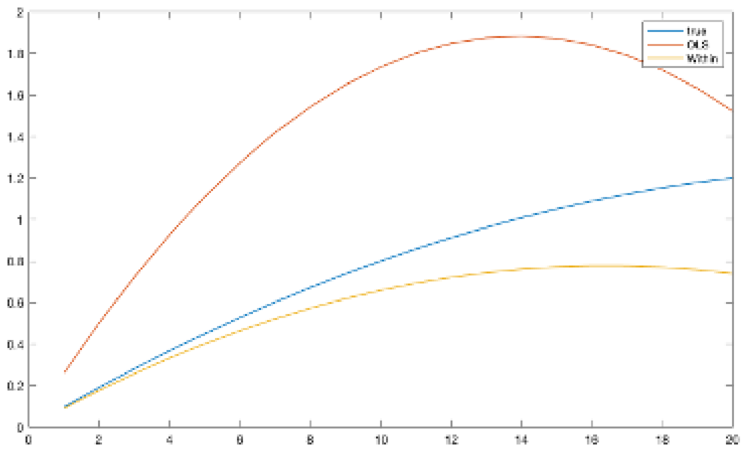
In learning environments where the choice in the current period (e.g., stay/quit, different occupations) contains information about the previously realized idiosyncratic components in the outcome equation, the instrumental variable (IV) estimator in Gibbons et al. (2005) would fail to correct the selection bias. This is regardless of the specification determining the wages; in two extreme cases of the linear contract, wages fully dependent either on the realized or expected output production, the instruments can not control for the sample selection appropriately.

A simulation practice below supposes a spell on a job and illustrates the biased estimates of the returns to tenure/seniority in the fixed effects approach in the uncertainty and learning environment. The direction of the sample selection bias is hard to predict in advance, though; in line with job search and matching theories of job mobility (e.g., Burdett 1978; Jovanovic 1979), it is plausible that workers perceiving themselves to be productive in the current job are likely to have longer tenure, predicting a positive bias. However, conversely, workers perceiving productivity may quit their current jobs at low

[31] The wages in their paper are assumed to be dependent only on the expected productivity (i.e., a linear contract with the piece rate being equal to zero).

tenures, which contributes to a negative bias in the tenure effect. It depends on the set of parameter values, but in a reasonable scenario yielding relatively short spells in the initial occupations, a comparison of wages for workers with different job tenures in a learning environment will understate the returns to tenure.

Figure 10: Estimated wage growth from different estimators:
True vs *OLS* vs *IV with within transform (Gibbons et al. 2005)*



E. CCP estimator employing Type I Extreme Value shocks

Assuming the i.i.d. Type I Extreme Value distribution (TIEV) of the utility shocks in the structural model provides the following convenient closed-form. To see this, let's assume that $Y_1, \cdots, Y_J$ are independent, non-identically distributed extreme value random variables with their location parameters $\mu_1, \cdots, \mu_J$ and common scale parameter σ. The distribution of Y_j is given by $P(Y_j \leq y_j) = \exp[-\exp\{-(y_j - \mu_j)/\sigma\}]$. Then, its mean and variance are $E[Y_j] = \mu_j + \sigma\gamma$ and $(\pi\sigma)/\sqrt{6}$, respectively, where γ is the Euler's constant. This family is max-stable:

$$P(\max\{Y_1, \cdots, Y_J\} \leq c) = \prod_j \exp\{-\exp(-\frac{c-\mu_j}{\sigma})\}$$

$$= \exp\Big[-\exp\Big\{-\frac{c - \sigma \log \sum_j \exp(\mu_j/\sigma)}{\sigma}\Big\}\Big] \tag{18}$$

From the expression, it can be found that $\max\{Y_1, \cdots, Y_J\}$ follows the extreme value distribution with location parameter $\sigma \log \sum_j \exp(\mu_j/\sigma)$ and scale parameter σ.

The structural model in Section 3 assumes that utility shocks are distributed i.i.d. Type 1 Extreme Value with scale parameter σ_u and location parameter $-\sigma_u \gamma$. Then, $\{v_j + \varepsilon_j^u)\}_{j=1,2,...,J}$ are extreme value random variables, with which the max-stable property yields that $\max_j\{v_j + \varepsilon_j^u\}$ also follows a TIEV distribution with the location parameter $\sigma_u \log \sum_j \exp\{(v_j - \sigma_u \gamma)/\sigma_u\}$ and scale parameter σ_u. It provides that, for an arbitrary alternative $q \in \{1, 2, ..., J\}$,

$$\begin{aligned} E_{\varepsilon^u}[\max_j\{v_j + \varepsilon_j^u\}] &= \sigma_u \log \Big[\sum_j \exp(\frac{v_j - \sigma_u \gamma}{\sigma_u}) \Big] + \sigma_u \gamma \\ &= \sigma_u \log \Big[\frac{\sum_j \exp(v_j/\sigma_u)}{\exp(v_q/\sigma_u)} \exp(v_q/\sigma_u) \Big] \\ &= -\sigma_u \log P_q + v_q \end{aligned} \tag{19}$$

Hence, with this particular distribution, the ex-ante value function can be rewritten with respect to the conditional value function associated with an arbitrarily selected choice q and the corresponding adjustment term which is a function of the CCP.